e-Research in the Life Sciences: From Invisible to Virtual Colleges

Lucy Power
Keble College

Thesis submitted in partial fulfilment of the requirements for the degree of DPhil in Information, Communication and the Social Sciences in the Oxford Internet Institute at the University of Oxford.

Trinity Term, 2011.
91,400 words.

Abstract

Lucy Power, Keble College

e-Research in the Life Sciences examines the use of online tools in the life sciences and finds that their use has significant impact, namely the formation of a Scientific/Intellectual Movement (SIM) (Frickel & Gross, 2005) complemented by a Computerisation Movement (CM) (Kling & Iacono, 1994) which is mobilising global electronic resources to form visible colleges of life science researchers, who are enrolling others and successfully promoting their open science goals via mainstream scientific literature. Those within this movement are also using these online tools to change their work practices, producing scientific knowledge in a highly networked and distributed group which has less regard for traditional institutional and disciplinary boundaries. This thesis, by combining ideas about SIMs and CMs, fills a gap in research that is typically confined to treating new tools as a part of scientific communication or in specialist areas like distributed collaboration but not in terms of broader changes in science.

Case studies have been conducted for three types of online tools: the scientific social networking tool FriendFeed, open laboratory notebooks, and science blogs. Data have been collected from semi-structured interviews, and the online writings of research participants. The case studies of exemplary use by scientists of the web form a baseline for future studies in the area.

Boundaries between formal and informal scholarly communication are now blurred. At the formal level, which peer-reviewed print journals continue, many academic publishers now also have online open access, frequently in advance of print publication. At the informal level, what used to be confined to water-cooler chat and the conference circuit is now also discussed on mailing lists, forums and blogs (Borgman, 2007). As these online tools generate new practices they have potential to affect future academic assessment and dissemination practices.

Acknowledgements

I am very grateful for the inspiration and guidance offered by my supervisors, Ralph Schroeder and Eric T. Meyer, with whom it was always a pleasure to meet and discuss (or argue) points around the social and technical factors shaping science. Thank you also to my enthusiastic, knowledgeable and indefatigable research participants for sharing their time with me.

I am indebted to my parents, Mary and Des Power, who fully understand the dissertation process, and whose editorial and psychological support helped me get to this point. I also very much appreciated the supportive environment of Keble College and in particular the support of my friends in the Keble MCR, as well as the time spent mucking about in boats on the Isis with the Kebelles. My sincere thanks also go to the Keble Association and the OII for providing assistance with travel funding for conferences.

On a personal note, Kat, thanks for being there for me. Lastly, I'd like to thank all of the past and present members of Schönheitspotential—you know who you are.

For Jack, Caitlin, Lucy and Ella, the next generation of Powers—
may you all have the opportunities I have had.

Table of Contents

Abstract iii
Acknowledgements v
Table of Contents ix
List of Figures xv
List of Tables xix
Chapter 1 Introduction 21
1.1 Theoretical grounding of this thesis 24
1.2 Research Questions 28
1.3 e-Research impact expectations 30
1.4 Methods 37
1.5 Scope of this research 39
1.6 Research Rationale and Contributions 41
1.7 Definitions 44
1.8 Thesis Structure 46
1.9 Background to the research 47
1.10 Literature Review 50
1.11 Conclusion 66
Chapter 2 Research Methods 68
2.1 Introduction 68
2.2 Preliminary work and research origins 69
2.3 Research design and methodology 71
2.4 Research Instruments and Procedures 87
2.5 Data analysis 109
2.6 Ethical considerations 116
2.7 Conclusion 121
Chapter 3 Online Scientific Networking: Making the Invisible College Visible 122
3.1 Introduction 122
3.2 Relevant literature 125

3.3 What is FriendFeed? 140
3.4 Research Participants 143
3.5 Research participants' Normal Use of FriendFeed 147
3.6 The Life Scientists Group 149
3.7 Network expansion online - from Biogang to FriendFeed 151
3.8 Further Detail and Discussion 157
3.9 Conclusion 175
Chapter 4 Open Notebook Science: Data in the Public Sphere 179
4.1 Introduction 179
4.2 Chapter structure and scope 181
4.3 The role of laboratory notebooks in science 183
4.4 Relevant literature 187
4.5 Open Laboratory Notebooks 209
4.6 Research participants 210
4.7 Inside the Open Laboratory Notebooks 216
4.8 Results and Discussion 256
4.9 Conclusion 270
Chapter 5 Scientific Blogging 273
5.1 Introduction 273
5.2 Relevant literature 286
5.3 Research Participants 311
5.4 Participant's Science Blogging 314
5.5 Blog Networks and Aggregations 329
5.6 PepsiGate - Scienceblogs network controversy 336
5.7 Arsenic Life - a blogging controversy 341
5.8 Results and Discussion 351
5.9 Conclusion 360
Chapter 6 Synthesis 362
6.1 Introduction - Science and Technology advancing together 362
6.2 The Open Science Scientific/Intellectual Movement 368
6.3 Computerisation and scientific work practices 379

6.4 Synthesising Theories - SIMs and CMs 384
6.5 Collaboration and Communication - Networking and Net-work. 386
6.6 Roles - Entry and Participation 397
6.7 Gatekeeping and Gatekeepers 407
6.8 Status and Rewards - Costs and Benefits of Participation 411
6.9 Conclusion 418
Chapter 7 Conclusions and Implications 421
7.1 Contributions 429
7.2 Limitations of this research 431
7.3 Implications for further research 434
Appendix A List of Research Questions 438
Appendix B Invitation Emails 439
Appendix C Schedule of Interviews 441
Appendix D Consent Form 442
Appendix E Interview Questions 444
Appendix F Content Codes 447
Appendix G List of Articles in Popular Science Publications 449
References 450

List of Figures

Figure 1. From theory to theoretical strengthening, the case study process.73
Figure 2. Embedded case research design (after Yin, 2003, p. 40). 76
Figure 3. Cases and Synthesis:relationship of case studies to theory. 80
Figure 4. Distribution of research participants 88
Figure 5. Screenshot of Google Reader Interface. 108
Figure 6. Example of research aspect and case coding in text 110
Figure 7. TAMs Analyzer coding window. 111
Figure 8. TAMs Analyzer Search Window 112
Figure 9. TAMs Analyzer Search Results Window. 113
Figure 10. FriendFeed screenshot showing *The Life Scientists* group 140
Figure 11. A "like" in FriendFeed. User clicks on Like link on post. 141
Figure 12. Commenting in FriendFeed. 141
Figure 13. Sharing on FriendFeed. 142
Figure 14. FriendFeed query asking for programming skills. 161
Figure 15. Discussion on FriendFeed. 163
Figure 16. ISMB Conference 2008 and 2009 FriendFeed groups 167
Figure 17. FriendFeed discussion about SubMeta funding 171
Figure 18. Jean-Claude Bradley announces SubMeta funding award 172
Figure 19. A page from one of Isaac Newton's experimental notebooks 189
Figure 20. Screenshot of an experiment on the UsefulChem wiki 222
Figure 21. The *dosol* experiment request spreadsheet on Google Docs 229
Figure 22. The Synaptic Leap OLN blog 232
Figure 23. Mat Todd Blog-based OLN 234
Figure 24. Cover of Nature Chemistry, October 2011, v3(10) 237
Figure 25. OpenWetWare home page. 239
Figure 26. Kochlab Open WetWare OLN experiment page 241
Figure 27. Results section of a sample Kochlab OLN page. 242
Figure 28. Screenshot of part of Maloney's open notebook dissertation 244
Figure 29. Cameron Neylon's LaBLog 247

Figure 30. Refrigerator temperature in a Southampton laboratory 248
Figure 31. The RRResearch Blog 250
Figure 32. Martin Johnson's Open Notebook at UEA. 254
Figure 33. Idea-saving in a short blog post 315
Figure 34. Sense-making blog post 316
Figure 35. Turning into a product 317
Figure 36. Blog post link in Twitter (Neylon, 2011a) 318
Figure 37. Blog post link in FriendFeed (Neylon, 2011b) 318
Figure 38. Screenshot of ScienceSeeker.org home page, July 2011 330
Figure 39. Example of Researchblogging.org icon in use on a blog post 332
Figure 40. Screenshot of Scienceblogs.com Life Sciences home page 337
Figure 41. Rosie Redfield's blog, RRResearch 342
Figure 42. Felisa Wolfe-Simon's Tweet 349
Figure 43. Combined SIM and CM theory concepts 385
Figure 44. Patterns of adoption of online tools observed in this research 388
Figure 45. Todd and Neylon conversation in Google + 416

List of Tables

Table 1 *Stages of Scientific Collaboration (from Sonnenwald, 2007, p. 671)* 61
Table 2 *Research Participants* 90
Table 3 *Interview information* 97
Table 4 *Types of Online Textual Data* 106
Table 5 *Ensuring validity in case studies* 115
Table 6 *FriendFeed research participant detail (n = 16)* 144
Table 7 *FriendFeed research participants (Data snapshot 18 July 2011).* 146
Table 8 *Open Notebook Science research participant details (n = 11)* 214
Table 9 *Open Notebook Science Claims Logos* 219
Table 10 *Scientific Blog Types and examples (non-exhaustive)* 279
Table 11 *Research Participants (n = 20)* 313
Table 12 *Examples of Scientific Blog Aggregations (August 2011)* 331
Table 13 *Scientific Blog Networks* 334
Table 14 *Summary of comments on Redfield arsenic blog post* 344
Table 15 *Illustrative Recruitment Activities and Products in the Open Science SIM* 377
Table 16 *Examples of patterns of online tool use by order of adoption* 390
Table 17 *Technology use by networked scientists (mentioned in interviews)* 401
Table 18 *Goals and Networked Activities in scientific online tool use.* 426

Chapter 1
Introduction

Science changes primarily through epistemological paradigm shifts (Kuhn, 1962), through evolving work practices—in this case due to computerisation (Kling & Iacono, 1994), and through scientific/intellectual change driven by resources being mobilised in the manner of social movements (Frickel & Gross, 2005).

This thesis investigates the use of online tools in the life sciences and finds that their use has had significant impact, namely in the formation of a combined Scientific/Intellectual and Computerisation Movement (SIM/CM) for open science, which is mobilising Internet-enabled tools to form coherent groups of life science researchers who are enrolling others and successfully promoting their goals in various forums, including via mainstream scientific literature. Those engaged with this movement are also using these online tools to change their work practices, producing scientific knowledge in highly networked and distributed groups which transcend traditional institutional boundaries. Although using an entirely informal, ad-hoc collection of tools in a network, these scientists display similar behaviours to formally-conceived and created collaborative network organisations; that is, sharing data, contributing to knowledge sharing via online networks, and cooperative co-creation of content, "facilitating cooperative work towards shared goals" (Dutton, 2008, p. 216). The scientists' behaviour signals the extension of online tools designed for social

networking and communications back into the scientific domain—a reversal of earlier computer-based innovations that transferred from scientific domains to the public (such as the Internet)—and also signals a viable social transformation in scientific ways of working.

The long-term impact of online tools in e-Research, defined in this thesis as is the use of information and communication technologies to aid research practices, is as yet unknown, but the indications from online communities studied in this thesis are that the scientists involved in this SIM/CM are consciously seeking to shift the existing boundaries of the life sciences, most notably in the areas of communication, scientific networking and the production of scientific knowledge. New and innovative forms of current research practices are emerging in online scientific communications through potentially transformative practices which open up the formerly hidden data of the laboratory, and, for those involved, bringing to the fore a range of unintended outcomes regarding their involvement in a prominent and public Scientific/Intellectual and Computerisation Movement.

Radical changes in scholarly communication have occurred previously. The introduction of Gutenberg's printing press in the fifteenth century made possible the production and distribution of printed reports of scientific exploration and endeavour, and this dissemination gradually reshaped scientific work practices, over time allowing them to evolve into the academic communications systems and work practices that are in use today. The influence of the Internet and in particular the World Wide Web

may be having similarly far-reaching effects on academic work. Although there is some evidence that email has had significant impact on scientific communication practices (Walsh, Kucker, Maloney, & Gabbay, 2000; Duque et al., 2005), the impact of other online tools has not been investigated to the same degree. The present research acknowledges both the social and technological factors shaping the observable changes—the social in the way that these scientists interact openly and display collaborative behaviours in contrast to the secrecy surrounding scientific activity until formal publication either in print or at conferences of peers, and the technological through the easy availability of computer technology and software that makes instant communication around the globe both affordable and practicable.

It should be noted that this move by some to open science is not taking place in isolation. The scientific use of online tools depicted in this thesis is but a single part of a larger group of changes happening within digital scholarship, in the form of scholars and librarians pursuing open access to scholarly literature (including those disciplines for whom preprint publication is now a normal part of scholarly production, such as the branches of physics which use the arXiv), and a resurgence of public interest in science, marked also by new approaches to online citizen science such as the Zooniverse[1] group of projects which allow the public to contribute astronomical data (Nielsen, 2011; Willinsky, 2006; Wilbanks,

[1] http://www.zooniverse.org/; see also Fortson et al., 2011.

2006; Fortson et al., 2011). What is different, and new, about the activities described in this thesis is the scientists' leverage of current Internet technology to expose science at the laboratory bench level, exposing much higher levels of data, of experimental method, and of the thought processes surrounding knowledge production in science.

1.1 Theoretical grounding of this thesis

The scientists studied in this thesis are aiming to alter the intellectual landscape of the life sciences by making all scientific discussion and discovery available to anyone with an Internet connection. This transformational goal challenges scientific orthodoxy, and the scientists' actions and interactions can be described using the theory of Computerisation Movements (CMs), proposed by Kling and Iacono (1988) and the General Theory of Scientific/Intellectual Movements (SIMs), proposed by Frickel and Gross (2005).

To model this shift to open science theoretically, I look to the agents of this change—the formation of a Scientific/Intellectual Movement (SIM) combined with elements of a Computerisation Movement (CM). Now that computers are ubiquitous in research settings, and highly functional software is just a click in a browser away from the user, the social choices made by scientists around their technology use are a relevant focus for research.

This combination of theoretical viewpoints is complementary because although both theories help to account for the conditions which drive

changes in science, each has gaps which the other can fill: Frickel and Gross's SIM theory provides an account of the factors, including the marshalling of resources, which drive intellectual change, whereas the CM theory considers the ways in which technologies are utilised within the social system to change work practices.

Frickel and Gross suggest that SIMs facilitate the "institutionalization of new social forms across the sciences and humanities" and provide a way of initiating changes in the "organisation, production, diffusion and transformation of ideas" (p. 225). Their General Theory of Scientific/Intellectual Movements is designed to describe social conditions that facilitate the emergence of SIMs, which if successful, in time, "gain adherents, win intellectual prestige, and acquire … institutional stability" (Frickel & Gross, 2005, p. 205). SIMs, in Frickel and Gross's account, emerge through marshalling resources behind an agenda that challenges established ideas or scientific orthodoxies (for example, acquiring or using existing publication channels, securing positions within professional bodies, and consolidating groups in disciplinary specialisms).

While Frickel and Gross examine the generation of a new intellectual movement, in this thesis what I will examine is not just an intellectual movement, but a movement which simultaneously wants to practice research in a new way; that is, with new technologies.

Hence, this theory needs to include how new technologies are introduced into ways of working together. The theory of Computerisation

Movements (CMs) proposed by Kling and Iacono (1988) acknowledges that the use of computers is a process involving social choices (investment, equipment, expertise) and technical choices (hardware and software)—that is, it looks beyond the conception of computerisation as simply the adoption and use of computers and information and communication technology tools, and acknowledges the social interactions that frame their use.

CMs are defined as "a kind of movement whose advocates focus on computer-based systems as instruments to bring about a new social order" and thus harmonise with the General Theory of SIMs (Kling & Iacono, 1988, p. 228). Together these theories support the argument that new tools and new ideas thus go together, an argument which is examined in detail in Chapter Six, Synthesis.

The CM and SIM approaches complement one another because they both address change at multiple levels—the micro level of individuals and laboratories, the meso level of academic institutions and inter-institutional collaborations among groups or movements within and between academic disciplines; and finally impact, both potentially and realised, at the macro level of global socio-technical systems. In addition, both allow the researcher to consider both technological and social developments. Both CM and SIM theory acknowledge that change is precipitated by the enrolment of both people and things (such as technologies) en route to making new phenomena more robust and eventually commonplace.

As noted above, problems for technology adoption in the sciences revolve around social as well as than technical issues. Hence what is required in a thesis which examines new online tools is a theoretical framework which does not exclusively foreground or favour either the technological or the social aspects. The Socio-Technical Interaction Network (STIN) approach aids this analysis as it explicitly avoids this bias by giving equal analytical weight to the social and technical (Meyer, 2006).

Complexity can be managed and explained by using ideas from the STIN framework to inform the research, as these ideas allow analysis of a comprehensive set of activities, including human motivations, social interactions and skills, organisations, equipment and protocols. STIN acknowledges that "technology-in-use and a social world are not separate entities—they co-constitute each other" (Kling et al., 2003, p. 54) and "emphasizes the mutually shaping relationship between research and collaboration practices, and the tools developed for the purposes of formal and informal communication and knowledge sharing" (Fry & Talja, 2007, p. 117). The STIN framework ensures identification and examination of all of the actors in a system—from the social to the technological. For example, Kling, McKim and King (2003) include eight entities in their STIN model of the scholarly journal production system: Internet technologies (WWW, computers, URLs), publishing formats (PDF, XML), articles, and actors: reviewers, authors, editors, publishers and scholarly societies. Following

this model, each case study in this research identifies and investigates the core technology, data flows and actors.

1.2 Research Questions

The present research examines whether the use of online tools is revolutionising, reinforcing and/or having unintended consequences for the research practices of academic life scientists. Such broad enquiry must necessarily be bounded by discrete research questions, which are listed below. This thesis examines seven aspects of scientific work practices; scientific roles, collaboration, open data, scientific communications, gatekeepers, effects on career status and rewards, and interaction with the media and public, with specific research questions examining each aspect.

Consistent with the theoretical frameworks discussed above, the questions below are designed to uncover the key changes wrought by the use of online tools by examining the core technologies in use, the core groups of people involved, the choices they make about technology, the nature of their interactions, and the incentives and barriers they face in driving forward their intellectual vision of open science.

1. Do online tools extend the professional roles of academic life scientists beyond their traditional ones?
2. Do online tools facilitate inter-institutional and interdisciplinary collaboration for academic life scientists, and if so, how?

3. How do academic life scientists use online tools (social networks, open laboratory notebooks, blogs) to disseminate different types of scientific data?
4. Does the use of online tools by academic life scientists circumvent gatekeepers in science and/or create new ones?
5. How do academic life scientists see that their engagement with the public has changed since they adopted online tools?
6. Do the various forms of online dissemination being used by academic life scientists yield rewards and status comparable to traditional outputs, or are there new mechanisms developing for reward and recognition of these activities?
7. Where in the chain of collecting and disseminating research materials are the incentives and barriers to open practices in the use of online tools by academic life scientists?

The research questions into these seven aspects contribute to answering the overall question of the thesis, whether the use of online tools is revolutionising, reinforcing and/or having unintended consequences for the research practices of academic life scientists. Essentially I conclude in this research that the scientists studied in this thesis are involved in a Computerisation and Scientific/Intellectual Movement which is successfully re-shaping their own research practices, and, in turn, also drawing attention to its activities via processes of promotion and enrolling

others, and subsequently provoking reactions and resistance from the rest of the scientific community.

Taken together, the answers to these research questions will provide an overarching view of the changes taking place in science due to the use of online tools, which can be seen as a coherent SIM/CM promoting open science.

1.3 e-Research impact expectations

Each embedded case within the main case study describes the core information and communication technology in use, and identifies, via analysis of seven aspects of scientific work practices, features of the SIM and CM emerging from the data. The STIN framework described above has guided identification of the seven research aspects within each of the three embedded cases, which connects the research questions to the theoretical framework outlined at the beginning of this chapter. The set of expectations about the impacts of e-Research in the life sciences is elaborated below by means of a short discussion of each research aspect.

Roles of scientific researchers

The roles of academic life scientists as laboratory researchers are bounded by norms which enable the production of research data through sets of specific experimental processes from hypothesis formulation, data collection and analysis by methods approved by their peers, through to journal article publication and/or conference presentation. Until fairly recently, the role of researchers in life sciences did not extend to activities

such as data curation and management, systems administration and programming, understanding and using legal and administrative systems, and/or more interaction with programmers and librarians, who traditionally fill these roles. The advent of the World Wide Web, and the increasing amounts of scientific data available online, appear to be guiding researcher involvement in these activities since online tools have made them accessible.

So, research question one asks to what extent are the roles of scientists extending beyond laboratory-based boundaries as they engage with using online tools which allow them to conduct research by publishing their initial thoughts and data world-wide rather than waiting until their final product has received the imprimatur of a journal?

Scientific collaboration practices

Collaboration is increasing across all disciplines (Wuchty, Jones, & Uzzi, 2007), but although many scientists collaborate, there is evidence that some find it more time-consuming and difficult to collaborate with colleagues at a distance than with those they can regularly meet face-to-face (Cummings & Kiesler, 2005). Despite the time costs of constant communication, scientists using online tools are more likely to engage in activities that keep them in daily contact with geographically distant colleagues (signalling presence through information broadcasting and aggregating tools such as blogs, RSS readers, FriendFeed or Twitter). Thus, the second research question asks whether the use of online tools examined in this thesis show that these practices result in a continuous and more self-aware process of reviewing

techniques and findings and allowing trustful conversations where ownership and recognition for interim results encourages greater collaboration and reflection among a larger cohort in an extended geographic and cultural sphere? Additionally, does the use of these online tools overcome the "tyranny of distance" that hampers distributed collaborative practices?

Openness of scientific data and discourse

Traditional collection and dissemination methods—from laboratory notebooks to journal articles—have mostly been a closed system, with all aspects of the research process confined to the laboratory and specified collaborators and not shared outside the formal research output of the peer-reviewed journal article or conference presentation. Some scientists are now choosing to open their closed systems to a wider range of their peers and/or the public, but this sharing ranges from partial to full practice of what its proponents have called open science. Whether the activity is sharing thoughts, protocols and methods via blogs, online video or text, sharing data via online databases and wikis, or sharing formal research outputs via open access publishing—all require a weighing of the costs and benefits of this openness by the scientists involved. This thesis asks what impact the sharing not only of ideas, but also the ways of working involved in open practice of research, will have on life scientists' day-to-day activities, their collaborations, their publication records and other aspects of their research and professional careers. Ultimately a key question will be whether opening previously closed systems accelerates the process of

scientific knowledge production and dissemination; that is, does openness confer greater benefits than not being open?

Informal and formal scientific communication practices

Formal scientific publications, including the well-established route of peer-reviewed journal article publication, are firmly entrenched as part of current life science practices. Journal publishers have embraced online publication, and a large majority of journals are available for online viewing or PDF download (provided the relevant institution or individual holds a subscription to the journal). In recent years, publishers have also begun to provide links to downloadable supplementary data such as high-resolution images, data tables and more detailed explanations of methods. However, some scientists have taken this trend several steps further and are publishing all or some of their experimental methods, data and results on the public Internet. What effect is the release of this information having on the networks, collaborations and working practices of the scientists who release it and those who see and perhaps also use it? Will such practices render formal scientific publications surplus to requirements—or at least an after-the-fact prize or award that increasingly has less to do with the way progress is communicated in science, similar to the role of preprints in some physics communities?

The roles of gatekeeper institutions

A series of gatekeepers have evolved to guard access points along the path from initial idea, through the process of research, to publication. Research funding bodies, institutions (as employers and funders), peers,

departments, industry standards bodies, and editors/publishers all act as gatekeepers in various stages of the research process. Scientific research can result in two kinds of data: some data are judged not to be of acceptable quality and consequently unpublishable; other data are used in the testing of hypotheses, and the writing of papers, which, providing they pass through the gatekeepers involved in the review process, are polished and eventually published. In research question four, this thesis will examine whether the widespread use of online methods is still following these paths, circumventing them in some ways or whether it has instituted a new set of gatekeepers.

Scientific career status management and rewards

The traditional successful academic scientific career has involved a journey from successful completion of a doctorate to promotion through tenured positions to professorial ranks. This journey required an understanding of current scientific funding priorities, the writing of successful grant proposals, the presentation of papers at relevant conferences and frequent publications in high-impact journals. These kinds of activities are evaluated in different ways across different academic boundaries (geographical or institutional), via job reviews and tenure committees, but now also through such enterprises as the UK's Research Assessment Exercise, Australia's Excellence in Research for Australia programme, and others. These research evaluations tend to measure output in traditional forms such as articles in journal publications. However, it should be noted that this career path is increasingly unattainable for many

academics; as for example in the USA, tenured positions make up only thirty percent of available jobs, while contingent and part-time faculty make up the rest (Knapp, Kelly-Reid, & Ginder, 2010). Now may be a good time to be pursuing alternative methods for publicising scientific work, such as can be done using online tools. Thus, research question six asks whether the rewards for using online methods are equivalent to the traditional gatekeeper-guarded publication route, and/or do these methods enhance chances of success at publishing? Do peers, employers and funding bodies recognise the online dissemination of research and discussion of those scientists who use the Web to disseminate their work? Are groups pressuring for or against online dissemination? Are we in an interim period where skills to succeed in both the areas of traditional academic life of papers and the less formal online environment are necessary?

Scientific engagement with the media and public

Traditionally, research is published in specialist publications available only to subscribing libraries and individuals, with research reported in newspapers, magazines, broadcast media and the like being limited to exceptional cases. But the researchers examined here aim to open up this process, making their work available to all. One of the distinguishing characteristics of the use of online tools, particularly blogs, is that the content they contain is open to the public, and usually allows for commentary and discussion. Research question five asks: what impact is this engagement with readers having on the scientists who blog? What are the implications of these activities for how the public engage with science?

Through an examination of scientific roles, collaboration, open data, scientific communications, gatekeepers, effects on career status and rewards, and interaction with the media and public across the three embedded cases (scientific networking, open laboratory notebooks, and science blogging), it can be seen that the changes that are being promoted via e-Research tools add up to an agenda (sometimes more and sometimes less implicit) throughout all these areas. This agenda embraces more openness, more sharing and collaboration, and redrawing the boundaries of communication by making them more transparent. Investigating these areas will therefore contribute to the overall research question of whether this engagement with online tools is changing the boundaries of traditional scientific practice, and to what extent engaging with these tools allows the participating scientists to form coherent groups which aim to challenge normal expectations of scientific practice. Thus the thesis will assess the extent to which these activities are revolutionary, reinforce existing practices, or produce unintended consequences. Examining these seven areas and answering this overall question will therefore make a contribution to identifying whether a more significant change in science is taking place – the question posed at the outset; namely, whether change in science should be seen in terms of revolution or evolution from one paradigm to another, whether work practices change in unintended ways, and additionally whether this shift can be seen as a successful or emergent

resource mobilisation behind a Computerisation and Scientific/Intellectual Movement to challenge existing ideas and practices in the life sciences.

1.4 Methods

A case study approach has been used to gather data for this thesis. Three online tools have emerged as the logical embedded cases within the overall case study of the consequences of the transfer of the computerisation of everyday life into the scientific domain.

The embedded case studies are ordered within an overall narrative which reflects the scientific research process; from information seeking, involving finding, filtering and assessing information; to laboratory work, involving data and information management and arrangement; and lastly the dissemination of research, involving publication, discussion and assessment of research outcomes. These processes are often intertwined, and as will be seen, the scientists involved also take advantage of hyperlinking capabilities to connect tasks across tools within the research process timeframes. The first case, FriendFeed, an online scientific social networking site, serves as an information finding and filtering tool for scientists, while the second case, Open Laboratory Notebooks (OLNs), looks at scientists managing their data and information openly. The third case, scientific blogging, examines the impact of online dissemination and subsequent online discussions. Although they progress the overall research process narrative described above, these cases are also heavily interrelated. In scientific practice these tasks often overlap, as participants may also use

each tool to support tasks across all three phases of the research process, given that the online tools in use are not solely designed for scientists, and offer a great deal of flexibility of use.

Each of the three embedded case studies examines the technologies the research participants are using, their work practices and the institutions they operate within, and the interactions they have with the public, their colleagues and peers. Several types of data are drawn upon in each of the cases: semi-structured interviews with participants, analysis of blog posts and web site content, and aggregated usage and linking data from the relevant web sites.

The Methods chapter explains in detail the rationale for choosing the three particular case study units of analysis and the empirical approach. This approach consists of closely examining how people came to use the technology, what kinds of communication they are enabling with it, how this affects their scientific work practices, and how it relates to the seven research questions. It also discusses ethical considerations arising from the research process.

For example, an examination of the interactions of scientists on FriendFeed is informed by data gathered from semi-structured interviews, as the participants themselves are the most appropriate sources of information about their motivations, the perceived outcomes and the discovery of unintended consequences of their actions. Interviews are also

the best source of data for sourcing details of the social processes taking place during the use of online tools.

Several data sources also complement one another, and additional sources of data come from the blog posts, comments and Twitter feeds of the participants, as these contain and make public their comments and discussions about other issues relevant to this research—such as participants' ways of finding information, how they find jobs, their perceptions of how what they do online affects their academic life, the practice of open science, and their own communication practices. Another level of analysis is provided by aggregated data from FriendFeed, which characterises the volume, chronology and authorship of posts and comments, as well as the social networks in which participants operate. Similar approaches and data sources describe the environments and practices within open laboratory notebook use and science blogging.

1.5 Scope of this research

Case study research is useful in leading to making connections to theory and in developing broader generalisations such as how these case studies connect to a conceptualisation and combination of the theories of Scientific/Intellectual and Computerisation Movements as is done in Chapter Six, Synthesis even though case study research in itself cannot be used to make general assumptions about a population (Yin, 2003). I have chosen to investigate three particular instances of online tool use: FriendFeed, a networked information aggregator, open laboratory

notebooks, and science blogs. FriendFeed and science blogs have been chosen because of their online prominence and relatively large user bases. Although there are many online tools, both purpose-built for science and appropriated for scientific use, they do not appear to have the popularity or the widespread use of FriendFeed and science blogs among the open science SIM/CM.

Open laboratory notebooks were selected because they represent an exemplar of what is considered by practitioners "the logical extreme of transparent approaches to research".[2] There are a limited number of open notebook practitioners in the life sciences, and the open notebook embedded case study will examine most of them. Despite their limited use so far, open laboratory notebooks have garnered considerable attention in the scientific community, and are regarded as potential models for enhanced scientific practices of data sharing and dissemination not just in the life sciences, but also in other disciplines (such as astronomy and mathematics). Within the embedded case studies, this thesis will not focus on patenting or intellectual property issues in science, or scientific innovation and competition in the scientific commercial sector, as these areas have been extensively studied elsewhere and fall outside the scope of this thesis (Hope, 2008; David & Hall, 2006; Lievrouw, 2004).

---

[2] Open Notebook Science. (n.d.) Retrieved from Wikipedia, 10 May 2010: http://en.wikipedia.org/wiki/Open_Notebook_Science, rephrased from Bradley, JC., (2007, December 13). Chemistry Crowdsourcing with Open Notebook Science [Web log post] Retrieved from http://usefulchem.blogspot.com/2007/12/chemistry-crowdsourcing-with-open.html

Sampling methods for this thesis were not limited by geographical location or national origins of the participants, although interviews were limited to those who could speak or write English. Details of specific limitations imposed by the methods chosen are explained in detail in the Methods chapter. Generally, the participants were eager to participate in this research and extremely cooperative, as might be expected given their participation in a movement which seeks to change science by making research processes and knowledge outputs more open.

1.6 Research Rationale and Contributions

We do not yet know where the current era of information access and communication will ultimately lead and what it will mean for science and scientists over the longer term—so we need to identify emergent changes at this stage, and the way to do this is to talk to practitioners and undertake systematic data collection. We need to discover what impact these technologies are having on involved scientists' daily work and research outputs, and how their activities are shaped or influenced by the types of online tools available to them. We also need to examine external influences, such as peer collaborators, institutional settings and disciplines, and professional and public responses to their increased openness.

As will be seen below in the assessment of literature in this area, several authors have pointed out that there is a gap in the research to date, particularly with regard to qualitative assessment of what the various developments in online tools and novel modes of scientific communication

entail for researchers. The present research aims to fill that gap by interviewing scientists who are using these technologies extensively and forming groups to promote more widespread use of online tools and to change scientific communication practices.

The present research is further aimed at understanding the changes taking place in life science research methods due to the use of online tools, assessing the social significance of these changes, and identifying their impacts on research practices. The case studies of scientists' use of the web (either as individuals or in groups) during the period of research will form a baseline, an assessment of a SIM/CM in-the-making, on which future studies will be able to build. In addition, any disadvantages or unintended consequences of their web-based activities will be explored with participants with the aim of providing insights into how the technology can be more seamlessly integrated into scientific work practices.

This thesis will also contribute original research in the sociology of science as it examines new and emerging trends in technology adoption in the sciences from the perspective of research processes and work practices. The use of online tools and data is different to email communications and significant in itself because of several key features; the fully public nature of the communications, the persistence of discussion, and the one-to-many and many-to-many nature of online discourse. Consequently this thesis will not specifically examine use of email lists or informal scholarly communication via email. Although email lists facilitate one-to-many

broadcasting and conversation, they are different from the areas to be investigated in the sense that one needs to sign up to a list to keep track of the active threads—and once a thread dies it is usually not revived. Communication via web tools is fully indexed by search engines, and thus is more public and findable than email lists, which may often be private (Fry & Talja, 2007). In addition, it is also persistent, in that each conversation, via a blog comment thread or via wikis, can be sustained, suspended or revived at any time.

There are many stakeholders in the scientific research process, including publishers, research councils and institutions. Institutions serve multiple roles as employers, ethics committee convenors and commercialization advice providers. This research acknowledges the influence of these stakeholders on scientific research, particularly in funding and publication scenarios, and this influence is examined to some extent in the research question about gatekeepers in the scientific process. However, as the focus of this research is the use of online tools by scientists, a detailed treatment of the role of these stakeholders is not required. For further research on stakeholders in these areas, see Willinsky (2006) on publishing, Borgman (2007) on scholarly communications, Evans (2004) on the commercial/academic biotechnology relationship, and Dougherty, Meyer, Madsen, Van den Heuvel, Thomas, & Wyatt (2010) on researcher engagement with web archives and libraries.

1.7 Definitions

Throughout this thesis, particular terms are frequently used which should be explained by their meanings within the bounds of this thesis. They are also defined at first point of use in the text, but collated here for ease of reference:

This thesis focuses on research practices involving ICT use in science. **e-Research**, also referred to as e-Science, cyberscience, cyberinfrastructure and e-infrastructure, is defined as "powerful and versatile computer-based and networked systems for research" (Dutton & Jeffreys, 2010, p. xi). Borgman (2007) similarly characterises these terms as describing "technologies to facilitate distributed, collaborative, information-intensive forms of research and learning" (p. 20).

Since all definitions of e-Research and e-Science involve the use of information and communication technologies to support research practices, this thesis will define e-Research as the support of research processes, such as formal and informal communication and collaboration, by means of online tools. e-Research and e-Science are used interchangeably within the literature on this topic to encompass both top-down and bottom-up initiatives for ICT use in science. The present research is not about programmes or sets of initiatives dependent on top-down initiatives and policies and so it would be inappropriate to use the labels of particular programmes or initiatives.

The **life sciences** are defined as the areas of science that study living organisms, and can include a wide variety of fields, such as systems biology, biochemistry, genomics, microbiology, and zoology.

Open science within this thesis means public sharing and communication of scientific data and theory, ideally extending the openness to all stages of the research process, with the goals of finding collaborators, removing redundancy and unnecessary repetition, and speeding up scientific knowledge production.

This definition is based on the ideals of scientists interviewed for this research, whose personal definitions of open science (and there is some variety among them) are discussed within the case study chapters. References to the **open science movement** within this thesis therefore refer to the elements of a SIM and a CM which are being used to drive the goals of open science. Note that this is not the definition of open science used by Merton (1973), David (2004) and Fuchs (1993), who define it as the *current* structures of knowledge sharing in science, albeit structures that they expect to evolve alongside the technology (Schroeder, 2007).

Science 2.0 is defined as the use of Web technologies to facilitate open science practices and dialogue.

Various acronyms are also used throughout, also spelt out at first use, and collated here for reference:

SIM - Scientific/Intellectual Movement

CM - Computerisation Movement

OLN - Open Laboratory Notebook

ONS - Open Notebook Science

1.8 Thesis Structure

This chapter, Chapter One, introduces the scope and boundaries of the research, presents the research questions, contextualises the choice of research case studies and briefly describes the methods used, and reviews literature around the origins and growth of modern life science, the influence of disciplinary cultures on scientific ways of working, scientific collaboration, and status and rewards in science. This introductory literature review examines research on broader issues in the life sciences; as detailed below, each embedded case study chapter begins with a review of relevant literature for that case, to support the rich description and narrative of each embedded case. Chapter Two details the methods used to gather and analyse data for this research. Chapters Three to Five present detailed case studies of online social scientific networking, open notebook science, and science blogging. Each case study chapter is structured in a similar way—the relevant literature for the case is reviewed, the data collection methods are briefly re-iterated, the case is described and then analysed using the seven thesis themes of scientific roles, collaboration, open data, scientific communications, gatekeepers, effects on career status and rewards, and interaction with the media and public. Chapter Six synthesises the discussions presented in the cases to and relates them back to the theoretical frameworks to provide an answer to the overarching

research question. Lastly, the Conclusion will summarise the findings and describe the substantive contributions, limitations and implications of this research, as well as recommending potential future research directions.

1.9 Background to the research

Before entering this investigation, it will be useful to examine the scientific, technological and social conditions leading up to the present situation. The World Wide Web of bioscience is barely two decades old. In November 1992, Tim Berners-Lee's catalogue of the world's Web sites showed just nine Bioscience sites.[3] Most of these, such as OMIM (Online Mendelian Inheritance in Man), the Genbank nucleic acid database, the REBASE restriction enzyme database and the BIOSCI mailing list are still in active use. In the intervening years, biological information sources and research aids on the Web have proliferated to the point where the term data deluge (Hey & Trefethen, 2003b) is in common use. Data management and preservation is now a point of concern for many scientific disciplines, as recent research highlights both the need for training in data manipulation, as well as infrastructure to store and manage data (Dutton & Jeffreys, 2010; Welsh, Jirotka & Gavaghan, 2006). In addition, as the Web has matured, so also have the opportunities for sharing and extending the practices and outputs of scientific research via the use of online networking, sharing and communication tools.

[3] http://www.w3.org/History/19921103-hypertext/hypertext/DataSources/bySubject/Bio/Overview.html

The original "read-only" Web of static documents has developed into the "read-write" Web, also known as Web 2.0, where user-contributed content and comment is easily facilitated, data is widely shared, and online communication and networking is commonplace. Proponents of Web 2.0 say that it is now gradually evolving into the Semantic Web, or the web of linked data as well as documents, which "will fulfil more of the Web's potential, in that it will allow data to be shared effectively by wider communities, and to be processed automatically by tools as well as manually" (Berners-Lee et al., 2006, p. 18). In this environment, life scientists are continuing to reconfigure their working practices using Web tools. It will be argued that the Web facilitates flexibility of contribution to the professional and public discussion of science, greater interdisciplinary collaboration, less experimental repetition and failure, and enables the connection of disparate and distributed data and wider dissemination of raw data.

The life sciences, defined above as the areas of science that study living organisms, can include a wide variety of fields, such as systems biology, biochemistry, genomics, microbiology, and zoology. Currently, work in the Life Sciences commonly deals with large datasets and involves conduct of collaborative work with other groups, and the output of this research is often of interest to the media and the public. Additionally, the rise of applied natural sciences over the course of the twentieth century has left applied life sciences at the forefront of scientific enquiry, particularly with

regard to receiving funding (Gabler & Frank, 2005). For these reasons the Life Sciences are a significant and fast moving field employing new methods and thus a good choice for a study of research methods involving ICTs.

Working practices in life science laboratories still display many of the hierarchies and structures described by earlier scientific ethnographies such as Latour and Woolgar's *Laboratory Life* (1979), and other laboratory-based studies by Knorr-Cetina (1981) and Lynch (1985). Knorr-Cetina (1999) further examined changes in molecular biology and physics cultures, and states that the life sciences, as represented by molecular biology, are centred around the researcher as an epistemic subject. That is, "the laboratory, experimentation, procedures, and objects obtain their identity through individuals" (Knorr-Cetina, 1999, p. 217).

Since these studies were published, the situation has changed considerably as the Web has become a platform which allows laboratories to connect and further their research—among other things, by sharing data, working practices and results in formal and informal online discussions. Some scientists contribute to, curate and discuss data shared online. Laboratories are exposing the previously unseen and private routine details of their experiments online through "open laboratory notebooks" in wikis and/or blogs, and thousands of scientists blog publicly about their own work, and comment on others' work.

In this environment, many scientific fields, notably physics, are using top-down "big science" (de Solla Price, 1963) to manage their data analysis and dissemination—for example, initiatives such as the UK's Grid service[4] and the various projects at the European Organisation for Nuclear Research (CERN), such as those surrounding the Large Hadron Collider.[5] The life sciences have also responded to these opportunities by embracing e-Research technologies where, for example, scientific documentation can be created and edited in real time by multiple collaborators. As will be demonstrated in this thesis, this extends across the entire research life cycle, from research thinking through to experimentation and dissemination of results.

1.10 Literature Review

Perspectives on change in science vary according to the approach used to assess them. Information scientists examine changes in science by looking at how citation, authorship and attribution patterns are changing. Both information and computer scientists also look at changes from a human-computer interaction (HCI) perspective, that is, how the screen in front of the individual life scientist is changing and how it enables them to do things differently in terms of information retrieval. This thesis focuses on the invisible college[6] of informal scientific interactions, which is being made

[4] http://www.ngs.ac.uk/

[5] http://public.web.cern.ch/public/en/LHC/LHC-en.html

[6] Invisible colleges are the informal social interactions of scientists (de Solla Price, 1961; Crane, 1972). This concept is discussed more fully in the Chapter Three literature review.

visible via the use of online tools, and is thus not restricted to examination of citation or authorship trends, or to individual user practices. This thesis examines a group of life scientists, who, it is argued, form a conscious Scientific/Intellectual and Computerisation Movement attempting to reshape the boundaries of science practice, communications and knowledge dissemination via the use of online tools.

The literature on how change occurs in science is vast. Within this chapter I concentrate on the changes most relevant to the scope of the present research. In each case study in the following chapters, namely scientific online networking, open notebook science and science blogging, I will also address the previous research particular to aspects of that case. Chapter Three, Online Scientific Networking using FriendFeed, discusses research on invisible colleges and informal scientific communications, online communications such as mailing lists, and research on social networking sites. Chapter Four, Open Notebook Science, covers research on laboratory life and practices, the history of laboratory notebooks, and research around electronic data collection and retention. In Chapter Five on science blogging, the literature review discusses research on blogging and bloggers, the small body of literature about academic blogging, and scientific communications with the public.

e-Research in the life sciences has elements that could be analysed from the vantage point of several social science disciplines, all of which have contributions to make, but none of which exclusively covers the research

problems defined earlier in this chapter. Library and information sciences and social informatics contribute scientometric data, including webmetric and bibliometric data and research about how scientists publish and communicate formally and informally (Kling, McKim, & King, 2003; Borgman, 2007; Bar-Ilan, 2008; Barjak & Thelwall, 2008). Apart from the general theory of Frickel and Gross (2005) discussed above, there are a number of case studies that explore the formation and survival of specific groups of researchers (Parker & Hackett, 2012) and the construction of epistemic cultures (Kuhn, 1970; Whitley, 1985; Knorr-Cetina, 1999), including within the life sciences (Knorr-Cetina, 2007). Within sociology, theories about how technological innovation spreads (Rogers, 2003) and how social networks are used to seek and share information (Granovetter, 1983) provide a framework for explaining how ideas and communication technologies spread through a population.

Investigators in the fields of computer-supported cooperative work and computer-mediated communications (CMC) have researched collaboration practices (Cummings & Kiesler, 2005) and electronic communications, particularly email (Walsh, Kucker, Maloney & Gabbay, 2000; Duque et al., 2005). Key findings from this research indicate that email contributes greatly to overcoming coordination issues in research (such as arranging schedules and meetings) but does not necessarily help to solve culture issues (such as asking for advice or making decisions), or show a net positive effect on collaboration. The present research reaches beyond email

to examine the effects of other types of online communication on collaboration.

Cummings and Kiesler (2005) studied large project collaborations and found that distance can cause difficulty for successful collaboration, but the negative impacts can be reduced by concerted communication and coordination efforts. Further studies of distributed research collaboration and the practices of collaboratories (also known as virtual research environments) elaborate the conditions supporting successful and causing issues within unsuccessful collaborations, which are explored in more depth in Section 1.10.3, Collaboration, below (Finholt, 2002; Sonnenwald, 2007; Olson, Zimmerman & Bos, 2008).

There is little research connecting this rich area of study with the changes currently taking place as scientific researchers adopt new communication strategies and models of data sharing using online tools. As will be shown in the science blogging chapter, for instance, there are as yet few studies on the use of blogs as a communication and dissemination tool by scientists. Most recently, Borgman (2007) and Hine (2008) produced detailed studies of changes in scholarship due to the influence of digitisation, but neither examined specific online activities such as blogging or scientific networking.

1.10.1 Origins and growth of modern life science

Studies of how science is organised and how scientific fields develop and interact attempt to identify patterns within particular disciplines and

across science as a whole (Whitley, 1984; Hackett, Amsterdamska, Lynch & Wajcman, 2007). Some previous research has shown that issues for technology adoption in scientific research do not revolve around technical implementation issues so much as social issues, which are a crucial challenge to effective distributed and interdisciplinary collaboration (Cummings & Kiesler, 2005; Borgman, 2007). These challenges have been studied in various ways, but there is as yet little research on the potentially transformative changes generated by use of the Internet, specifically on the possibilities of online interaction and collaboration.

Research on the sociology of science and e-Research technologies can inform the present work. For example, the case studies and thematic chapters in the collections by Dutton and Jeffreys (2010), Jankowski (2009) and Olsen, Zimmerman and Bos (2008) explore the benefits and constraints of advancing information and communication technologies (ICTs) on research practices. However, there has not to date been a substantial body of work that deals with distributed, multi-institutional, Web-based collaborative projects or activities such as science blogging, open laboratory notebooks and data sharing and curation.

There are also a number of specific studies about how the Internet and related technologies are affecting scientific practices and outputs. In a 2006 European-wide survey of informal academic involvement with Internet technology across five disciplines, Barjak found a positive correlation between involvement with the Internet (maintaining some kind of personal

Web presence) and productivity in formal scholarly publishing, characterised in his study as research productivity (Barjak, 2006). In a related finding, Vakkari (2008) administered a nation-wide survey of users of the Finnish e-library FinELib, and found that scholars felt that access to online research literature improved the quality of their work via faster access to a wider range of literature, and also prompted more publications in international journals.

Further research is needed to confirm these findings with measurement of objective variables such as time spent looking for resources, the increase or decrease of the variety of journals now available online and what influence types of informal academic communications, such as blogs and social networking, are having on the generation and spread of new academic ideas and research questions. This kind of research will clarify how use of the Internet translates into higher formal research productivity.

In the life sciences, despite a comprehensive annual listing of new and existing databases in the Journal of Nucleic Acids Research, there has been little research on the social impact of databases (Galperin, 2008). Both Leonelli and Evans have told the story of a model organism *Arabidopsis Thaliana* (a plant) and assessed its impact on the field through study of an online database (TAIR, The Arabidopsis Information Resource) (Evans, 2004; Leonelli, 2007). However, these two studies focused on other issues, namely epistemology and the commercial/academic biotechnology relationship, rather than the use of the online database as a research tool.

1.10.2 Disciplinarity

Many studies acknowledge that there is distinct variation among academic disciplines, something that academics embedded within their own disciplines often overlook. Several studies will be useful in examining disciplinary organisation in relation to life science research practices. Becher's (1989) investigation into higher education disciplines found several dimensions of disciplinary organisation, one determined by the cognitive characteristics of the discipline, and the other by the social dimensions of disciplinary organisation, defined by the intensity of interactions with others and people-to-problem ratios. Within the present study, it appears that the scientists involved in using online tools have been able to overcome potential cultural barriers to interaction by concentrating on mutually shared goals such as performing science openly. Whitley (1985) used similar concepts of "mutual dependence" and "task uncertainty" to characterise the types of knowledge practices in a discipline, such as knowledge and theory sharing, co-authorship, reliance on technology, and collaboration within and between organisations.

Whitley's (1985) theory of the intellectual and social organisation of scientific fields distinguishes between the intellectual (theories, topics, and research objects) and the social practices (funding, reputation and rewards) of academic disciplines. Whitley asserts that co-variance between these two can be explained by task uncertainty (which is whether the knowledge discovery tasks of a field involve pre-determined and documented techniques and research objects, or are more ad-hoc and contingent on

research outcomes) and mutual dependence (which is the specific usefulness and interoperability with other data of researchers' results (functional dependence) and the research goals of the field, including high level coordination of research programmes and outcomes (strategic dependence)). The interactions of task uncertainty and mutual dependence, according to Whitley, shape knowledge production in the sciences.

Fry & Talja (2007) applied Whitley's theory to the production, relevance and use of informal scholarly communication on the Internet: namely mailing lists, home pages and other digital resources. They interviewed 72 academics across seven disciplines, including ten from environmental biology. The biologists were largely laboratory-based, and Fry and Talja found that they had low degrees of task uncertainty (as is common in experimental science fields) and high strategic and functional dependence (Fry & Talja, 2007). Fry and Talja found that these biologists were less likely to produce and share resources on the Internet, particularly pre-prints of papers, and this finding is echoed in the discussions of the life scientists profiled in this thesis, who are attempting to mobilise open science and open publishing of results on the Internet against this prevailing trend in their discipline. The authors cite low task uncertainty as a driver for researchers not needing to communicate outside their research groups; it is possible that the data deluge may be increasing task uncertainty across many biological disciplines and driving some to communicate and

disseminate their research more widely, thus increasing their functional dependence on others' data (Fry & Talja, 2007).

Bettencourt, Kaiser, Kaur, Castillo-Chávez and Wojick (2008) investigated the emergence of new scientific fields over time via bibliographic analyses, and found that the evolution of new fields is not dissimilar to epidemiological and population growth models, although they moderated the epidemiological model to include typical characteristics of scientific interaction, such as conferences, workshops, PhD programs, and similar (Bettencourt, Kaiser, Kaur, Castillo-Chávez, & Wojick, 2008). They found that scientific ideas spread much more slowly than diseases, but that the spread of ideas is fostered by sustained and repeated contact via these typical scientific interactions. The sustained and repeated online contact between scientists in this research may thus be a conduit for the spread of open science ideas.

Hesse, Sproull, Kiesler and Walsh (1993) surveyed 257 oceanographic scientists to investigate how scientific disciplines network to advance science and found that scientists most active on the networks available at the time were also the most productive: "Scientists who use the network more also produce more papers, receive more professional recognition from their peers, and know more physical oceanographers" (Hesse, Sproull, Kiesler, & Walsh, 1993). It will be shown in this thesis that online tools are allowing the sustained and repeated contact necessary for the spread of ideas, publication, professional recognition and consider that they have a

wider network of colleagues; a fact that may also contribute to the potential success of the associated Scientific/Intellectual Movement.

Gabler and Frank (2005) seek to explain the rise and fall of natural science disciplines over the 20th Century. They propose that the popularity of life science disciplines today derives in part from the relationship that scientists have to the phenomena they are investigating; that is, whether they are viewed as "active investigators in a dynamic universe" (for example, biology, chemistry and physics) rather than participating as "passive observers" of phenomena (for example, astronomy, botany and zoology) (Gabler & Frank, 2005, p. 183).

These shifts in popularity are also observed by Powell, White, Koput, and Owen-Smith (2005) in their investigation of network ties in biotechnology over a twelve-year period. Powell et al. mapped the formation, stabilisation and dissolution of network ties among scientists, institutions and private business to see how the "opportunity structures" of the field were affected by these changes (Powell, White, Koput, & Owen-Smith, 2005, p. 1134). They found that for new entrants to the field, trend-following and choosing partners based on similarity were significant, but their findings did not support the hypothesis of "winner-takes-all". The strongest finding was that a diversity of choices and partners, which they called multiconnectivity, was the best predictor of successful tie formation. This finding is also, as we shall see, supported by the research presented in this thesis (although this thesis deals mainly academic researchers, whereas

Powell et al.'s study focuses on links with commercial biotechnology firms), as multiple examples of successful interdisciplinary collaboration are discovered within this research, and disciplines are being reconfigured online (Dutton, 2005).

In comparison to life sciences, online activities in physics have been studied fairly widely, particularly with regard to their publication practices, which are unusual among scientists (Moed, 2007). The arXiv[7] physics, mathematics, computer science, quantitative biology, quantitative finance and statistics preprint archive, and similar repositories, serve as a "universal digital library" for most physics fields (Nentwich, 2004). Physics is also making use of distributed (Grid) computing for tackling massive amounts of data. For example, the data output of the Large Hadron Collider at CERN is analysed using widely distributed Grid-based resources. However, physicists are usually only handling data from one source, the output of the experiment they are currently running. Life scientists are often dealing with distributed data from many sources and many experiments, and attempting to bring it together to form a coherent whole. Solving the associated problems of data normalisation and standardisation form part of the reasoning behind life scientists promoting open data. Thus we will need to see if the publication and data practices of physicists are shared by life scientists.

[7] http://arxiv.org/

1.10.3 Collaboration

Issues in distributed collaborative work, within science and beyond, such as communication and coordination difficulties, have been extensively studied. Sonnenwald's 2007 *Annual Review of Science and Technology* chapter on collaboration studies identified four stages of scientific collaboration from the existing literature: foundation, formulation, sustainment and conclusion. Within those stages, many inhibiting and facilitating factors were identified from the research literature, beginning with scientific, political, socioeconomic, resource accessibility, and social and personal networks. These factors remain in place as the collaboration progresses, while new factors also emerge in further stages, such as the emergence of ICTs and intellectual property considerations in the formulation stage.

Table 1

Stages of Scientific Collaboration (from Sonnenwald, 2007, p. 671)

	Stages of Scientific Collaboration			
	Foundation	Formulation	Sustainment	Conclusion
Factors	Scientific	Research vision, goals & tasks	Emergent challenges	Definitions of success
	Political	Leadership & organisational structure	Learning	Dissemination of results
	Socio-economic	Information & communications technology	Communication	
	Resource accessibility	Intellectual property & other legal issues		
	Social networks & personal			

These stages and factors stress that historically, collaboration is a linear process, and provide a useful metric against which to measure the less formal collaborations emerging within the open science movement studied in this research. Sonnenwald's overview also encompasses a number of more specific aspects of collaboration, which can now be discussed in more detail inasmuch as they have also been addressed by other scholars.

Cummings and Kiesler (2005) studied large project collaborations and found that distance can cause difficulty for successful collaboration, but the negative impacts can be reduced by concerted communication and coordination efforts. Walsh and Maloney (2007) also found that disciplinary culture and coordination were problems in scenarios where there were high levels of scientific competition due to an unwillingness to freely share data and ideas, and that phone and face-to-face interaction did not significantly reduce those issues. Another interesting (and unexpected) finding was that asynchronous communication (in this case, exchange of email with individuals or groups) "may be critical for keeping collaborations on track"(Walsh & Maloney, 2007, p. 725). Walsh and Maloney also suggest that "technologies that provide regular information on progress, that allow answering quick questions without necessarily demanding immediate responses, and that allow processing off-line, may be key to facilitating scientific group work" (2007, p. 725). The present research investigates the kinds of tools described by Walsh and Maloney, and also investigates

whether an individual scientist's approach to sharing can overcome these disciplinary hurdles.

Olson, Zimmerman and Bos's *Scientific Collaboration on the Internet* (2008) includes several biomedical science case studies, which examine top-down initiatives such as the National Institute of General Medical Sciences (NIGMS) collaborative project research grants of the early millennium and the Biomedical Informatics Research Network. These project overviews found that although many projects used a range of available technologies to facilitate collaboration, it takes time to foster widespread use of these technologies to build a "culture of team science" (Rogers & Onken, 2008, p. 212). The research fostered by these grants also resulted in project Web sites and widespread availability of the data produced, even for those outside the funded projects (Rogers & Onken, 2008).

Shrum, Genuth and Chompalov (2007) examine "big science" project collaborations in physics, but focus on multi-institutional projects taking place without the use of distributed computing, viewing the structure of collaborations via the available documented data production practices (Shrum, Genuth, & Chompalov, 2007). Shrum et al. wish to identify the organisational forms that "integrate or segregate structural elements of collaborations; these forms shape the autonomy of member scientists in producing knowledge" (Shrum et al., 2007, p. 4). Although studying a different area of science, this thesis similarly examines online data

produced by scientists, in a highly informal online setting where the autonomy of individual scientists is assumed.

Hine's 2008 study of systematics examines how taxonomists work with ICTs, in particular the recent transformation of the field into a "cyberscience" (Hine, 2008). Hine used ethnographic and historical data to support her findings that although there was strong historical continuity between the aims of systematics pre-and post-"virtual culture", there was also a need for conscious reassessment of how those aims could be achieved using ICTs. The present research, while paying attention to Hine's call to "bear in mind that these technologies [the ICTs in use] both form sites for the enactment of culture and are cultural artefacts that come to carry meanings for their users", also extends Hine's findings by asking whether the actions of the researchers studied in this thesis are attempting to reshape the aims of their disciplines by focusing on extending research openness (Hine, 2008, p. 53). Hine's findings suggest that in-depth case studies of changes in life science fields are needed to unearth the complexity of changes across fields when new ICTs are adopted, and the case studies in this thesis will contribute to this effort.

1.10.4 Status and Rewards in Science

Status and rewards in science have mostly been investigated via bibliometric studies, which examine authorship and citations and attempt to determine the structures of academic co-authorship networks. Cronin (2001) has drawn attention to the increasing levels of what he terms

"hyperauthorship" or multiple authorship *in extremis*, which is supported by the findings of Wuchty, Jones, & Uzzi (2007), who report increasing dominance of authorship teams in all academic disciplines (Cronin, 2001; Wuchty et al., 2007).

The question of how to appropriately allocate publication credit is still open as Birnholtz (2006) shows in an examination of the problems around authorship and obtaining credit for large physics collaborations which may involve hundreds or even thousands of collaborators, impossible to list in a paper. He concluded that most physicists still think that informal attribution is not enough recognition for academic career reasons, but the problem of how to properly attribute credit has not yet been resolved. Like Cronin, he suggests an alternative model for attributing credit, such as is used in film productions, of listing contributions by specific role. Online, as we shall see, this debate about attribution is flourishing, particularly as the job of data curator is becoming seen as a valid role within the life sciences.

Other studies examine status in academia via elite interviews, talking to successful academics to determine how they navigate the social worlds of academia, showing that the basis of interactions between disciplines are shaped by opinions and beliefs about the worth of those disciplines (Lamont, 2009). Zuckerman (1977) investigated success in science, as measured by interviews with members of the scientific elite, Nobel Prize winners. Zuckerman found that contrary to popular conception, Nobel Prize winners are rarely lone geniuses, but like most scientists, including

those in the present study, collaborate with others to produce the work that ultimately wins the prize (Zuckerman, 1977).

Research to date thus shows that the areas of life science, disciplinary cultures and status and rewards in science are well-examined, often from the perspective of top-down research infrastructures such as large collaborations and "big science" projects. However, what this research does not cover are gaps which the present research will fill—the growth of online collaboration with the use of novel tools, and how the adoption of these novel modes of scientific communication across the entire research process is changing science.

1.11 Conclusion

This chapter has presented the background for the thesis and the gaps in research that are to be filled. A combination of Scientific/Intellectual Movement and Computerisation Movement theory provides the framework for investigating the impact of e-Research in the life sciences.

This investigation is framed by seven research questions into particular aspects of scientific work, introduced in this chapter, which cover scientific roles, collaboration, open data, scientific communications, gatekeepers, effects on career status and rewards, and interaction with the media and public. This chapter has also described and justified the methodology and scope of the thesis. The first literature review examined the background to life science research, research disciplines, collaboration, status and rewards in science, to be followed by reviews of relevant literature within each case

study chapter. Chapter Two, Research Methods, will provide a detailed description of the research.

Chapter 2
Research Methods

2.1 Introduction

Using embedded case studies, this thesis studies the activities of scientists via the distinct but interconnected phenomena of their scientific social networking, use of open laboratory notebooks and science blogging. This chapter describes the research design, and the methods used to enrol participants, gather data and perform data analysis. A rationale for the methods used is provided. Difficulties and solutions generated during data gathering, processing and analysis encountered during the research process are described.

As stated in the Introduction above, the three case study chapters of this thesis (Chapters 3, 4, and 5), are structured in a question-and-answer style format, with the detail of each case described before the seven research questions are posed and answered. The Discussion of each case study chapter considers the findings. This case study reporting style connects with the methodological approach described below, which uses the seven research questions as the basis for making sense of the data obtained during the course of this research (Yin, 2003). Chapter Six synthesises the case study findings and the theoretical concepts of this thesis.

This chapter opens with a description of the intellectual catalyst for this work, why it is important to science research, and the development of the research process. Following that, the research design and methodologies are

discussed, including details of sampling, oral and email interview protocols, and the use of online documentation and discussions. Details of data analysis techniques are then provided, including information about data processing and the qualitative computer-assisted coding methods used to classify data. Lastly, a number of ethical considerations for the present research are discussed.

In respect of ethical considerations interviewee anonymity (or rather, participants' collective rejection of it) is of special interest, and other information to do with ethical matters is provided about data protection and copyright relevant to Internet research and digital data storage. This chapter and its appendices aim to provide enough information to reliably reproduce this research design. With qualitative methods, particularly interviews, exact reproduction of the research itself is not possible, as it is a detailed snapshot of a particular time period.

2.2 Preliminary work and research origins

It may be useful for the reader to understand the origins of my interest in life science research. Before undertaking this DPhil research, I had practical work experience related to this topic. I worked as a project manager on two European Union Fifth Framework-sponsored projects that aimed to develop search engines for scientific papers and microscopic image collections (*e-BioSci* and *BioImage*). As a project manager working for an online academic publisher (Ingenta plc.), I attended all of the project meetings at locations including scientific centres such as the European

Molecular Biology Organisation (EMBO) in Heidelberg, Germany, the Sanger Institute in Hinxton, near Cambridge, and the Institut National de Recherche en Informatique et en Automatique (INRIA) and the Centre National d'Etudes Spatiales (CNES) in Nancy, France. Talking to the scientists and bioinformaticians working on the project, I became aware of the issues life scientists were facing with regards to data handling, publication, storage and search methods. These issues revolve around the fact that scientists in many fields are producing a digital data deluge (Hey & Trefethen, 2003b); and the numbers of research papers keep increasing; for example, the Royal Society reports that the number of publications worldwide has risen from 1.09 million to 1.58 million in the five years between 2002 and 2007, a 45% increase (Royal Society, 2011).

Beginning DPhil work at the Oxford Internet Institute (OII) in October 2007 further sensitised me to this topic. Several studies and papers published by the OII were relevant, including Dutton's work on social research and emerging digital technologies (Dutton, Carusi & Peltu, 2006) and David et al.'s work on openness and e-science (David, den Besten & Schroeder, 2006; David, 2004).

As part of my coursework in the early stages of the DPhil, I did a pilot study of the topics of blog posts on ResearchBlogs.com (an aggregation of blog posts that discuss peer-reviewed research) and ScienceBlogs.com (of which more below in Chapter Five, the scientific blogging case study). I attended Cameron Neylon's two-day workshop *Blog based approaches for*

electronic laboratory record keeping[8] (in Abingdon 28/29 February 2008), which is discussed in Chapter Five, Open Laboratory Notebooks. I met Jean-Claude Bradley, an open laboratory notebook user, open science advocate and participant in this research, at the 2008 e-Research conference in Oxford, and the discussion with him allowed me to identify several ideas for potential cases and particular research participants. Collectively, these activities helped solidify my research topic and approaches to data collection and analysis for the thesis.

2.3 Research design and methodology

Qualitative methods can be used to build up a rich picture and strong narrative of actions within social worlds. The credibility of qualitative research relies on several factors: that the methods chosen have been rigorously and mindfully followed; that all aspects of validity for the selected methods have been considered; that attempts to generalise are to theory, rather than populations, and thus provide a basis for further study.

Paying attention to the factors which produce reliable and valid research is important, as a potential criticism of qualitative research is that it may lack the reliability of quantitative methods as the data it produces is more dependent on the interpretations of the researcher, and the sampling methodology is usually not as representative. Factors which reduce this risk include the use of a consistent method to classify the data, such as

[8] http://blog.openwetware.org/scienceintheopen/2008/02/29/placeholder-post-for-workshop-comments/)

computer-assisted qualitative content analysis. Several other checks on reliability can also be used, such as ensuring construct validity by checking the findings with the research participants, using purposeful sampling methods to select the most relevant cases, using the replication logic of gathering data from different sources, and comparing the evidence found in the research to the existing literature. These and other checks of research validity are explored in more depth in Section 2.5.2, Procedures for handling validity and reliability.

2.3.1 Justification of research design

What drives research is the knowledge that something is happening which is worth exploring for a variety of reasons. Intrinsic focus on something "just because it's happening" does occur, but for most research, there is an instrumental reason for investigating a given phenomenon (Stake, 2005); for example, finding out the what, why and how of certain conditions that affect scientific output. To make the research worthy of pursuit, the researcher must surround the questions about what is happening with a methodological framework which allows a focus on the 'right things', or sensitising concepts (Blumer, 1954), capture of relevant data, and analysis of that data in a manner which brings the researcher back to answering the research questions, as well as figuring out whether the research supports its theoretical propositions. This section describes the chosen framework for this research, and assesses it in relation to the questions being asked.

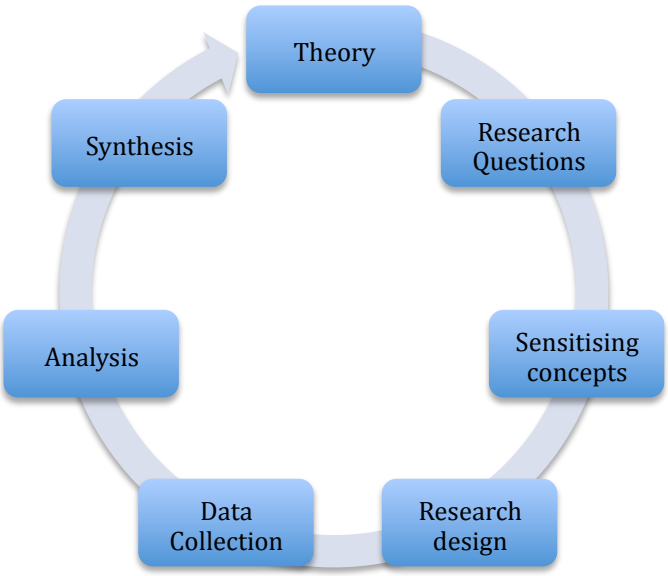


Figure 1. From theory to theoretical strengthening, the case study process, via readings of Yin (2003), Blumer (1954), Parkhe (1993) and Patton (2002).

The study of online research practices and their effects on scientific knowledge production is a nascent field with few prior studies. There is little description of the ways in which academic scientists are working with online technologies, and in particular there is little information on what technology choices they are making, and how this affects their daily work. Therefore, the methodology chosen for this research needed to allow access to the range of behaviours and decision-making that are involved in online scientific interactions. For this reason a qualitative approach was chosen in order to unpack the concepts involved.

2.3.2 Background to qualitative research

Social science research aims to understand particular social phenomena and to make sense of the social worlds occupied by people and their activities in particular contexts. Abstraction in social research is achieved by

conceiving the social world in terms of "classes of objects and of relations between such classes" (Blumer, 1954, p. 3)—just as it is in physical science research. The methods applied to achieve this abstraction are therefore the drivers of success or failure of the research. Blumer aimed to allow social scientists to approach abstraction with the introduction of sensitising concepts, designed to focus thinking about the theories framing the research.

Blumer, writing in 1954, characterised qualitative social scientific research as having "grave shortcomings" when compared to physical scientific research in terms of theory development (Blumer, 1954, p. 3). However, since that time, a wealth of methodological research has aimed to tie particular methods to particular types of questions in social research, and to bolster social scientific theory development.

One way is to start with existing knowledge as those familiar with experimental method would do and to see where the data takes already existing conceptualisations. Another is the grounded theory approach whose proponents argue that existing conceptualisations can blind the researcher to the enormous variety and stimulation of the social world which, if ignored, means that new paths are not only not trodden but not even imagined (Glaser and Strauss, 1967).[9]

[9] This references Glaser and Strauss's first joint publication on grounded theory; this thesis is not using grounded theory so I will not attempt an explanation of Glaserian grounded theory versus Straussian grounded theory. See Grbich (2007) for a full explanation.

A third way is to try, as this thesis does, to have the best of both worlds, and to use existing knowledge of how people use technologies and emerging ways of communicating using computers where the boundaries may stretch further than we can imagine at present. Thus, existing research can sensitise us to data illustrating earlier conceptualisations but can also allow us to become aware of new developments emerging out of current scientific practices along with rules of interaction.

In the 1990s as well as into the present century, qualitative methods have become refined and use technology to support the entire qualitative research process. The present research was able to collect data from the online thoughts and discussions of scientists, use Skype to conduct and record interviews, use Express Scribe software to transcribe them, and use the computer assisted qualitative analysis software TAMS Analyzer to do qualitative textual analysis. Computer software therefore played a central role in this study of computer-assisted scientific knowledge production as part of the method, as well as being a subject of the research.

2.3.3 Embedded case design

The present research is designed to analyse the use of online tools in science, with reference to three embedded cases—online scientific networking, the use of open laboratory notebooks, and science blogging. The three embedded cases are chosen for their focus on a range of scientific working practices which represent the scientific research process; from

information seeking, to laboratory work, and to the dissemination of research.

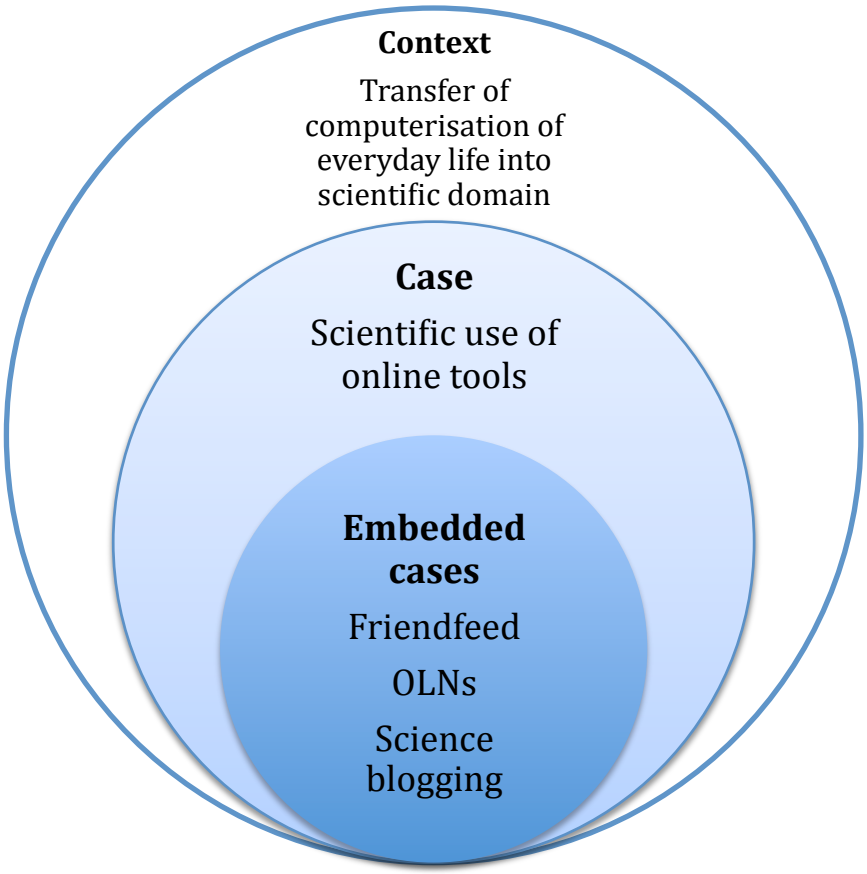


Figure 2. Embedded case research design (after Yin, 2003, p. 40).

As Yin (2003) observes, a common definition of case studies wrongly cites the *subject* or *topic* of the case study as the definition of the case. Yin's own definition shows that certain research problems have two clearly defined facets which make them suitable for case study approaches; firstly, a *scope* which (a) "investigates a contemporary phenomenon within its real-life context", (b) where the "boundaries between phenomenon and [real-life] context are not clearly evident" (that is, the phenomenon is intertwined with everyday life, and not something that could be extracted and tested by experiment); and secondly, *technical conditions* which (a) have "many more

variables of interest than data points" (that is, multiple contexts and sources of data), (b) rely on "multiple sources of evidence, with data needing to converge in a triangulating fashion", and (c) benefit from "the prior development of theoretical propositions to guide data collection and analysis" (Yin, 2003, p. 13-14).

In relation to this research, these case study elements are easily identifiable. The phenomenon of use of online tools for research is indeed *contemporary*. Little work has been done to date that examines the use of online tools in scientific research, although there is a rich body of literature on epistemic cultures, and a growing amount of scholarship on scholarly communications since the advent of email (Koku, Nazar & Wellman, 2001; Hine, 2008; Knorr-Cetina, 2007; Borgman, 2007).

Context in this case refers to the use of these tools within the everyday work practices of scientists. Such use is not yet an established practice for many scientists, and for those who are using online tools to change the way they produce scientific work, the boundaries overlap with traditional ways of working. A discussion of how an emerging phenomenon can be studied is provided below in the Research Time section.

With regard to *boundaries*, the cases studied in this thesis are appropriate for the case study approach for two reasons: firstly, the practices being studied are not widespread, which constrains the size of any potential sample; secondly, the use of online tools is enmeshed in the daily work practices of the scientists, and cannot be extricated from this context and

disciplinary culture. However, each embedded case study is also clearly bounded by the tool being studied, which helps to decide what to include and exclude within each case.

The *variables of interest* in these cases range from macro to micro: from institutional rules to the daily work practices of individual scientists. The variety of behaviours and contexts within the cases warrants rich description of the kind recommended for case studies.

This study uses *multiple sources of evidence* from a variety of data sources to corroborate the case study narrative and examine the research questions. These sources include semi-structured interviews, email discussions, blog posts and comments, and FriendFeed posts and comments. More detail of these sources is provided in Section 2.4, Research Instruments and Procedures.

For *theoretical propositions*, as is made clear throughout this thesis, this research is guided by the seven aspects of change in scientific work, which form an analytic framework for discussing the impact of online tools in science across the three case studies. The seven aspects that frame and inform the research instruments described below are in turn informed by the theories driving this research, using a combination of the Theory of Computerisation Movements (CMs), which acknowledges that the use of computers is a process involving social choices (Kling & Iacono, 1994), and the General Theory of Scientific/Intellectual Movements (SIMs), first proposed by Frickel and Gross (2005). Frickel and Gross's SIM theory

provides an account of the factors, including the marshalling of resources, which drive intellectual change, whereas the CM models consider the ways in which technologies are utilised within the social system to change work practices. Major changes in scientific ways of working need to be driven by an organisation, such as a SIM, which uses technology to organise itself and enrol new participants in the movement.

Little work has been done to date on the impact of online tools on scientific working practices, as demonstrated by the paucity of research literature in this area. It could therefore be tempting to succumb to purely exploratory and descriptive research, but as a researcher I am conscious of avoiding "mindless empiricism", characterised as "the elevation of technique over all other considerations" (Adams & White, 1994, p. 573). This is achieved by being not purely descriptive, but also engaging in theoretical tying-together—moving from the case descriptions to answering the research questions. Within each case study the question and answer method approaches this, but a synthesis chapter (Chapter Six) is also used to focus on the links between the research questions and theory.

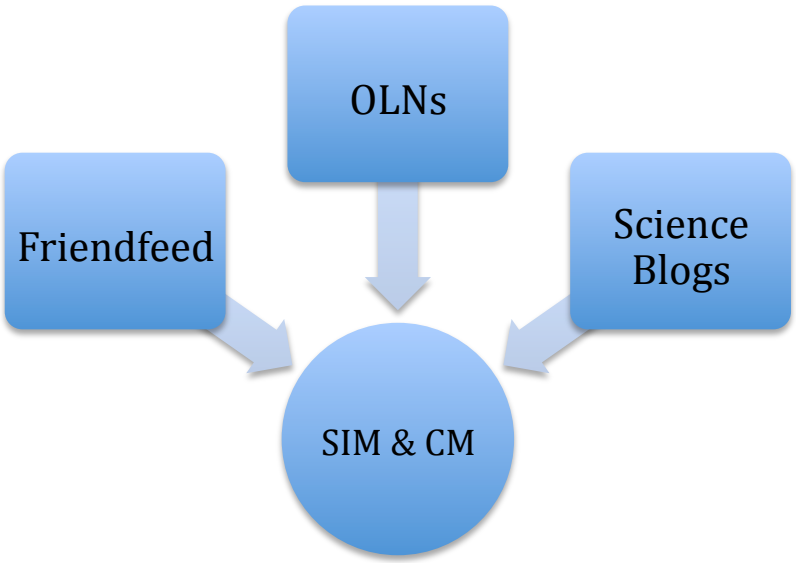


Figure 3. Cases and Synthesis - relationship of case studies to theory synthesis.

Interviews were designed to gather data on the use of online tools in work practices and settings, and the seven research aspects (collaboration, communication, status and rewards, work roles, interaction with the public, scientific gatekeepers, and open science) shaped the design of the interview questions. For example, the question construction style of "what has changed about [*topic of interest*]" helped to tease out information about communications, use of computers, and open science. Further details of interview structures and protocols are provided below in the Research Instruments section.

The units of analysis for this research are the use of three online tools in life science research settings. Therefore, the focus of data collection is on the work practices and experiences of scientists who are using such online tools. This focus is narrowed to the examination of three particular types of tools: (1) scientific social networking through the FriendFeed *The Life*

Scientists group, (2) the use of open laboratory notebooks in six scientific laboratories, and (3) communication with peers and public via science blogs, in particular the challenges and controversies encountered while blogging. Most data is collected from two main sources – interviews with scientists heavily involved in one or more of the above activities, and other textual data taken from sources written by the research participants, such as blog posts, comments, FriendFeed posts and comments, and some Wiki pages.

The research aspects of change in the roles of scientific researchers, collaboration practices, openness of scientific data and discourse, scientific communication practices, the roles of gatekeeper institutions, scientific career status management and rewards, and scientific engagement with the media and public are sensitising concepts for this research, and inform the departure point for content analysis, as will be explained below. Each of the three online tools was selected because it showed evidence of the activities described in the seven research aspects of the thesis, as well as activity at different organisational levels, from the individual scientists through to the laboratory group, collaboration or discipline. FriendFeed and science blogs also have relatively large user bases from the scientific community, and open laboratory notebooks, although their user base is small, exemplify the practice of open science.

Parkhe proposes that case studies are an essential part of the early development of any field of study, in order to capture the 'soft' concepts

which "permit deeper understanding and sharper delineation of concept domains" (Parkhe, 1993, p. 230). The present research aims to examine the growth and development of a small group of scientists using online tools for everyday work, networking, and activism, who were identifiable by their actions and displayed differences between their activities and approaches to scientific practice and the larger scientific population using more traditional methods. Little would have been gained by surveying or experimenting on a random sample of the scientific population – quantitative methods would not have helped build a rich picture of the open science movement, which, although small, has had measurable impact on parts of science.[10] A previous study of the impact of digital resources in the humanities found that specialised communities have small but active communities of users, which can nevertheless produce considerable impact in the field (Meyer, 2011).

As Stake (2005) puts it, "the allocation of attention to contexts will be based partly on the distinction between intrinsic and instrumental purposes"(p. 64)—that is, the researcher knows which contexts to focus on based on the research questions derived from theory, rather than the study promoted by an intrinsic interest in the topic as a singular phenomenon. The present research studies a social process—a rich mix of complicated interactions across time. Although interactions between individuals in a

[10] See, for example, later references to *Nature* editorials and articles published in the last few years by Open Notebook Science practitioners cited in Chapters Three and Four, and listed in Appendix H.

network and the size of a network can be measured using methods such as social network analysis (Barabasi, 2003), the diagrams thus produced will not help to answer the research question of what has changed in the social world of science since online tools became available. Case study methods, which aim to gather rich data about the development of particular social phenomena, are thus the most appropriate method of inquiry for this research.

A case study is "instrumental" when it seeks to understand something wider than just what is being examined (Stake, 2005). Thus, the role of theory in these case studies is to focus the case on particular sensitising concepts (Blumer, 1954). Instrumental cases can be selective about context to answer the research questions, whereas an intrinsic study must pay attention to all contexts. For example, details of exactly how scientists update their social networking sites may be useful for a study with a Human-Computer Interaction (HCI) focus; but for a study like this one, the actual content of what they post in these online spaces is likely to reveal much more. This is because their online and interview content contributes directly to understanding what is different about their scientific work practices. This is not to say that the scientists' HCI practices are irrelevant to this study—they contribute to the understanding of how online tools have been integrated into their everyday work practices—but in terms of the key sensitising concepts of this research, they are of secondary interest. For example, participating scientists' remarks about the time it takes to learn

and use these tools is germane to understanding the potential barriers to uptake and use by others.

"Sensitising concepts" continue to inform how social scientists interpret their research data. "Sensitising concepts offer ways of seeing, organising, and understanding experience; they are embedded in our disciplinary emphases and perspectival proclivities" (Charmaz, 2003, p. 259). Another important facet of case studies is how the contexts, activities and characters of the case are described. Case description must necessarily be detailed and thorough; since case studies are bounded by time and other contexts particular to them, such as social and economic conditions, or the possibility of Hawthorne effects,[11] it is necessary for the researcher to capture much detail of the singular events within a case study (Bailey, 1992). Detailed description is not only warranted but is essential if there is any anticipation of an attempt to translate the situations described to other settings or people—in this case, if the research hopes to generalise to a theory that particular actions by scientists form part of a Computerisation and Scientific Intellectual Movement—one needs to be able to understand which particular contexts and conditions led to the actions of the participants.

[11] Essentially, the knowledge of being studied may make research participants act in ways that they would not do if unobserved. See Section 2.6.1 for my comments on non-participation in the online tools I am studying. The name originates from Landsberger's 1950 study of the Hawthorne electric works.

2.3.4 Research Time

The present research took place over the years 2007-2011. Data continued to be collected and analysed right up to the period when the thesis was finalised. This illustrates one of the difficulties of studying a phenomenon in active development – so much keeps happening that it is difficult to know where to "draw a line" under the research. In late 2010 and the first few months of 2011 several new initiatives in online open science were launched. For example, new externally-supported blogging platforms such as *The Guardian* online science blogs, websites like *Figshare* (open sharing of figures, datasets and multimedia), *Colwiz* (online collaboration), and *ScienceSeeker.org* (a science blog aggregator) were all launched in late 2010.[12] Also, in December 2010 the NASA arsenic bacteria story[13] developed, which perfectly encapsulated the kinds of discussions happening around scientific blogging and its interaction with formal scientific publishing and peer review, so although it was discovered late in the research, it was included in the blogging case study.

Similarly, within already-interviewed cases, important developments continued to emerge. For example, I interviewed Steve Koch of the University of New Mexico in August 2010. At that time he stated that there

---

[12] http://www.guardian.co.uk/science-blogs, http://figshare.com/, http://www.colwiz.com/, http://scienceseeker.org/

[13] The NASA Arsenic bacteria story was a publication controversy—a peer-reviewed research publication by NASA researchers found that a certain strain of bacteria could grow in a non-Earth environment, which quickly became a popular media story along the lines of "Alien life", publicised by NASA. The research was questioned strongly by the scientific blogosphere, which found much to criticise in the research. Full discussion of this story can be found in Chapter Five, Scientific Blogging.

were no external citations of his laboratory's open notebooks. In late January 2011 I was scanning reports of the Science Online Conference which had taken place on January 13-16th, and noted that Koch had written about his work being found and cited by researchers previously unknown to him. I contacted him about it via email and he told me the story, and he also wrote about it in a new blog post on February 5th.[14] This development was important to include in the reporting of the Open Laboratory Notebooks case. Eventually, though, it was necessary to stop online exploration and discovery and to concentrate on writing the thesis, although I am sure developments did not stop during this period.

Similarly, this thesis does not attempt to cover every incident and development in online tools for e-Science between the years 2007-2011, simply because there have been too many to write about. For example, the FriendFeed case study concentrates on one FriendFeed group, *The Life Scientists*, and some of the activities of those who used it extensively in 2009-2010. There were thousands of posts and comments, by the research participants and many other scientists, in other related FriendFeed groups, but limiting the study to the origins and activities of one group allowed for a more distinct narrative and the discovery of interesting concepts such as "bursty work", described in the FriendFeed case study in Chapter Three. Despite this proliferation of information and ongoing changes and developments, the thesis conclusion will provide an overall assessment of

[14] http://stevekochscience.blogspot.com/2011/02/open-data-success-story.html

the current significance of this phenomenon as a whole, even if it is still growing and changing.

2.4 Research Instruments and Procedures

2.4.1 Sampling

2.4.1.1 Theory and background

A wide variety of purposeful (also called purposive) sampling methods exist (Patton, 2002). Purposeful sampling aims to gather samples which are selected to yield rich data for case studies, and is different from probability sampling, which seeks to represent a larger population by sampling randomly. The case studies in this thesis used a purposeful sampling approach, seeking individuals within each case who appeared to be deeply involved with the online tools of this research.

The present research seeded the sample with a few scientists who could be considered critical cases within the online activities studied by the cases. Cameron Neylon and Jean-Claude Bradley are examples of these scientists. They are widely known within the open science movement, both blog about their and others' work, use Open Laboratory Notebooks, and are heavy users of FriendFeed and other scientific social networking tools such as Twitter. As an example of using the snowball technique, both Neylon and Bradley helped identify further participants for research by naming them in interviews.

As well as asking them both to recommend interview subjects I also used my seed scientists (Neylon and Bradley) to identify people of interest

to this research by looking at their FriendFeed profiles and postings, and their blog and twitter postings. From these sources it was easy to identify their most common correspondents and informally identify their online networking peers. From this informal survey I drew up a list of potential research participants across each case. For each potential research participant I collected preliminary information via online searches (contact details, discipline, position, institution, location, addresses of blogs, FriendFeed and Twitter). Every participant was asked to provide the names of other potential research participants, and I found that this quickly led to saturation – that is, the same people were named multiple times by multiple participants. In my email approach to the next scientist in line I would mention the previous participants who had recommended them (see Appendix E for sample emails). Further details of interview approaches and protocols are available below in Section 2.4.2, Interviews, and in the relevant appendices.



Figure 4. Distribution of research participants

Sampling was not limited by geography. Although it might have been possible to limit this investigation to UK scientists only (and thus give a higher likelihood of face-to-face interviews), it would have meant not capturing many of the active online scientific networkers and researchers in the population of interest. Accordingly, researchers were interviewed face-to-face in London and Oxford, and via Skype in the USA, Poland, Australia, Hungary, France, and more distant parts of the UK.

Research participants represented a range of scientific career positions. Table 2 summarises the career positions of the participants, which are reasonably equally distributed across the six categories, although this was not consciously sampled for, as is discussed in Sampling Strategies, below. The sample is dominantly male; which seems to be consistent with genders in science. 2006 statistics from the National Science Board (USA) show that 40% of doctoral degrees in science and engineering are awarded to women. Munger (2010) estimates that one third of the 2100 blogs registered on Researchblogging are women (some are pseudonymous with no gender indication). However, this is not an area of focus for this research; for research on questions of gender differentials in scientific practice see Keller (1978); Long (1990); Fox (2001); Stack (2004) and Leahey (2006).

Table 2

Research Participants

Position	No.	Friendfeed	OLNs	Blogging
Undergraduate student	1		1	
Graduate student (PhD)	4	2	2	3
Postdoctoral Fellow	2	1	1	2
Mid-Career Scientist (Lecturer, Senior Lecturer, Associate Professor)	8	8	4	8
Professor	4	2	3	4
Non-Scientist (with science PhD or MSc)	3	3	0	3
Total	22	16	11	20

Table 2 also displays the amount of overlap between the various cases, and how most research participants were able to contribute to each case. Most participants kept a blog, and many of them were also active FriendFeed users. The OLN users were substantially both bloggers and FriendFeed users. Nearly all OLN users also used FriendFeed and had a blog. None of the research participants solely used FriendFeed; they either had a blog, and/or an OLN.

2.4.1.2 Reaching sample saturation

Case study sampling should target information-rich sources, that is, the "validity, meaningfulness and insights generated from qualitative inquiry have more to do with the information-richness of the cases selected and the observational/analytical capabilities of the researcher than with sample

size" (Patton, 2002, p. 245). This section discusses what criteria were used to decide when enough data had been collected.

As mentioned above, I used a snowball sequential sampling method, which has resulted in a high degree of overlap between cases, the specifics of which are discussed in the Research Participants section of each case study chapter. The population being drawn upon is academic scientists in the biosciences and some biochemists, members of a large population (approximately 817,000 researchers as of 2007 (UNESCO, 2011).[15] Not specifically by design, but due to the small numbers involved, the Open Laboratory Notebooks case is as far as can be ascertained, a considerable proportion of those practicing Open Notebook Science; that is, eight of fifteen (53%) based on a Wikipedia page maintained by practitioners.[16] The FriendFeed case includes all of the OLN users and a snowball sample of other users who were active in the Life Scientists group on FriendFeed, and the Science Blogging case includes those who blog from the first two cases and a snowball sample of science bloggers.

---

[15] Note that this is an extremely difficult number to estimate. UNESCO divides scientists by fields, and estimate 545,000 scientists in 'Natural Sciences' in 2007, and this figure does not include 'Medical and Health Sciences'. UNESCO figures are missing for many countries, including the United Kingdom, the United States, and Australia, so this figure is very likely to be a significant underestimate. Despite this missing data in their official statistics, UNESCO's 2010 Report on Science reports total numbers of scientists in all fields for 2007 in the United States as 1.425 million (UNESCO, 2010). The Royal Society, using figures from the same UNESCO source, estimates total numbers of scientists in all fields at 7.1 million in 2007 (Royal Society, 2011).

[16] http://en.wikipedia.org/wiki/Open_Notebook_Science, based on people listed in Active and Partial/Pseudo categories, does not include notebooks requiring logins or group sites which support the concept (e.g. Open Wetware is listed). Accessed 20 April 2011.

Obviously, I have not sampled a statistically significant proportion of science bloggers and FriendFeed users, but that was never the aim of this sampling approach. The present research aims to examine the cases of FriendFeed use, open laboratory notebooks, and science blogging within the context of the theoretical propositions that online activities are supporting the formation of a Scientific/Intellectual Movement and a related Computerisation Movement. Hence, once enough information had been collected on these cases to provide for a rich narrative of their activities, and when a point of saturation or repetition of information was reached in interviews, data collection could cease.

2.4.1.3 Participant withdrawal

One participant in this research withdrew post-interview. This participant was a student who indicated that they believed their participation in this research could have a negative effect on their then current job hunt. Although they had participated in open research themselves, they wished to draw a distinction between open research and talking about open research. Given the topics of our discussion it would have been difficult to anonymise this participant due to the context in which their work would have been discussed. It is regretful that this occurred, as this participant was the only one who felt their online participation might have had a negative effect on their scientific standing, although this statement was somewhat contradicted by other evidence offered within our discussion. Perhaps it could have been avoided if I had

decided to anonymise everyone participating in this research, although I think this would have been difficult due to the specialised nature of their online activities, and also would have made it impossible to quote from any blog posts or comments, as Google and other search engines are very good at identifying specific phrases and providing their context.[17] Anonymisation for the sake of a single participant would have been difficult, and would have meant withholding a lot of the rich description from the case studies.

2.4.2 Interviews

2.4.2.1 Development, pilot and maturation of interview questions.

Semi-structured interviews were undertaken to supply both primary and supplementary data sources for all three of the case studies. Interviews were piloted with two scientific colleagues from the University of Oxford and modified as a result of these findings and comments.

Interview questions were designed to stimulate discussion of the main research aspects and the participants' involvement with technology and the open science movement. There were four broad areas of discussion in the interview:

- Participant background and path to current position.
- Everyday routine work and use of technology within their current position, as well as changes in routine across time (for example, comparing graduate student days to now).
- Questions about use of blogs, open laboratory notebooks, and/or

[17] "The goal of Blog Search is to include every blog that publishes a site feed (either RSS or Atom). It is not restricted to Blogger blogs, or blogs from any other service" Google (2011). Retrieved from: http://www.google.com/support/blogsearch/?hl=en

FriendFeed.

- Discussion of open science and science in Web 2.0 (including the participant's definitions of these terms).

Formulation of questions was modelled after Rubin and Rubin's model of main questions designed to elicit overall experiences, which lead to the use of follow-up questions and probes, to build on the initial answers and get a deeper understanding of the phenomenon and context being explored (Rubin & Rubin, 2004). Again following Rubin and Rubin's guidelines, *Tour* and *Minitour* questions (asking the participant to guide the interviewer through particular experiences) were used to encourage the participants to provide unfiltered answers, as well as comparison or contrast questions—asking participants to sort experiences into opposites (best/worst) then following up with questions about why they were put in those categories (Rubin & Rubin, 2004).

Questions aimed to elicit reflections on participant's activities in science and on the Web, and draw out the ways in which using the Web had impacted upon their scientific work, without specifically referencing the idea of an open science movement.[18] For example, the opening question asked them to discuss their work history:

Can you describe the path that led you to your current job?

Which led to a question about their current work:

[18] Participants were aware from the research information sheet I provided that I was investigating online activities like blogging and OLNs with particular reference to open science, but it was not specifically stated that the theoretical driver for this thesis was the investigation of the formation of scientific/intellectual and computerisation movements. See Appendix D for details.

Can you describe your average daily work routine?

Which in turn led to discussion of the changes in work routine and computers:

Could you tell me how things have changed with your work and computers between your student years and now?

Questions about specific online activities were introduced after these explorations of background, and the focus of the next set of questions depended on which activities the participant was most engaged with, which was established by prior online investigation or discussion with the participant. The examples below show some questions designed to explore the participants' involvement with scientific blogging:

Can you talk me through how you got started with blogging – late December 2003, right?

Can you step me through your process of writing a blog post, from initial inspiration to writing?

Similarly, with open laboratory notebooks:

Can you talk me through how you got started with your open notebook?

What processes and people were involved in setting it up?

What have been the best and worst things about the open notebook?

What makes [worst thing] so difficult? What makes [best thing] so enjoyable?

Interviews ended with discussion about open science, collaboration, dissemination of work, and an appeal for recommendations of further participants:

How do you find collaborators for projects? Has anything changed about that?

What does the term "open science" mean to you?

How would you describe "Web 2.0" to someone else?

Can you think of anyone else I should talk to?

Most participants were able to recommend colleagues or online connections for further interviews, as noted in the Sampling section above.

2.4.2.2 Conducting interviews

The interview protocol for both face-to-face and telephone was as follows:

1. Participants received and retained a study consent form supplied beforehand (either hardcopy or emailed PDF) (See Appendix D).
2. Participants read and signed the consent form (or emailed back with consent)
3. Consent procedures, including anonymity, were reinforced orally by the interviewer at the start of the interview
4. The interview questions were asked and participants were encouraged to answer in full and with frankness
5. The option of interviewee anonymity was again emphasised at the end of the interview.

Requests for interview were made via email (example email texts are supplied in Appendix B). Positive responses were received from every person who was approached. In some cases the interview could not be scheduled immediately, so was held later at a mutually agreed date. The

full interview schedule is provided in Appendix C. Average interview metrics are reported below in Table 3, Interview information.

From these explorations I created a spreadsheet of potential participants which listed their contact details, links to their online activities and details of their current work and institutional affiliations, where available.

Each participant received a copy of the research information sheet and consent form (Appendix D). Participants were asked to complete the consent form with some basic information. Each participant was also asked if they wished their interview to be anonymised at the start and finish of each interview. In cases where consent forms were emailed, the consent was obtained via an email message from the participant's email account. Follow-up questions were asked via email.

Table 3

Interview information

	Total	Mean	Range
Interviews	22	-	-
Length of Interviews (hh:mm)	13:21	00:45	00:30 - 1:09
Length of Transcripts (thousands of words)	138	5	3 - 9

The research did not include other qualitative methods such as participant observation. Most of the participants are extremely diligent in recording, broadcasting and discussing their activities in the online forums that they are active in.

2.4.2.3 Telephone Interviews

Most of the interviews for this project were conducted using Skype. One overarching feature of my research was that I was interviewing academics or those working in the private sector who had gained at least a PhD, so had an understanding of academic research processes. Interviews with these types of participants display "characteristics particular to academic elite groups that most facilitate the strategies discussed ... [below] a willingness to talk articulately, and often at length, about a particular topic and a willingness to participate in, and an understanding of, research processes and values" (Stephens, 2007, p. 213). Below I have summarised and commented on the issues which should be negotiated and considered during elite telephone interviews (Stephens, 2007).

Interruption - as there are no visual cues, I quite often interrupted the participant during the course of the interview if I wished to probe further on a topic they had already moved on from, or if I did not realise that they had not finished speaking. As Stephens also found, none of the participants seemed to interpret this negatively.

Topic control - Stephens characterises this as unsolicited utterances that alter the flow of conversation. Although I acknowledge these frequently occur, a more useful characterisation would be *steering* - moving the conversation in the directions required to cover the interview questions, as well as allowing for serendipitous ideas and examples to emerge. I felt that I had good control over the flow of interviews, even though at times the

conversation needed to shift to a new topic, for example moving the conversation from FriendFeed to blogging.

Lack of visual communication - related to the issues above, the inability of both parties to always accurately judge when to interrupt or keep talking was observed, although I did not think it created any particular issues with the flow of the interview conversation.

Articulation (being able to hear and understand each other) - was not usually an issue, although during two or three interviews the Skype "line" became sufficiently crackly that my participants could not hear me. In these cases a hang-up and re-dial solved the problem.

Holding the telephone (may hinder note taking) - Not a concern as I was using a hands-free Skype headset.

Controlling the environment (of the interviewee) - Stephens (2007) suggests asking the participant if they are ready to speak, which I always did at the start of the interview.

Bringing preparatory materials - I usually had a printout of the interview protocol as well as a browser window open with the participant's various online presences. This proved useful more than once as I could quickly find a blog or FriendFeed post as they referred to it in conversation, and ask further questions about it if necessary. Some interview notes were taken by hand, although I mostly relied upon the recording devices and just listened and responded during interviews.

Recording - recording using Skype was easy. The software used would automatically start recording after a designated period of time (10 seconds). Recording was orally confirmed as okay with each participant, although it was also mentioned on the information and consent form (Appendix D).

2.4.2.4 Recording and Transcription

Telephone interviews were recorded using a Skype plugin for Macintosh called Skype Call Recorder.[19] Face-to-face interviews were recorded using an iPhone. I transcribed the majority of the interviews for this project myself, using the free Macintosh software Express Scribe,[20] which allowed audio playback at slower speeds and had user-assigned hotkeys which assisted transcription, such as "go back 5 seconds" and "add timestamp".

Kvale (2009, p. 177) describes transcription as an "interpretive process", whereby transcription is the transformation of oral discourse into a written discourse. The researcher should be aware of this comment and able to justify any transcript inclusions or omissions. Particular expressions of irony and humour are often not rendered visible in a written transcription. However, being the interviewer and the transcriber does assist with this, as I was able to recall the tone of the conversation and refer to my notes about particular points in the conversation which may have communicated a stance on an issue, or will have "the social and emotional aspects of the

[19] http://www.ecamm.com/mac/callrecorder/

[20] http://www.nch.com.au/scribe/index.html

interview situation present or reawakened during transcription" (Kvale, 2009, p. 181).

Six interviews were transcribed by an external transcriber recommended by my supervisor. In the case of external transcription I always listened to the audio again whilst following the transcript to ensure it was accurate. My own transcription protocol was based on my prior knowledge of the level of content coding I wished to do afterwards. Although spoken text was not transcribed *verbatim*, only "ums", "ahs", and repetitions and interruptions to conversations (such as external phone calls to the participant, breaks and re-starts to conversation if re-dialling was required due to technical issues with Skype) were not transcribed, and I used shorter sentences for the questions I asked routinely in each interview, for example: "*LP: b&w things about OLNs?"*(in transcript) means *"What have been the best and worst things about using open laboratory notebooks?"*

Long pauses, laughter and sighs which I considered significant in terms of communicating tone were transcribed as [pause], [laughs] or [sigh]. The external transcriber also included these indicators, and they were not instructed to leave out ums, ahs, or repetitions, as I decided it should be my decision as the researcher where and how any adjustments to the record could or should be made.

2.4.2.5 Email interviewing

Although most interviews were conducted via recorded Skype discussions, at an early stage of the research one of the non-native English

speakers requested an email interview instead, as they felt they would be able to give more complete answers in writing than orally. I was happy to comply with this, and the data from this interview was quite similar to the recorded and transcribed interviews. There is a wealth of methodological research literature on email interviewing, often in the context of studies employing both email interviews and other types of online interactions, synchronous (such as instant messaging, where a typed conversation takes place, usually within one session) or asynchronous (such as email or forums, where a lengthy delay may occur between questions and answers).

Recent papers by Kazmer & Xie (2008) and Burns (2010) provide a full overview of the obstacles and benefits of email interviews. Several criteria which the interviewer should be aware of apply when email-interviewing that differ from face-to-face and telephone interviewing, including paying attention to the flow and momentum of the written conversation, invoking previous responses where necessary, following-up quickly, and being alert to conversational tone, and making sure the participant understands the questions (Benford & Standen, 2010).

Many of the concerns raised around email interviewing are based on issues around preservation of privacy, particularly in cases where the research participants are perceived to be either very protective or not protective enough of their privacy, for example autistic adults, teenagers with mental health issues, and HIV-positive adults (Benford & Standen, 2010). These concerns do not apply to the research participants in this

project. Email has been a professional tool for them for most, if not all, of their scientific careers, and they spend a lot of time discussing scientific issues with collaborators via email (Cummings & Kiesler, 2005).

Given that much of the data in this research is textual in origin (see Table 4, Types of Textual Data), and often reflective of the issues under discussion in this thesis, it is reasonable to assume that those interviewed by email are at least as reflective via email as they are in their public postings. Bloggers often write about a wide variety of topics related to their online activities and some of my questions required them to synthesise their views on how blogging, for example, fitted into their personal schedules.

Related topics included queries about the way they incorporate the use of technology into their daily work, how that has changed during the course of their careers, how they got started with blogging, open notebooks and FriendFeed, and various details of their involvement with these specific technologies. (A full list of interview questions is available in Appendix E). Many bloggers discuss their backgrounds and interest in blogging on their blogs in any case; for example, in November 2008 a "blog meme" circulated among some science bloggers, titled "Why do we blog and other important questions?", answered by many on the original *Nature Networks* forum,[21] but also by individual bloggers in their blogs, in over thirty separate posts.[22]

---

[21] http://network.nature.com/groups/nnbloggername/forum/topics/3392

[22] http://blogs.nature.com/mfenner/2008/11/30/why-do-we-blog-and-other-important-questions-answered-by-34-science-bloggers

The approach I adopted for email interviewing was to start with a short introductory email (Appendix B), usually referencing someone they were likely to know who had already participated in the study. It then provided the option for a telephone interview if that was preferred, and included the first couple of "starter questions" of the email interview. A PDF copy of the research consent form (Appendix D) was attached to this email. If the respondent chose a phone interview, I scheduled that with them using email and proceeded as described above in Section 2.4.2, Interviews. In the case of an email interview I waited for their first response and then added follow-up or probe questions where required, also referring to the previously attached consent form and asking for their consent in written form. In the case of slow responses I would have sent prompts up to three times, once per week, on different days of the week, after doing some online research to ascertain that the respondent was not at a conference or on vacation (which is often announced on Twitter). If I had been able to locate a blog post by the respondent on a question topic, I would reference that in the question and ask for clarification or validation of that topic.

I used my oii.ox.ac.uk account for these communications, and included a comprehensive email signature with my contact details at the OII. I used a "threaded" email client (Gmail) to read the email, because it puts all emails with the same subject line in a common "thread", which allows easier reading of email conversation. In tandem with this, I used an individual

Subject line for each invitation so that each conversation stayed in its own thread.

2.4.2.6 Online Documentation and Discussion

Each case study yielded a set of additional data from participants' Web 2.0 interactions. The main sources of data are detailed in Table 4, Types of Textual Data, below. Most useful data came from personal blog posts and FriendFeed posts and subsequent discussions. Other data examined in the course of this research included project web sites, press releases, conference papers and journal articles, photographs, podcasts (video and audio) and some personal communications such as email or instant messenger logs.

Table 4

Types of Online Textual Data

Text	Form	Use in this research
Blog post	Text posting of almost any length	Similar to interview data in that it often outlined a position on a certain topic or gave specific examples of online interactions.
FriendFeed post	Text or multimedia either directly posted to a FriendFeed profile or "fed" in via automatic RSS feeds from a blog, Twitter, or Slideshare account	Discovering topics of most interest in the group; discovering who talked to whom online.
Like	User has clicked a link on a FriendFeed post indicating that they "like" it.	Discovering which topics most interested *The Life Scientists* FriendFeed group.
Comment	Short comment on a blog post or FriendFeed post (limit of up to around 50,000 characters, but an average comment is much shorter)	To indicate interest in a topic; sometimes provided links to more lengthy commentary or relevant article or resource elsewhere.
Tweet	140-character (maximum) text string	To indicate interest in a topic; sometimes provided links to more lengthy commentary or relevant blog post or resource elsewhere. "Live tweeting" at conferences also showed what was happening at a certain place and time.
Retweet / @	A form of Twitter commentary – re-posting someone else's tweet from user's own account, or @<username> - responding to someone else's tweet	Indicative of Twitter "conversations" between users, also showed topics of interest to a particular user.
Wiki entries	Text entries on Open Laboratory Notebook wikis	Examples of online laboratory notebooks - often resembling a paper notebook in terms of description, but also including links to other data such as graphs, videos, or a related "conversation space" such as a FriendFeed post.

To find this data, I searched for the participant on Google and examined the links from their personal blogs and their FriendFeed accounts. Other text data was gathered from RSS (Really Simple Syndication) feeds of blogs and blog aggregations. RSS is an alternative way of reading text on the Web. It is a simple format containing the text and a small amount of metadata such as a title, author and source (often a blog). It allows people to centralise their reading on the Web as they can access many items of interest via one tool (desktop or online) which aggregates posts from many blogs. In effect the data comes to the reader, rather than the reader visiting many different URLs. RSS is a "push" technology, in the sense that the blog post RSS can be updated whenever the author of the text publishes their work. Google Reader (shown in Figure 5) is an example of an online RSS reader.

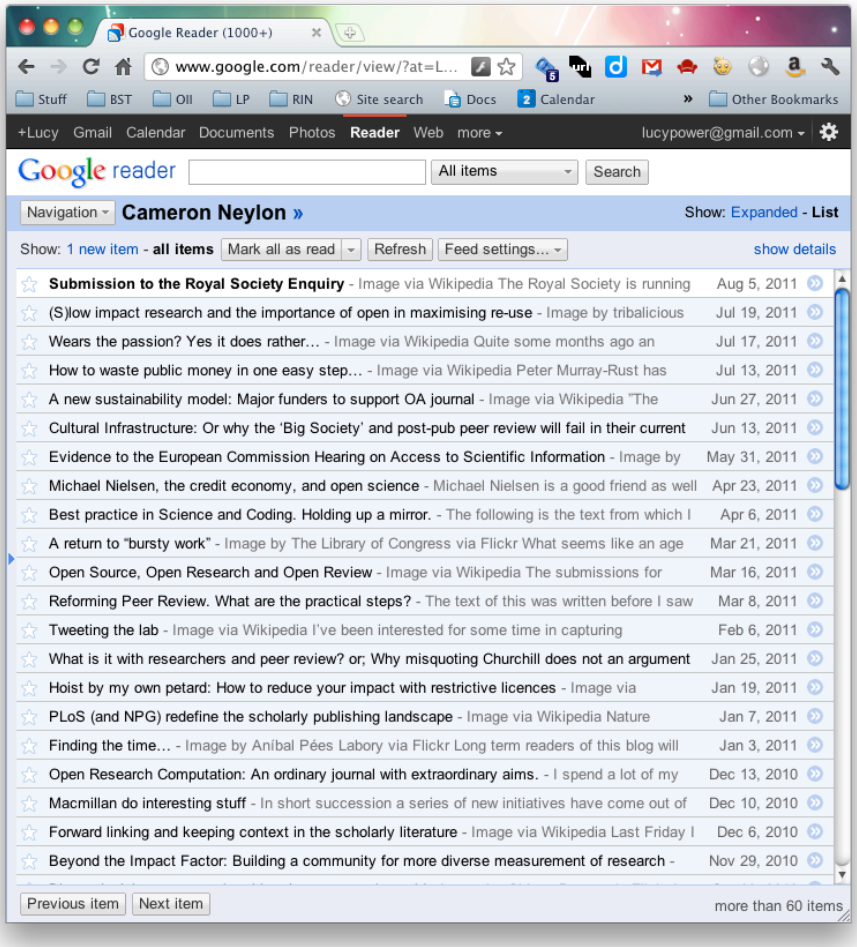


Figure 5. Screenshot of Google Reader Interface. In the live browser interface, clicking on a post heading will show the full text of the post.

Individual blog posts which comment on the issues of interest to this research have also been collected as data. This was done by simply cutting and pasting the text from the blog post or Google Reader interface, and recording the date of the post after the text was imported into TAMS Analyzer.

2.5 Data analysis

2.5.1 Data processing using qualitative analysis software

As Weitzman and Miles (1995, p. 3) state, "computers don't analyze data; people do". However, computer-assisted qualitative analysis software can be a powerful tool for several reasons: making it easier and faster to code data, allowing searches across data for particular codes or strings of text, and allowing re-coding or combining of codes where deemed necessary. This section will explain the tools and procedures used in content coding.

Coding categories were mostly pre-determined by the previously identified thesis research aspects, as well as the cases. This "start list" of codes based on the thesis research aspects was useful in guiding the subsequent interviews, as it helped "tie research questions or conceptual interests directly to the data" (Miles & Huberman, 1984, p. 64). Due to the fact that each research participant was often involved with two or more of the cases, different parts of the interview needed to be coded with both the relevant case and the relevant research aspect. The coding categories gave structure to the interview transcripts, and made it easy to find quotes or conversations about particular issues, and allowed within-case and cross-case analysis of thesis research aspects via search.

CN: Yeah. And by that stage there was...
{theme>communication}{case>OLN}And it sort of built, everything built on itself, I mean it's probably one of the best examples in some senses of the value of that – we made it open but talked about what we were doing, made that open, and people responded, and corresponded, and became part of a community, a wider community, and that community grew through mutual support and had a lot of the same interests which held it [together]{/case>OLN}{/theme>communication}

Figure 6. Example of research aspect and case coding in text

After transcription, the transcripts were imported into TAMS Analyzer[23], a free qualitative analysis tool for the Mac. TAMS Analyzer allows the researcher to add codes to plain or RTF text, in the form of a "markup"—that is, a code at the beginning and end of a piece of text. In TAMS Analyzer a code takes the format *{codename}codeable text{/codename}*. Due to this structure, it was possible to nest different codes within others. For example, if someone was talking about a particular case (for example, blogging) and also clearly referring to one of the thesis research aspects (such as scientific status and rewards), these codes could enclose the paragraph or sentence, and a later search would allow the extraction of all information relating to a particular research aspect across a particular case. See Figure 6 for an example of this coding. Codes relating to cases, research aspects, theory, people and particular interview questions (for example, "routine" for daily work routines) were pre-determined, and other codes such as "secrecy" and "twitter" were added as emergent codes during the coding phase as it was noticed they were frequently talked or written about. A full list of codes is available in Appendix F.

[23] http://tamsys.sourceforge.net/

Each participant's transcript was coded as described above for easy identification of research aspects and key discussions, as were included blog entries, comments and FriendFeed entries. Figure 7, below, shows a typical coding window for an interview. I selected relevant text with the mouse, then double-clicked on the code name in the left hand window to apply it to that text.

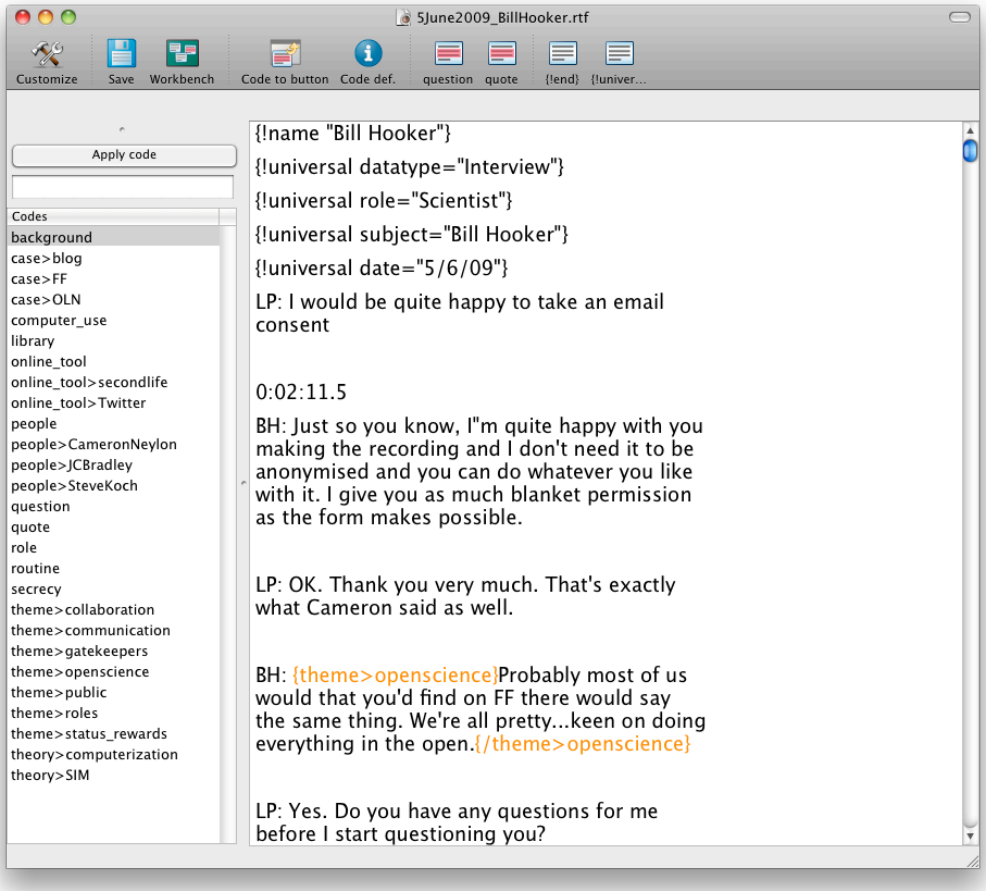


Figure 7. TAMs Analyzer coding window. Coding markup is shown in orange.

Searches (Figure 8) could be conducted across the coded content by single codes or combinations of codes. For example, to find all content coded with the OLN case and the communication research aspect, I typed

"case>OLN+theme>communication" into the search box. Searches could also be saved for future reference.

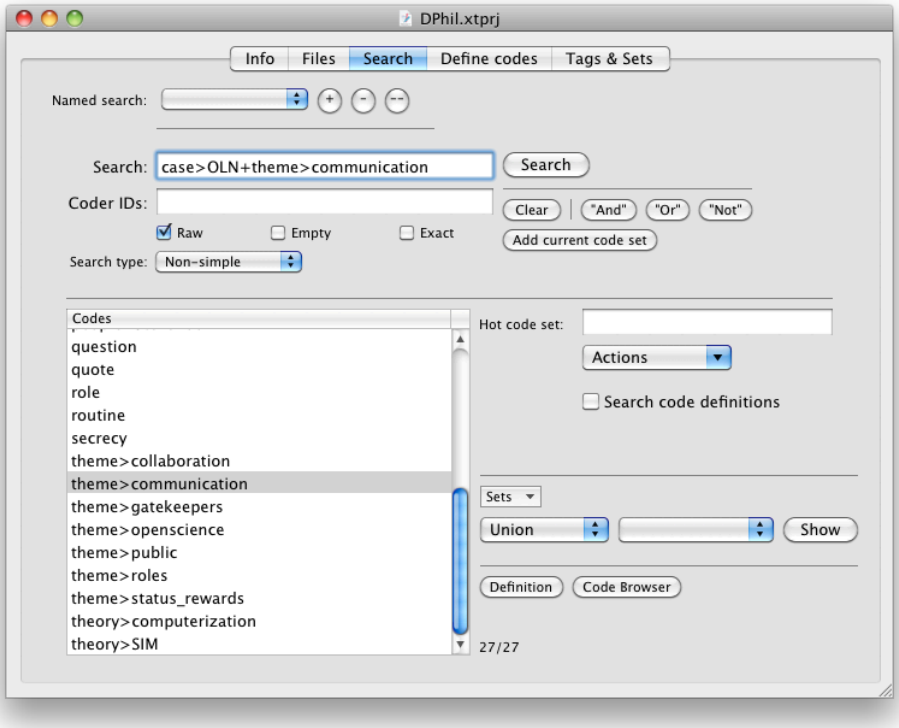


Figure 8. TAMs Analyzer Search Window

Search results (Figure 9) showed the data retrieved in a top horizontal window, and the document reference, with metadata including data types, people's names and roles and the interview date. Including some metadata in each file was useful, as it enabled searches like, for example, "*show me all of the Grad Students in the OLN case who talked about status and rewards*". Double-clicking on the coded data in the top window opened up the coding file for that data (see Figure 7), showing the codes in the context of the interview or blog post or comment.

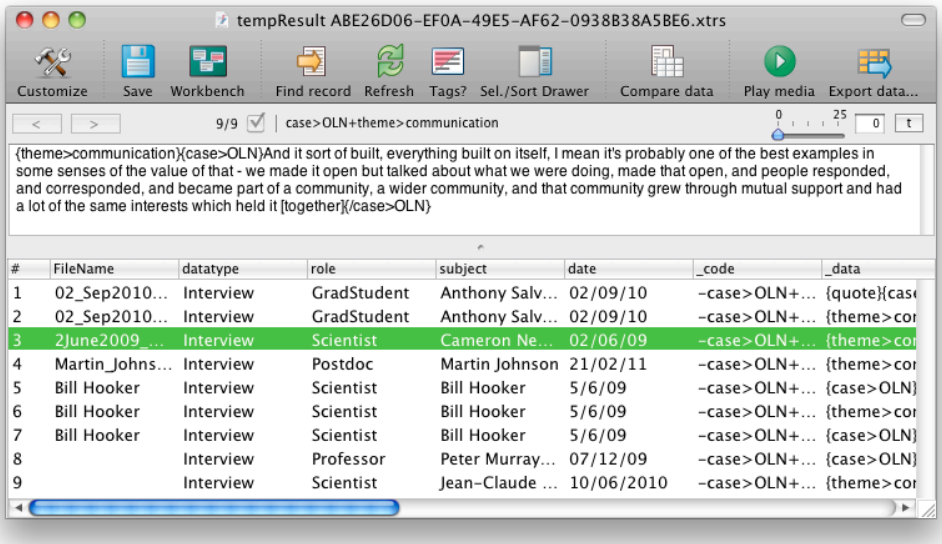


Figure 9. TAMs Analyzer Search Results Window. Search is combining theme and case.

2.5.2 Procedures for handling validity and reliability

Construct validity, which tests the validity of the methodological instruments chosen against the theoretical propositions underlying the research, was ensured by using different data sources, such as semi-structured interviews, email discussions, blog posts and comments, and FriendFeed posts and comments (triangulation of data), making explicit links from the research questions to data and conclusions (aided by the seven research aspects) and having participants review case study chapter drafts. Each case study chapter was emailed to the relevant research participants, including those who were not directly quoted but had added to the overall case background. These participants are listed in the tables within each case. The email asked them to review the chapter, checking for any inaccuracies, and also welcomed further feedback. All involved

responded positively, and the corrections and suggestions included such things as a correction of a wrongly-expanded conference acronym, an incorrect job title, clarification around some of the times at which various science blog networks were founded, and links to further potential references of use (for example: further examples of problem-solving via FriendFeed and several new publications relating to open science work).

Details of strategies for ensuring each criterion of validity and reliability are provided by Yin (2003) and Parkhe (1993). Table 5, below, provides an outline of each strategy, and the Comments column states how each measure was achieved.

Table 5

Ensuring validity in case studies

(adapted from Parkhe, 1993, & Yin, 2003)

Criterion of research quality	Strategy	Research Phase	Comments
Construct validity	Corroborate sources of evidence	Research design	Semi-structured interviews, data from blog posts, comments and other online discussions
	Establish chain of evidence	Data collection	Explicit links from the questions asked to data collected and conclusions drawn
	Have research participants review draft	Composition	Each chapter was e-mailed to the relevant research participants and feedback was welcomed.
Internal validity	Within-case and cross-case analysis	Data analysis	Less important as this is not an explanatory case.
External validity	Purposeful sampling from defined population	Research design	Purposeful sampling methods described within this chapter
	Use replication logic	Research design	Following recommendations from Yin (2003) and other qualitative methods texts referenced in this chapter.
	Comparison of evidence to literature	Data analysis	Each case chapter contains a review of research literature particularly relevant to that case.
Reliability	Develop consistent content coding for all data	Data collection	Content codes developed in advance of coding based on research aspects; some "emergent" codes allowed.
	Use case study protocol	Data collection	Reliance on standard texts for guidance (e.g. Yin, 2003, Patton, 2002).

2.6 Ethical considerations

In online research, ethical issues need to be considered across three planes:

- Data collection and storage - bandwidth/network access and copyright issues (discussed in Section 2.6.2, Data storage and copyright, below)
- How a "participant" is constituted in online research, and the "public" versus "private public" debate (discussed in Section 2.6.1, Participation, information and consent, below)
- How participants react to their own words, read long after they were uttered or written, interpreted in the context of research (discussed in Section 2.6.1, Participation, information and consent, below).

Data gathering techniques for this research included:

- Blog post, comment and FriendFeed textual data sampling (copied and pasted from RSS feeds and imported as text documents to TAMs Analyzer)
- Interviews (transcribed and imported as text documents in TAMs Analyzer)

Data collected during the course of this research was evaluated and stored using the approaches described below, in accordance with the University of Oxford's Central University Research Ethics Committee (CUREC) guidelines.

2.6.1 Participation, Information, and Consent

Regarding informed consent, there is a range of opinions as to whether it is necessary to obtain informed consent for all types of web crawling and research data retrieval. Hookway summarises the three main positions as:

- "Fair game - public domain" - the information is public and there is no need to obtain consent in any situation
- Diaries/Blogging are often written with an expectation of privacy on the part of the author, and should be treated as such
- The situation is unclear and should be assessed on a case-by-case basis (Hookway, 2008, p. 105).

The Association of Internet Researchers (AoIR) suggests in its ethical guidelines that the question to ask in this situation is, "Do participants in this environment assume/believe that their communication is private?" (Ess & Committee, 2002, p. 7). If the answer is unambiguously "No", then the blog authors are participating in a public discussion and would be aware of the accessibility of their posts.

Although I agree with the guidelines stated above, in my own research I applied more stringent rules. For blog posts and FriendFeed comments, I did not seek specific permission from the users (although they are all from participants who had given consent via one of the methods detailed above). Occasionally in conversation during interviews the participant or I would refer to one of these posts as a probe or extension of the discussion topic.

2.6.1.1 Author membership of social networking sites for research purposes

Although most of the scientific FriendFeed users had public profiles, in order to see some profiles and postings, I needed to join FriendFeed and add potential participants to my own friend feed. Users added this way could opt to automatically receive an email, instant message or text message (the site default was an email) informing them of my actions, to which their most common response was to add me back. I similarly joined Twitter (@pownet) and most scientists I followed also followed me back. I also kept a website and profile on the Oxford Internet Institute site which described my research and listed my presentations and publications. However, I consciously chose not to blog, use FriendFeed or Tweet about specifics of my research, with the reasoning that any discussion of my research-in-progress might in turn end up being discussed by the research participants, and influencing the attitudes of future research participants.

Towards the end of my research I posted a few questions on FriendFeed regarding particular issues (for instance, I was unable to find documented instances of scientific "scooping" in the research literature and wondered if anyone else had), and was gratified to receive a few "likes" as well as some useful responses. In retrospect I have reflected upon whether the research would have benefited from my active participation in these communities by precipitating discussion on issues around my research questions, and although I think it would have offered some further insight, I believe that the unprompted online musings and thoughts of the research participants provides rich data in any case.

Yin (2003), suggests that where possible, full disclosure of people and locations in a case study is desirable. Identifiable cases provide two helpful outcomes; firstly, that the reader may recall information about previous knowledge of this case or similar cases, and secondly, that the case may be reviewed and checked more easily, and in this thesis, can be corroborated via use of the Internet. Thirdly, disclosure is easier for the researcher, as they do not have to keep track of various anonymisations across a lengthy thesis. For these reasons, I am grateful that my research participants did not wish to be presented anonymously. All but one of the participants were prepared to use their real names and institutions in the thesis (the exception to this withdrew from the research as discussed above), some with the caveat that the relevant chapter should be shared with them before the thesis was submitted (which was done with all participants whether or not they had expressed this wish; see above regarding a discussion of validity in Section 2.5.2).

It could be argued that making participant names and institutions public is actually furthering the agenda of the open science movement, as this thesis will become searchable on the public internet after it is complete and published on the University of Oxford's institutional research repository (ORA). I would interpret this potential perception differently by saying that this research is designed to report neutrally the outcomes of the participants' activities by the use of a methodological framework. Where in later chapters, findings are positive or negative, this reflects the views of the

research participants with regard to the impact of online tool use on their work. One of the strategies of this SIM is to gather more support for openness, so that their preference for openness in relation to research *about* their research is in keeping with this.

2.6.2 Data storage and Copyright

As an Internet researcher I collect and store not only self-generated data such as interview transcripts, but also data collected from other sources mentioned above, such as blog posts and wiki pages.

The copyright issues for storage of retrieved data are the same as those faced by search engines and the Internet Archive (Thelwall & Stuart, 2006). RSS feeds are supplied on the common understanding that they can be used and stored by third-party software (online and offline). Under the UK's 1988 Copyright Designs and Patent Act, research use of a literary work (such as a blog post) is allowed under the Fair Dealing clause (UK Intellectual Property Office, 2007).

Questions about impact on users (or WWW servers) that can arise regarding online data gathering include how much of the site's resources data collection employs, who is paying for hosting and bandwidth, and whether data collection will affect the availability of a Web service. The answers depend on whether the site's authors are paying for hosting themselves or using a free advertising-supported service such as *Blogspot*, *LiveJournal* or *Wordpress.com*, and also how much data is being retrieved. Data retrieval impact for all web crawling and website viewing was

negligible for this project, as it usually only involved one page-view per retrieval, of less than 100 kilobytes per page (Thelwall & Stuart, 2006). In addition, the text of most blog posts was retrieved from *Google Reader*, which had pre-cached the content, thus even multiple viewings of the text were not drawing upon the original site's server. I am therefore confident that this research did not have significant impact on free or personal data servers.

2.7 Conclusion

Stake (2005) makes the point that "case study is not a *methodological* choice but a choice of what is to be studied", that is, the research topic dictates the choice of case study as a method (Stake 2005, p. 443).

This chapter has detailed the methods used to design, assemble and analyse this research. The next chapter will introduce the first of the three cases, scientific social networking via the FriendFeed site. The chapter on FriendFeed will examine how a nascent Scientific/Intellectual Movement, through the use of online tools, is reshaping some of the ways in which life scientists are collaborating and disseminating their work, making an "invisible college" visible in the process and exemplifying a mix of revolutionising, reinforcing, and unintended consequences in several of the key areas described in the Introduction to this thesis.

Chapter 3
Online Scientific Networking: Making the Invisible College Visible

3.1 Introduction

Use of online tools for scientific communication and collaboration is transforming the way scientific research is conducted. Using the Internet, the life scientists participating in this research are able to reach outside the boundaries of their laboratories to share knowledge, collaborate and publish with colleagues they have met only online. This case study examines the use of FriendFeed by life scientists. FriendFeed, which is described in detail below, is an online social networking and information filtering tool which aggregates online RSS feeds (blogs, Twitter, and other options including Flickr and CiteULike)[24] and affords group interaction via social networking features such as "friending" (adding a contact to your network), "liking" (indicating interest and/or approval of a topic) and commenting on content provided by others.

These uses of FriendFeed exemplify the growth of an online group with the characteristics of so-called "invisible colleges". Invisible colleges, described in more detail below, are defined within this study as a group who share a set of norms promoted and instilled through their online community interactions on FriendFeed, creating community by virtue of the way they communicate with each other, rather than by their physical

[24] At time of writing in July 2011, FriendFeed will automatically import from 58 sites, as well as any user-produced RSS feeds.

locations. Examining the FriendFeed interactions that have occurred will show that this shift to online science communications has had wide-ranging consequences for participating scientists as reflected across the seven research questions of this thesis.

Changes in the roles of scientific researchers are demonstrated via the fact that some of the most active participants in these invisible colleges are former academic scientists who have moved to non-academic jobs but still retain strong interest in the practice and products of science. With regard to changes in scientific collaboration practices, the major change is that scientists have used FriendFeed to actively collaborate on research: sharing methods and data and producing publications across institutions and disciplines.

Traditional gatekeeper institutions are being sidelined to some extent by the production of live conference reports and information-seeking outside institutional boundaries, while scientists are also using FriendFeed to source new funding opportunities and collaborate on funding proposals, thus also contributing to their own career development.

In formal and informal scientific communication, it will be demonstrated that scientists use FriendFeed to discuss issues relating to open data and open science, share and disseminate new sources of data, and crowd-source answers to specific questions about their research and research practices. They also use FriendFeed to discuss approaches to dealing with the media and public, and publicise events and issues important to them.

The unit of analysis of this case study is the use of the FriendFeed site by life scientists, bounded by data gathered from and about several members of a particular group on FriendFeed named "*The Life Scientists*". This group provides a natural boundary to the case because it was created by scientists for the "life scientist types on FriendFeed".[25] Scientists interviewed for this chapter are active contributors to this group, and they conceive of themselves as a group with shared norms and rules.

This chapter continues with a review of relevant literature about invisible colleges and scholarly information management and research on social networking sites, particularly FriendFeed. It will then contextualise and explain use of the site by life scientists. The FriendFeed website is described and interview data is presented to support the statements above, along with summary statistics collated from the FriendFeed site (methods of data collection are detailed in Chapter Two, Research Methods).

This will be followed by examples and discussion of FriendFeed activities and events related to the research questions of this thesis. This chapter concludes that although these online practices do not constitute a revolution in the way science is practised they do signal an important change in the way science is produced and communicated as invisible colleges of science are made visible for the first time, and research practices

---

[25] *The Life Scientists* group description: "A room for all the life science types on FriendFeed (and everyone we've co-opted). Topics tend to focus on bioinformatics and computational biology, but discussion from any area in biological sciences is welcome. Try not to spam and if you are associated with a service (and most don't know that), please say so". Retrieved from: http://friendfeed.com/the-life-scientists.

are opened to scrutiny and study. The communication shift it entails invites us to consider the "common experience of acquiring new means to achieve old ends", as Eisenstein described the impact of the printing press on scholarship (Eisenstein, 2005, p. 288). Analysis of scientific use of FriendFeed will show how these new means can similarly achieve scientific ends.

For this chapter, sixteen interviews with scientists were conducted. During the semi-structured interviews, participants were asked questions about how they found and began using FriendFeed, examples of their and others' use of it, and details of the connections and conversations they have initiated and found via FriendFeed. A full list of the questions is included in Appendix E.

3.2 Relevant literature

To set the stage for understanding FriendFeed, it is important to first understand the previous literature relevant to this case study: the literature on scholarly information management, the concept of invisible colleges, and research on FriendFeed and some relevant aspects of online social networking.

3.2.1 Invisible colleges made visible

I will argue in this case study that with scientific communications moving online, the invisible college has become visible, and for scientists who participate, this newly visible college consists of ties that are largely

internet-mediated and are therefore more networked and distributed than has been previously possible.

The ties that bind scientists into invisible colleges have existed for hundreds of years. Robert Boyle first referred to the Invisible College, as opposed to the visible colleges of Oxford and Cambridge at the time, in a letter to his former tutor written in October 1646 (Boyle, 2001). Boyle's invisible college was a precursor to the Royal Society, founded in 1662 by a group of natural philosophers who corresponded and occasionally met on matters of scientific interest at Gresham's College, London.

The concept of the invisible college is a useful term to describe the informal social interactions of scientists and was popularised in the modern era in a 1961 paper by de Solla Price, (1961), and used by Crane (1972), in her study of the development of new specialities in the literature of sociology and mathematics. Griffith and Mullins (1972) talk about the volunteer nature of science and its effective "loose" communication networks, and the challenges of science that drive these networks to form and consciously work towards certain scientific objectives.

Parker and Hackett (2012) also reference Griffith and Mullins in their description of "coherent groups" as "small, tightly woven groups of researchers oriented toward common intellectual goals who work together in opposition to current intellectual trends" (Parker & Hackett, 2009, p. 22). In other words, invisible college refers here to the way in which non co-

located researchers combine to form a community with particular norms and goals around a common research agenda.

Coherent groups or invisible colleges cannot thrive without informal communication technologies. Lievrouw and Carley (1990) reviewed the formal and informal communication processes of science and examined what they referred to as "telescience", the use of electronic journals, e-mail, fax and teleconferences for collaboration and sharing scientific knowledge. They characterised scientific communications (or the transfer of scientific knowledge) as having three distinct phases: conceptualisation, documentation and popularisation. Conceptualisation is the phase where ideas are developed and discussed. Documentation involves the formalisation of these ideas to share them among those who do not have social contacts in common but who do share a common body of scientific knowledge; this is the phase during which invisible colleges form. Finally, popularisation is the promotion of scientific ideas to the general public. Lievrouw and Carley hypothesised that the potential effects of telecommunications technology on scientific communications would change the way scientists practice science, and blur the differences among the three communication phases as the technology might increase scientists' spheres of influence by enabling greater access to collaborators, facilitate deeper knowledge-sharing with colleagues, and allow scientists to communicate more frequently than their colleagues using conventional communications channels, thereby enhancing their ability to communicate

with other like-minded groups (Lievrouw & Carley, 1990). Depending on the discipline, there could also be a change in the policies or norms controlling the dissemination of information, in the direction of more tightly-controlled releases of information. This effect, as we shall see, is not in evidence for the scientists studied in this research.

Scientific communication is further explored by Zuccala (2006), who uses Lievrouw's distinction between structure and process to examine definitions of invisible colleges and proposes a model for examining them. Zuccala first asks if invisible colleges are considered social processes in themselves, their effects measurable by informal or internal means (via qualitative assessment of communication processes) or part of the structures of scholarship, which could be measured via formal artefacts of science (publication and citations). Zuccala contends that they can have elements of both, in that they may produce formal scientific artefacts such as peer-reviewed journal articles, but communicate formally and informally with each other even if geographically distant.

To aid future analysis of both the formal products and informal social processes of an invisible college, Zuccala defines Information Use Environments (IUEs)—scientific workspaces, physical or virtual. She proposes a model for studying invisible colleges based on examining the subject specialty (disciplinary rules and research problems), social actors (communication processes, formal and informal) and the IUE (physical, human and technological resources). These categories are useful but are

focused on single disciplines, and thus do not account for highly interdisciplinary work such as is being studied in this case.

Caldas (2003) used a longitudinal network study of the connectedness of authors in several academic speech technology newsgroups to examine whether the concept of invisible colleges is supported by online interactions in these groups. Caldas asserts that an invisible college can be said to be present when there is evidence of informal interactions between researchers or the presence of remote collaboration. His data consisted of 28,185 newsgroup messages collected over an eight-year period (1992-2000), analysed quantitatively to assess the general levels of activity and location, and qualitatively via a content analysis of newsgroup discussions and network analysis of the social structure of the groups. Caldas found that inter-sectoral (university and industry) and international collaboration was supported, and that there was also support in the network analysis data for his hypothesis that participants in the newsgroups were forming what he termed "electronic invisible colleges" (Caldas, 2003, p. 55). Caldas concluded by saying that further content analysis of newsgroup messages was needed to determine whether these systems were also supporting the creation of scientific knowledge.

Invisible colleges are being made visible through changes and advances in information and communications technology, and as this chapter will show, are revealing details of interdisciplinary connections and collaborations, as well as the mundane, everyday details of scientific work.

3.2.2 Scholarly information management

Individual management of and response to information flows is not just a topic for the online information age. Blair (2003) documented the anxiety of scholarly authors attempting to keep up with new publications as far back as the sixteenth century. Blair argues that the impression of overabundance led to the popularisation of annuals, encyclopaedias, dictionaries and methods of note-taking and notating, selecting, sorting and storing data. Blair also holds that the early modern (1550-1700) "experience of overabundance of printed material not only fostered the diffusion and development of various aids to learning...but also affected the ways scholars worked" (Blair, 2003, p. 12). This overabundance, Blair argues, led to reading strategies that allowed scholars to consume greater volumes of relevant information, following Francis Bacon's 1612 advice:

> some books are to be read only in parts; others to be read, but not curiously; and some few to be read wholly, and with diligence and attention. (as cited in Blair, 2003, p. 14)

The concerns of the early modern scholars cited by Blair are echoed in the present day by those who participate in online scholarship as they use a variety of techniques and tools to manage the flow of incoming information via online subscriptions and email. This is compatible with the uptake and popularity of FriendFeed among some scientists, because as will be seen below, FriendFeed is essentially a sophisticated social information filtering service, or peer-filtered information-gathering. FriendFeed allows researchers to manage and disseminate relevant information among group

members. They are collectively dealing with information overload by spreading the filtering tasks over their network.

Scholarly information management is a major concern for contemporary life scientists. Not only is "big science" producing tremendous amounts of research data in many life science fields (Hey & Trefethen, 2003a; Borgman, 2007), but the average scientist is also faced with managing an enormous body of peer-reviewed scientific literature that grows increasingly accessible as distribution of information in the digital era becomes more efficient and widespread (Williams & Clark, 1992).[26]

Although Evans (2008) argues that digitisation has narrowed scientific citation patterns, others find that this narrowing trend may also represent greater selectivity by scientists, as there is evidence that scientists are reading a greater selection of the literature available (Tenopir & King, 2008); see further below. King, Tenopir, Choemprayong, & Wu (2009) also found that the more productive an academic was, the more research they claimed to be reading, thus indicating that a high level of information consumption is necessary for success in academia. As will be seen below, large amounts of information can be more easily managed and filtered via FriendFeed, which helps participating scientists stay informed about current issues in their disciplines and open science.

[26] Figures on the volume of scholarly output relative to researchers are difficult to ascertain. Mabe & Amin (2001) state that journal growth rates per annum are approximately 3.3%, and growth can be clearly linked with researcher behaviour and subject discipline trends.

Tenopir and King (2008) found that scientists are reading more overall. Their surveys, conducted between 1977 and 2005, show that "university science faculty on average:

1. read more [papers] ... [but spend] less time per reading
2. rely less on browsing and more on searching,
3. rely more on library provided articles than from other sources [such as personal journal subscriptions]
4. and, because they make choices based on what helps them get their work done, *will readily adapt to new technologies that are convenient to their information-seeking, reading, and work patterns*." (italics mine).

Tenopir and King found that science faculty predominantly search for information rather than browse through pre-categorised literature. If scientists engage with online technologies, they have to manage email and mailing lists as well as keeping up with preprints, commentary and discussion via blogs and online social networks. Peer-filtered information-gathering is one technique that can help manage this deluge.

Electronic information-seeking has also been investigated using organisational theory. Constant, Sproull and Kiesler (1996) investigated the hypothesis that intra-organisationally, the perceived quality of advice given over an internal electronic mail network can be predicted by the resources available to the respondent and their organisational motivation. This hypothesis was informed by Granovetter's (1973) theory of the "strength of

weak ties", which suggests that people close (with "close ties") to us are more likely to know similar things to us, therefore relative strangers sourced via our weak ties will be more likely to have the information we seek. In the context of this research, this suggests that it is important to have networks for close ties but also to have networks of those who are at the fringes of our interests and disciplines.

Constant et al. (1996) analysed requests for information sent over the internal email system of a computer corporation employing 11,000 people over a six-week period, and surveyed the 82 questioners and a selection of their responders. They found that the number of weak ties of an individual does not increase the likelihood of obtaining good advice, but that good advice comes from being able to connect to people with superior resources (access to information or expertise) via the weak ties. At the broadest level, the authors maintain that due to the longevity (>6 years) of the system they studied, an organisational culture that supports information sharing is required to bolster the technological system that enables it, although aspects of the technology (such as publicly viewable questions and answers) also serve to reinforce that culture.

This finding reinforced Granovetter's (1973) "strength of weak ties" theory and demonstrated that the concept of searching for, and finding, information via personal connections also works in an asynchronous digital communication environment, again bolstering the notion that online tools can support the formation of invisible colleges.

It is acknowledged that most online communities and discussion lists harbour varying numbers of "lurkers", people who read content but do not contribute (Nonnecke & Preece, 2000; Preece, Nonnecke, & Andrews, 2004; Bishop, 2007). Nonnecke and Preece (2001) conducted semi-structured interviews with 10 self-acknowledged lurkers in an email discussion list, and identified six broad categories of reasons for lurking:

- Personal (desire for anonymity, shyness, safety from giving offence, lower language skills)
- Relationship to group (still learning about the group, group knowledge is greater than theirs, receive information via others' questions and answers)
- Intention from outset (never had any intent to post, need to post, or not motivated to post)
- Group characteristics (high or low volume of content, poor user interface, style of moderation, response of group to de-lurking)
- Stage of membership (just joined group, about to leave group)
- External constraints (work related or time related) (Nonnecke and Preece, p. 7).

The studies above apply to email lists and discussion forums rather than social networking sites, where the connections between users are specifically revealed, although connections need not be reciprocal in FriendFeed, as detailed below.

Although Meadows (1998, p. 7) defined informal scientific communication as "often ephemeral and made available to a restricted audience only", a definition re-used by Borgman in 2008, the online communications examined in this chapter extend this definition, in the sense that the communications are findable and public, and that those using them accept that they are essentially ephemeral (because there is no formal archiving strategy in place, and no guarantees that the data will remain in place in the long-term), although they do discuss ways of archiving these communications.

The activities described below in this chapter call for a definition of informal scientific communication adapted for the online context which adds the characteristics of an idea being public, findable and traceable from its genesis to its outcomes. The research literature on these areas leads to the expectation that findings will include an obvious blurring of the boundaries between the traditionally closed world of the laboratory and the wider scientific community, that searching for information using network ties is likely to be successful, and that open online communities are likely to have a high proportion of lurkers, who absorb the online information but rarely participate in discussions, but who are informed and are perhaps being educated in ways that will assist them later in their careers.

3.2.3 FriendFeed and online networking

There exists a growing body of research on social network sites (SNS), including research on teenage SNS use (boyd & Ellison, 2007; Lenhart,

Purcell, Smith, & Zickuhr, 2010), levels of SNS adoption (Lenhart et al., 2010) and reasons for use (Brandtzaeg & Heim, 2009). However, most of these studies examine user behaviour on Facebook, MySpace or similar sites, and not SNS with distinct topic-oriented communities like FriendFeed.

Boyd and Ellison (2007) define social network sites as: "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site" (boyd & Ellison, 2007, p. 211). As we shall see, this definition fits FriendFeed well.

SNS use is common among the general population of Internet users. The Pew Internet Project reported that in September 2009, 47% of American adults over age 30 had social network site profiles, mostly on Facebook (73%) and MySpace (48%) and some on LinkedIn (14%), and 73% of the under-30 population have an SNS profile, and 45% of this group access that profile daily (Lenhart et al., 2010). Among academics generally, SNS use is perceived to be largely for personal social activity rather than for scholarly activities. A study released in 2010 by the Center for Studies in Higher Education, UC Berkeley, which conducted semi-structured interviews with 25 biologists at various career stages from graduate student to professor, found that in general, online scientific community networking was not

embraced by faculty or graduate students. For some faculty, this was due to field-specific traditions of secrecy about their work, or fear of their work being stolen. Others, although generally positive about technology, could not see social networking as something that would benefit their science, or cited a lack of interest or time to pursue any interest they had (Harley, Acord, Earl-Novell, Lawrence, & King, 2010).

Graduate students and young scholars were inclined to follow the conventional practices of their fields and scientific mentors. As one interviewee remarked, "It's actually quite easy to find counter examples of young researchers who are very conservative and older researchers who are much more progressive" (Harley et al., 2010, p. 281). Apart from this qualitative study no research on scholarly use of SNS was found in the literature search.

Three studies have examined FriendFeed using network analysis of the entire FriendFeed network with data scraped[27] from the site over several weeks. Gupta, Garg, Mahanti, Carlsson and Arlitt (2009) described the networks and activities of FriendFeed users. They retrieved data concerning the number of FriendFeed users overall and found almost 112,000 distinct users who posted over 10.5 million posts in a five-week period between September and November 2008. Of the more than one hundred thousand users, they examined which services users aggregate data from, who follows the FriendFeed feeds, and why. Gupta et al. found that the activity

[27] Scraping refers to using programs or scripts to access a web page and copy the data on it.

levels of users on FriendFeed appear to be independent of the number of followers a user has and the services they aggregate content from. They observed similar degrees of directionality between nodes in the FriendFeed network—that is, although "friending" (known as "following" on FriendFeed) is not by default bidirectional (as it is on Facebook), a similar proportion of users in FriendFeed had bidirectional linkages, that is, if someone follows you, you are likely to follow them back, although there is no obligation to do so.

A further study by Garg, Gupta, Carlsson and Mahanti (2009) used FriendFeed's application programming interface (API) to scrape FriendFeed's public timeline every five days between 26 February and 6 May 2009, and found the connections of the users discovered for each crawl, which yielded more than two hundred thousand users. This data was combined with similar samples done in September 2008 and January 2009, to provide data for examining age bias (in the sense of length of membership) in social network evolution. This study added the observation that there was a strong group affiliation bias in the evolution of the network—that is, users were drawn towards following other users with similar interests, and joining similar groups.

Celli, Di Lascio, Magnani, Pacelli, & Rossi (2010) collected data from the whole public FriendFeed for two weeks in September 2009. They collected over 10.5 million posts (however they did not distinguish between posts and comments) and half a million "likes". Their study was more detailed

than Gupta et al., in that they subdivided their data to examine "a culturally defined part of the network (the Italian-speaking sub-network)" (Celli et al., 2010, p. 348). They found that for the general FriendFeed population, a small proportion of the users were creating the bulk of the content, that most content did not receive comments, and that users who contributed the most content also received the most comments. They also found a "small but significant" number of users who only contributed comments (Celli et al., 2010, p. 349). Within the Italian-speaking sub-group (identified by computational clustering of the use of Italian language in posts), they described two distinct types of users—"Weak users" (a couple of entries per day which attracted few comments) and "Heavy users" (more than 10 entries per day which received a higher number of comments and likes). They conclude that further qualitative analysis of the entries and conversation topics of the two groups is required to understand why this occurs (Celli et al., 2010).

Thus far, there has been no qualitative research on user behaviour on FriendFeed, exploration of the reasons why people use it, or on usage by researchers or other specialist groups. Prior to discussing user activities on FriendFeed with regard to the research questions of this thesis, some description of FriendFeed and its use is required.

3.3 What is FriendFeed?

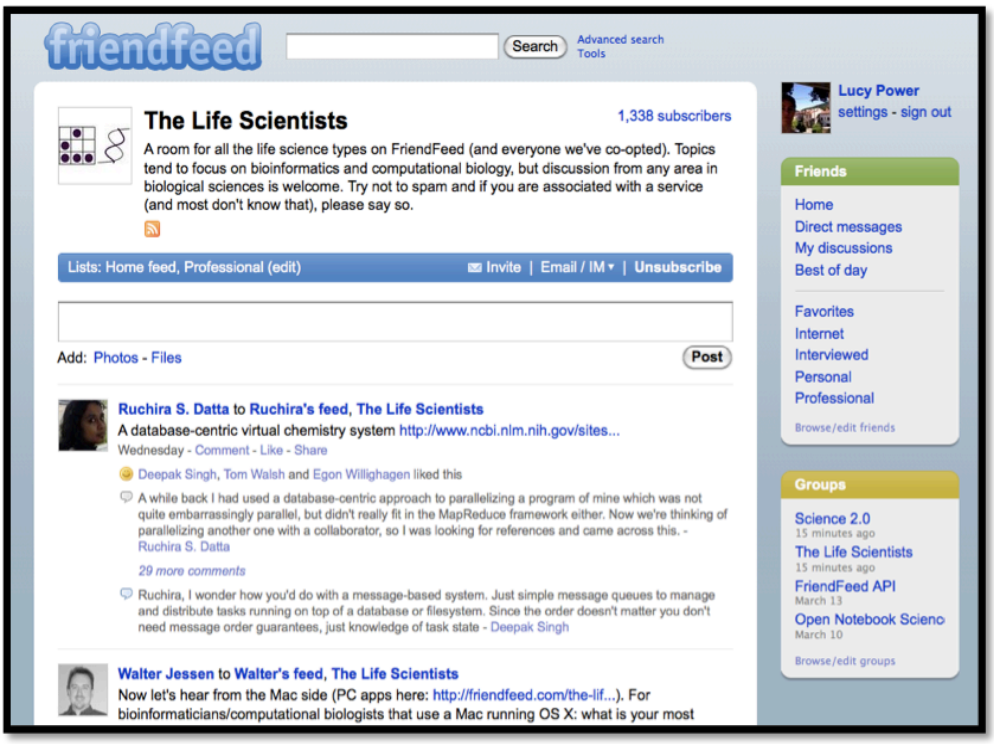


Figure 10. FriendFeed screenshot showing *The Life Scientists* group

FriendFeed is a Web-based RSS feed aggregator, launched in October 2007.[28] It can collate information from any online information sources that output an RSS feed, such as blogs, Twitter, wikis, and other sources such as items from a Google Reader feed. Users add the URL of any RSS feed to their FriendFeed profile and items from that feed will automatically be posted to their FriendFeed and any other groups they specify. These items are posted on FriendFeed chronologically, but will rise back up to the top of a user's FriendFeed if they receive "likes" or comments. There is also a feature called "best of the day/week/month" which will show the user's posts, or a group's posts ranked by numbers of comments and likes. Each

[28] http://www.friendfeed.com/

post is a discrete unit in FriendFeed, whether from an RSS source or created as an original post within the site. Other users of the site can "Like" (click on a "Like" link below the post), comment (click on the "Comment" link) or "Share" (via a provided URL, reposting to the user's FriendFeed or Twitter).



Figure 11. A "like" in FriendFeed. User clicks on Like link on post.



Figure 12. Commenting in FriendFeed. User clicks on Comment, writes comment and clicks Post button.

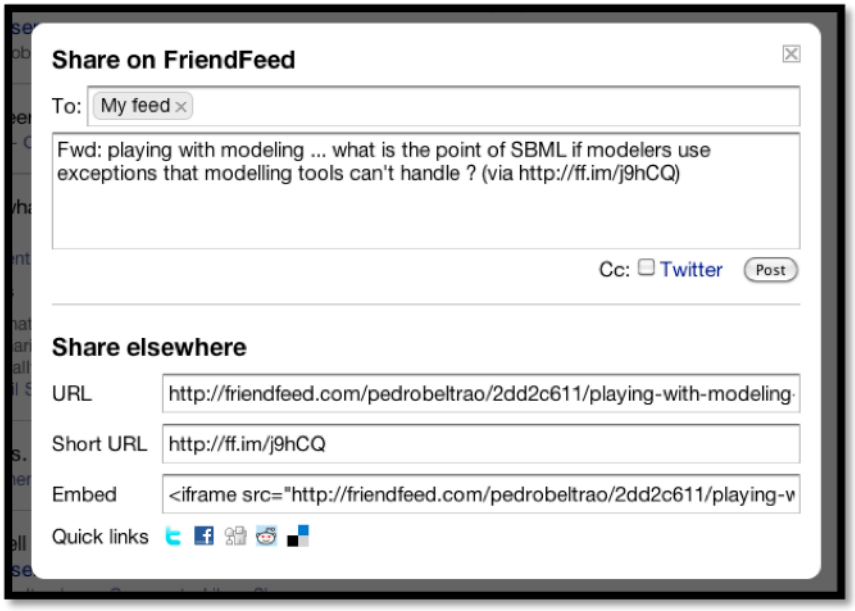


Figure 13. Sharing on FriendFeed. User clicks on Share, and share options are displayed.

By default, FriendFeeds are public. Users can set their FriendFeed to "Private", which will mean that they will need to approve other users who wish to subscribe to (follow) their FriendFeed. Across the whole site, approximately 12% of feeds are set to private (Gupta et al., 2009). None of those interviewed for this research have private feeds.

FriendFeed has a robust search function, which users have remarked upon as one of the best ways to search the "live web" (in particular, Twitter), and the commenting/liking system is fast enough for users to engage in live chat—that is, feedback from "liking" or commenting is near-instantaneous.[29]

Apart from the studies published by Gupta et al. (2009) and Celli et al. (2010), reviewed above, there is no general data on FriendFeed users or user

[29] http://scobleizer.posterous.com/the-second-life-of-friendfeed

behaviour. FriendFeed was acquired by Facebook in August 2009, and the service has subsequently been maintained but not developed, while it was at this time that Facebook added the "Like" button and the ability to comment on status updates to its own site, both of which are features originally developed within FriendFeed.

The next sections of this chapter introduce the FriendFeed research participants and describe their usage of the site as a personal information filter, as well as providing details of the ways they interact and collaborate using the site, and how their interactions constitute an online (in)visible college.

3.4 Research Participants

Sixteen scientists have been interviewed for this FriendFeed case study. Of the scientists interviewed, the mean number of followers (people subscribing to their FriendFeed) is 388, with a range of 29 to 1,334 as of 11 August 2011. In terms of average FriendFeed followers, this is very high, as Celli et al.'s sample found an average of 2.73 followers per user (Celli et al., 2010, p. 348). All include Twitter and one or more self-authored blogs in their personal FriendFeeds, as well as other services including Slideshare and CiteULike. Their commenting and liking activities are more variable, with an average of 4300 comments and a range of 153 - 10,634, and 3288 likes and a range of 10 - 10,955. The sixteen joined FriendFeed at roughly the same time, so these numbers can be validly compared; however there is no baseline user comparison data as the joining dates of other FriendFeed

users is not known, and obviously affects their commenting and liking statistics. Other groups the research participants have joined are also exclusively science-related, which indicates that they use FriendFeed solely for work purposes.

The research participants are all male, and their experience ranges from current doctoral student (2), recent postdoctoral fellow (1), mid-career scientist (8), Professor (2), and former scientist (3). They all hold doctorates in a scientific field, and several of them self-identify as bioinformaticians (4) or did their doctorate in bioinformatics (3). All participants expressed a common interest in open science and open data—it is an often-discussed topic in the FriendFeed Life Scientists group. There is some overlap between the FriendFeed research participants and the science blogger and open laboratory notebook groups, which are discussed in chapters four and five. Most of the FriendFeed participants have been blogging for several years and four of them are involved in open laboratory notebook initiatives in their workplaces.

Table 6

FriendFeed research participant detail (n = 16)

Name	Current Position	Education	Location	Biogang member
Berci Mesko	PhD student in genomics at University of Debrecen, Faculty of Medicine	MD at University of Debrecen, Hungary, 2009	Hungary	no
Bill Hooker	Senior Scientist at Calypte Biomedical	PhD in Molecular Parasitology, University of Queensland, 1997	Seattle, USA	yes
Cameron Neylon	Senior Scientist at Science and Technology	PhD in biochemistry, Australian National	Rutherford Labs, UK	yes

Name	Current Position	Education	Location	Biogang member
	Facilities Council	University, 2000		
Deepak Singh	Business Development Manager - Amazon EC2 at Amazon Web Services	PhD in Chemistry, Syracuse University, 2000	Seattle, USA	yes
Lars Juhl Jensen	Professor of Systems Biology, The Novo Nordisk Foundation Center for Protein Research, University of Copenhagen	PhD in Bioinformatics, Technical University of Denmark, 2002	Copenhagen, Denmark	yes
Neil Saunders	Statistical Bioinformatician, CSIRO Mathematics, Information and Statistics	DPhil Biochemistry, University of Oxford, 1996	Sydney, Australia	yes
Paweł Szczęsny	Asst. Professor of Bioinformatics/Genomics, U Warsaw	PhD in Biochemistry/Biophysics, Institute for Biochemistry and Biophysics, U Warsaw, date	Warsaw, Poland	yes
Pedro Beltrao	Postdoctorate in molecular biology, USCF	PhD in Bioinformatics, European Molecular Biology Laboratory, 2007	San Francisco, USA	yes
Richard Grant	Information Architect, Faculty of 1000	DPhil in Molecular Cell Biology, University of Oxford, 1994	London, UK	no
Pierre Lindenbaum	Bioinformatician at INSERM U915	PhD in virology and bioinformatics, Université Paris Sud (Paris XI)	Nantes, France	yes
Jean-Claude Bradley	Associate Professor of Chemistry, Drexel University	PhD in Organic Chemistry, University of Ottowa	Philadelphia, USA	no
Mat Todd	Lecturer in Organic Chemistry, University of Sydney	PhD in Chemistry, University of Cambridge	Sydney, Australia	no
Steve Koch	Assistant Professor of Biophysics and head of Koch Laboratory, University of New Mexico	PhD in Physics, Cornell University, 2003	Albuquerque, USA	no
Andrew Maloney	Doctoral student, Koch Laboratory	MS in Physics, UNM, 2008	Albuquerque, USA	no
Andrew Lang	Professor of Mathematics, Oral Roberts University	PhD in Mathematics, University of Missouri-Columbia	Oklahoma City, USA	no
Bora Zivkovic	Blog Editor, Scientific American	MS Zoology, NCSU, 2005	Chapel Hill, USA	no

Table 7

FriendFeed research participants (Data snapshot 18 July 2011).

Name	Subscriptions*	Subscribers	Comments**	Likes**	Groups***
Berci Mesko	391	792	466	1,101	23/1
Bill Hooker	97	252	7,769	2,862	30/0
Cameron Neylon	160	602	10,140	5,700	36/0
Deepak Singh	386	841	10,372	10,955	49/0
Lars Juhl Jensen	73	198	1,264	546	1/0
Neil Saunders[30]	252	350	6879	5282	41/0
Paweł Szczęsny	393	396	1,986	3,260	53/0
Pedro Beltrao	138	234	1,914	1,310	15/0
Richard Grant	71	100	505	71	11/0
Jean-Claude Bradley	303	366	4,213	4,640	39/2
Pierre Lindenbaum	163	373	2,989	5,192	36/0
Mat Todd	73	96	2,411	802	20/0
Steve Koch	159	164	10,634	2,606	34/4
Andrew Maloney	23	29	153	28	6/0
Andrew Lang	23	87	1,029	1,217	5/0
Bora Zivkovic	1,120	1,334	6,079	7,048	55/1

Key:

*Other FriendFeed users or groups they subscribe to

**Comments and likes made by these authors

***Number of groups they are a member of: Science or Academia related/Non-science related

[30] Saunders left FriendFeed on January 31, 2011, announcing it via a blog post (Saunders, 2011). Saunders found that over 90% of the items in his FriendFeed came from Twitter, and decided a better strategy was to publish via various services which supported archiving (e.g. his blog), and broadcast that fact via Twitter. The statistics above for Saunders only are from 22 March 2010.

3.5 Research participants' Normal Use of FriendFeed

Research participants use FriendFeed in several ways, either as their primary information filter and reader, or as a secondary backup to their primary sources, such as Google Reader. Their information management practices with relation to FriendFeed are to some extent governed by technology interface preferences they formed before starting to use FriendFeed. Everything fed into FriendFeed can in turn be an RSS feed itself (for example, items a user has liked or commented on can be separate feeds, and each group has its own feed). Some FriendFeed users will check the site remotely through standalone feed reader software. Based on the current sample, seven users use the FriendFeed interface as a primary information filter on a regular basis, and two prefer to use Google Reader to access FriendFeed.

"Primary readers" make FriendFeed a part of their workflow, and have changed their work practices by integrating most of their online activities into the FriendFeed interface, as discussed below. "Secondary readers" use FriendFeed as a backup or repository of all of the news items they are interested in, although they will still come to the site occasionally to read comments or participate in discussions.

It is common for FriendFeed users to have a large number of new items in their feed in the mornings, no matter what time zone they are operating in, since those active on FriendFeed are scattered around the globe. Every morning they will be greeted with fresh comments or likes on their own

FriendFeed postings from the day before, and a number of new items to be scanned or read.

> I guess this is one of the examples where these web tools come in handy. I guess the obvious is going through the eTocs, table of contents, but mostly it's also recommendations from other people in several websites, like Google Reader and FriendFeed for example. [In the] morning I go to a bunch of different websites. Well, I need my coffee before I start working. (Beltrao interview)

A typical daily workflow would include getting coffee, getting online (either at home, during their commute or first thing on arriving at work), and checking their centralised information sources (FriendFeed and/or Google Reader). As they scan through up to 300 fresh items, they will triage—some items will be "liked" or commented on, some opened in new tabs or bookmarked for later reading, and the majority not looked at again (that is, no action is taken post-scan). Then they check personal email and arrive at or begin work.

> I try and do it at various times through the day, rather than just letting it come to me all the time. So I'll do quite a bit first thing in the morning, when I get coffee, and again sort of lunchtime, and probably late afternoon. (Saunders interview)

Management of online information flows during the working day varies—some people will close all alerts and browsers while they concentrate on a particular work task, and allocate lunchtime or pre- or post-work time for feed-reading, responding or blogging. Others will keep up with information streams throughout the day, microblogging via Twitter and commenting on blogs and FriendFeed items.

> [On my commute] I can generally get through all of the reading of blog posts, online stuff, sort of formalised online stuff, if that makes sense, before I arrive here [at work]. I do a quick run over some of what's happened on FriendFeed overnight, and maybe a quick check on Twitter. So then I would arrive at the office a bit before 9 o'clock. Then the first thing I then do is actually try ... [to] get through all of the things that have been thrown up by all the stuff previously read that need ... an interface that isn't a mobile phone. So that will often be looking in a little more depth at blog posts, it will often be chasing down a few trails of information online, finding any other bits of information that need to go into emails, that kind of thing. (Neylon interview)

From these examples, it can be seen that FriendFeed and related online tools have been integrated into daily work practices. Part of the information filtering that takes place is provided by FriendFeed groups, which are described below.

3.6 The Life Scientists Group

FriendFeed contains specific topic-limited areas—in an earlier design iteration called "rooms"[31]—for example *The Life Scientists*, *Science 2.0* and *Open Science*, as well as groups which aggregate information on particular scientific conferences (*ISMB 2008*)[32] or attempt to source particular references (*References Wanted*). This chapter examines the activities conducted within a FriendFeed group named *The Life Scientists*, created in June 2008, whose August 2011 membership was 1,518. FriendFeed groups allow people with similar interests to pool their online findings, and discuss informally (and publicly) issues of relevance to them. Groups are used

[31] A simple way to distinguish early adopters of FriendFeed from later is whether they refer to groups as "rooms" or "groups". The design change happened in April 2009.

[32] ISMB is the International Conference on Intelligent Systems for Molecular Biology.

primarily for extending and marshalling discussion on topics of interest to the group, and *The Life Scientists* is updated many times a day by both its external feeds (for example, blogs, Twitter, CiteULike) and internal comments (links and discussion). The frequency of interaction among participants is high, and it has been adopted as a tool for dissemination and discussion by groups of life scientists.

From its inception in June 2008 through to March 2010 (655 days), a mean of 5.02 entries were posted per day, with a mean of 3.85 comments and 3.1 likes per post. Over this period the group's membership grew to 1334 members. To give some idea of average group activities, the 3290 posts were written by 254 authors, and commenters numbered 154 (making 854 comments in total), and Likes were posted by 202 group members. Fewer than a quarter of the current group members have posted, and even fewer have commented. This indicates that most members of *The Life Scientists* are lurkers rather than contributors, or, as characterised by Neylon, listeners: there are a "lot of people who are just listening" (Neylon interview).

However, since the group membership has grown very rapidly over the past year, it is possible that many lurkers fit into Nonnecke and Preeces's (2001) "early stage" membership of the group and are still working out what the group conventions are, since participants acknowledge that they exist:

> There's definitely a community. And it's definitely got a character and set of unspoken rules. (Neylon interview)

Some of these "unspoken" rules are in fact clearly enunciated—as mentioned above, the description of *The Life Scientists* says: "Try not to spam and if you are associated with a service (and most don't know that), please say so" (see Footnote 25). This sense of community experienced by the FriendFeed users supports the idea that there is an active invisible college centred on FriendFeed's use for online scientific networking.

Topics discussed in *The Life Scientists* vary widely, but rarely lose their scientific focus. Topics that are frequently observed include: open science and open data, the processes of academia (such as grant finding and writing), specific wet laboratory (laboratories handling biological matter, as opposed to *in silico* (computer modelling)) or bioinformatics questions or discussion of particular software or laboratory kits, and the sharing of items considered to be of general interest to the group, such as protein visualisations or general bioscience issues being reported in the press. *The Life Scientists* acts as an information-filter, a starting point for finding items of interest to the community.

3.7 Network expansion online - from Biogang to FriendFeed

Use of FriendFeed by over 1300 life scientists was initiated by a particular group of scientists who were the first adopters and founders of *The Life Scientists* group. They remain the most active users of FriendFeed, and the origins of their group and their move to FriendFeed are relevant to this discussion, as they demonstrate a shift in their group's boundaries, as they moved from a small, relatively closed group (writing public blogs but

not actively networked) to FriendFeed, an open, searchable and networked tool.

Before using FriendFeed, the life scientists involved followed individual blogs, and although conversation did (and does) arise on blogs via the comment mechanism, the blog technology was not networked, thus readers followed individuals rather than viewing a linked information network.

Deepak Singh created the FriendFeed *The Life Scientists* group. The first to populate this group were a set of scientists and ex-scientists named the Biogang. The Biogang is a group of life scientists who promote and practice the idea of collaborative problem solving in science using online communication tools and data sharing.[33] This section shows the origins and activities of the Biogang, and details its involvement with FriendFeed as an information distribution and problem-solving network. The Biogang identified several key projects which they wanted to work on and problems they wished to discuss, and they have been able to realise several of them using FriendFeed and the wider network that has grown up around *The Life Scientists* group.

Prompted by reflections by Singh on his blog, in a post named "*A global informatics collaboratory*", on 30 May 2008, and followed by a wikispaces (a free wiki provider) posting the next day[34], later transferred to

[33] http://openwetware.org/wiki/Biogang (Accessed 18 March 2010)

[34] http://biogang.wikispaces.com/

OpenWetWare, the Biogang was formed as a way to get science done, self-described in the wiki as:

> An informal, distributed collection of geeky life scientists who have come together to try and think of cool problems and ideas that can be solved collaboratively. While many of us are informaticians and modelers, the Gang welcomes all comers. We have a variety of backgrounds and experiences but all share a love of science, openness and belief in the power of the web to enable distributed collaboration and community. Neil Saunders summed our approach and attitude up very well: "*A biogang characteristic: lack of respect for institutional boundaries and restrictions?"*. (Singh, 2008)

The quotation above illustrates that the Biogang have internalised a key aspect of what Frickel and Gross (2005) consider to be essential to the formation of a Scientific Intellectual Movement (SIM), namely that they pose a significant challenge to "dominant ways of approaching some problem or issue" and involve "dramatic breaks with past practices" (Frickel & Gross, 2005, p. 207). This is a typical statement made by participants and many members of *The Life Scientists*—they see the Web as a medium enabling them to challenge closed ways of doing science and to make it more open.

Furthermore, there is also a sense of the community being enabled by its relationship to the technologies they use. Saunders commented on Singh's original blog post:

> Like you, I think this informal "biogeek" network that has formed is a wonderful thing. I get more pleasure and intellectual stimulation from my daily web interactions than I have from any of the universities at which I've ever worked. I guess the issue is: can we, as a community, generate something that others perceive as having value? Which means: tools or

> better still, discoveries/original research findings. (Saunders, 2009)

This quotation shows that the "Biogang biogeeks" fall between two strands of thought. On one hand they view themselves as "geeks" in the sense that they are heavily involved with technology for communication and as a tool for performing their work, but on the other hand they are also aware that their use of technology services a broader intellectual goal of generating useful scientific communication and creating a new and more productive scientific research environment.

This vision of a new scientific research environment is explained by Singh in an interview:

> The Biogang, in my mind, was this bunch of bioinformatics and computational biology geeks. And even non-comp-biology people who can get around and solve problems to help each other. .. a community of computer-savvy life scientists who can essentially collaborate—what I call "bursty work"—it's a short project. ... And I think that's actually how science should be done in general, even in a more formal context rather than have one big company trying to solve one big problem, you have a lot of little people all over the world, who can come together. (Singh interview)

The concept of "bursty work"[35] appropriated by Singh and the Biogang is similar to Parker and Hackett's "hot spots and hot moments"—a short, productive scientific interaction characterised as "those places and times of high emotion, creativity, and performance where transformative science is done" (Parker & Hackett, In press, p. 33). Parker and Hackett use

[35] Singh acknowledges borrowing the phrase "bursty work" from a tweet by Chris Messina (designer and open web advocate) on 7 Dec 2007, "It just dawned on me that startups are obsolete. Sustainable, distributed bursty work is the future". Retrieved from: http://twitter.com/#!/factoryjoe/statuses/479283692

"affectivity" in the sense of emotional group connections, which is further evidenced by Saunders' reflection below:

> I think in all of the places where I network there's always been this core of about 50 people I guess. A few are everywhere. We have a bit of a joke about it sometimes, actually. It's the usual suspects on all the social networks. (Saunders interview)

A list of Biogang members is available on OpenWetWare (19 members)[36] and a LinkedIn group (12 members).[37] Biogang members interviewed for this case study include Cameron Neylon, Deepak Singh, Neil Saunders, Pawel Szczęsny, Pedro Beltrao and Bill Hooker. They come from and work in diverse locations and roles (see Table 6, above). An informal survey on FriendFeed initiated by one of this chapter's research participants received 30 responses and shows that the group is a mix of wet and *in silico* laboratory scientists—15 bioinformaticians, 8 wet laboratory, 4 mixed and 3 other (non- or ex-scientists).[38]

The Biogang members got to know each other through their personal blogs. Most of them have been blogging about their scientific work for between three and seven years.

> We're sort of the old men of the scientific blogosphere, especially the bioinformatics blogosphere ... So that's how we all got to know each other. Then we all ended up in the same social networks, and when social networks took off, other people came. (Singh interview)

---

[36] http://openwetware.org/wiki/Category:Biogang_members. Accessed 18 March 2010.

[37] http://www.linkedin.com/groups?gid=1300137. Accessed 18 March 2010.

[38] http://friendfeed.com/the-life-scientists/c0dafafb/what-do-you-work-on-general-field-and-wet-dry-lab

The Biogang members form the core of the FriendFeed network, although many others also participate (there are 19 Biogang members and 254 authors in total in *The Life Scientists* group in March 2010).

The Biogang maintain a list of ongoing projects (currently fifteen) including ideas for scientific games, FOAF (Friend of a Friend) semantic web networks, various software ideas and a Bioinformatics Career Survey. Most of these are reposted ideas that originated in FriendFeed, but some are more extensive notes, sketches and nascent project plans. The most fully realised project is the Bioinformatics Career Survey, which was created in 2008 by Michael Barton, and promoted via the rest of the Biogang on FriendFeed and via blogs and email. It received 681 responses over the month of June 2008, and the data is freely available, with some analysis also being hosted on the OpenWetWare wiki.

The Biogang wiki provides evidence that Biogang activities have moved almost entirely to FriendFeed. Most of the Project pages have not been updated since mid-2008, and the main page has not been updated since May 2009.[39] The discussion that the Biogang used to have internally and on each other's blogs now takes place in the more open forum of FriendFeed, although to some extent it also still continues on the personal blogs where

[39] The Biogang wiki on Open Wetware was last updated in May 2009 (minor edits), and prior to that in November 2008 (adding LinkedIn).
http://openwetware.org/index.php?title=Biogang&action=history

many FriendFeed posts originate. There is further evidence on FriendFeed that "Biogang" has become a more general term of reference for "members of *The Life Scientists* FriendFeed group", as shown by posts that say "Welcome to the Biogang" to first-time posters/de-lurkers in that group.

Thus we can see that the Biogang members have a set of shared norms, and aim to challenge a number of ways of doing science, such as formal collaborations and limited interdisciplinarity, simply by using online tools such as FriendFeed to facilitate their "bursty work".

3.8 Further Detail and Discussion

The examples cited above show how FriendFeed is being used by members of *The Life Scientists* group, and that the group's activities originated from the Biogang, who found FriendFeed an effective medium to expand their network and promote their ideas of how science should be carried out. I have shown that there has been a shift in informal scientific communication for its users, and that online production of scientific work is possible using FriendFeed and complementary Web tools. The use of these tools has expanded the personal scientific networks of those using it—not only do they still have their traditional networks of co-workers and colleagues who meet locally or at conferences, but they have added online colleagues with whom they have daily conversation. Further discussion below expands on this by addressing the research questions of the thesis with specific reference to exemplary activities on FriendFeed.

3.8.1 Scientific Roles

RQ: Do online tools extend the professional roles of academic life scientists beyond their traditional ones?

Although the use of FriendFeed does not change an individual scientist's role, it does afford them new or alternative opportunities for information retrieval and cross-disciplinary conversation. *The Life Scientists* constitutes an informal network not defined by particular institutional or disciplinary boundaries, so the roles of the participants are not directly referenced or implicated in the types of interactions they have. For example, Deepak Singh, the founder of and key participant in *The Life Scientists*, is a former scientist and bioinformatician, who works in a quasi-related area (Amazon Web services) and still participates, supports and encourages academic scientists using FriendFeed. FriendFeed to some extent enables a flattening of hierarchy, in that most participants do not actively declare their status or job (although it can often be found in the "About" section of their blog or FriendFeed profile). Despite, or perhaps because, they do not overtly declare their professional interests, they are free to participate in a wide range of conversations and projects which do not necessarily relate to their narrower professional expertise.

The "visible college" of FriendFeed is largely confined to online activities. Within their own workplaces participants appear not to discuss their activities on FriendFeed, or do not feel it is worthwhile to encourage their workmates to join. Some participants mentioned their colleagues'

indifference to their online activities, for example Saunders: "[my] workplace was pretty much unaware of it [FriendFeed], and if they were aware of it, were not particularly interested in it" (Saunders interview). However, participants are happy to post work-related questions to the FriendFeed group, and are not concerned by their colleagues' lack of interest as they greatly value their online relationships, as was shown above in Section 3.7 when Saunders described the pleasure and intellectual stimulation he gained from daily interactions with his online colleagues.

Interactions within FriendFeed do not require the participants to change or extend their roles in terms of computing expertise. Most research participants commented that they found FriendFeed very easy to use. Discussions within FriendFeed's *The Life Scientists* group are likely to remain topic-specific. For example, a discussion about how to use a particular command to sort or filter genomic data will usually only receive comments from bioinformaticians, whereas a discussion about open data or open science often has input from science librarians as well as life scientists in many different specialities.

> I haven't physically been in a library building for years, so I hadn't given much thought to libraries as anything other than providing a website where I go to download my PDFs... But then you sort of think well, hang on, someone manages these websites, and does all of this kind of web librarian stuff, and of course that's what librarians do now, you know, they're web architects, basically. And we discovered all these people, who are really very interesting and have a lot of interesting things to say. (Saunders interview)

This quotation shows that there is a change in roles: as well as acting as each other's online librarians by using FriendFeed to filter information, they are interacting with the information science community, who by their own admission they have not had contact with for years.

The core participants in FriendFeed scientific conversation are largely members of the original Biogang, who met online via their blogs. This group consisted of people who were looking for an online tool which would enable them to share information and interact around it much as they would offline in the workplace with their local colleagues. FriendFeed enables this interaction on a global scale with online colleagues.

> LP: have you made connections to many people via FriendFeed?
>
> CN: Hundreds. Connections with the people who run the Science 21 conference in Waterloo last year wouldn't have happened without FriendFeed. Connections with ... the US information management / librarianship ... community wouldn't have happened without FriendFeed. (Neylon interview)

While individual roles may not be changing due to the use of FriendFeed, via its use participants note that their network of connections is expanding and interdisciplinary conversation is increasing, thus signalling some reconfiguration of their networks (Dutton, 1999, 2005). This reconfiguration affects group relationships and increases the chances for "bursty work" to occur. Examples of "bursty work" include the collaborations, both informal and formal, described below.

3.8.2 Online Collaboration

RQ: Do online tools facilitate inter-institutional and interdisciplinary collaboration for academic life scientists, and if so, how?

The interactions and conversations stimulated via FriendFeed have given rise to several examples of new collaborative work, much of it conducted almost entirely online. Some examples of online collaboration initiated on FriendFeed are given below. These examples also highlight the convenience of networking via FriendFeed, which makes it easy for those seeking collaborators to connect with others who have the specific skills they are looking for.

On October 2, 2008, a member of *The Life Scientists*, Andrew Su, posted a short request asking if anyone in the group was interested in helping out with some programming "in exchange for authorship on a future publication".



Figure 14. FriendFeed query asking for programming skills.

Less than a year later, on September 15 2009, an article was published in the Nucleic Acids Research database issue. As Su posted on his blog at the time:

> I'd also be remiss if I didn't also note the critical role online collaboration played in this effort. Of the seven coauthors on this paper, two I've met only once in real life, and two I've never met in person. We are spread over four cities, five organisations, and nine time zones. Initiating and executing

> this collaboration happened virtually entirely online, aided by the FriendFeed Life Scientists room and Molecular and Cellular Biology WikiProject at Wikipedia. It was an eye-opener in terms of how effective online collaboration can be done. (Su, 2009)

There are several papers that have been co-authored by people who have met only via FriendFeed (of which more in the section below on Gatekeepers), and others where crowdsourcing has played a part in data or idea-checking.

An idea that did not lead to publication but was heavily discussed and drafted via FriendFeed conversation and wikis was Cameron Neylon's suggestion that the group co-author a paper on the use of FriendFeed for a special issue of the Journal of Science Communication on peer-to-peer and user-led science (http://jcom.sissa.it/archive/09/01). Cameron Neylon posted a FriendFeed entry with a link to the journal's call for papers (see Figure 15 below), and a long FriendFeed discussion ensued about it, with 99 comments from 11 authors. They worked together to produce a proposal, which was ultimately not accepted, but resulted in a project proposal officially authored by six people from different institutions and disciplines, titled "A Friend in Feed is a Co-author Indeed: The social aggregator as a tool for user-led collaborative science" (Brembs et al., 2009).

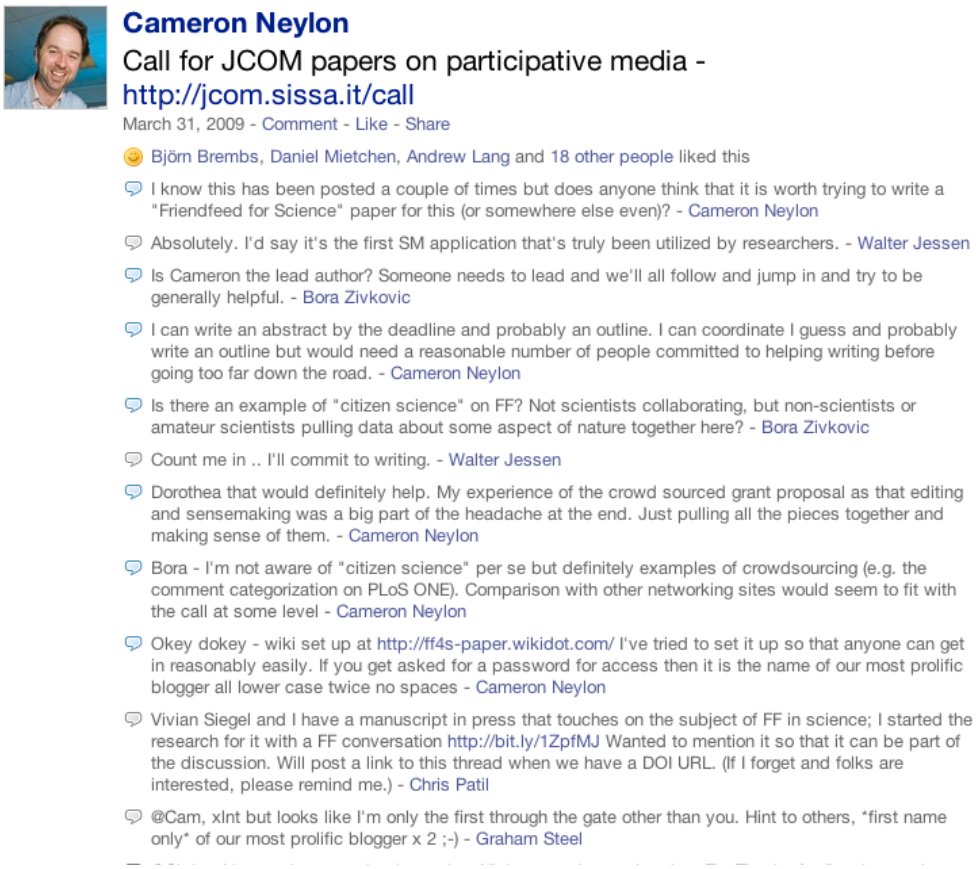


Figure 15. Discussion on FriendFeed. (23 likes, 99 comments) (Neylon, 2009).

Thus, there is evidence that members of *The Life Scientists* are finding and participating in collaborative work sourced solely online. However, at this stage it is unclear whether this is initiating more collaboration than it would in the normal course of their work, since it has been shown that scientific collaboration and co-authorship is increasing in any case, as Wuchty, Jones, and Uzzi showed a steady increase in author numbers across many disciplines in their analysis of almost twenty million journal articles (Wuchty, Jones, & Uzzi, 2007).

3.8.3 Science Communication

RQ: How do academic life scientists use online tools (blogs, wikis, open laboratory notebooks) to publish different types of scientific data?

FriendFeed's Like and Comment features enable a form of communication that has several interesting features: flexibility, openness and near to real-time interaction (a page will automatically update as new comments are added). The style of communication is informal. Research participants say that they like FriendFeed because of the "low barriers to entry" (click to "like"). As Saunders remarked:

> We realised the great thing about FriendFeed was the conversational aspect ... It seems to fill some niche that people were looking for in terms of scientists chatting amongst themselves, but that chat being centred around specific items, rather than just chat. (Saunders interview)

In a blog post about FriendFeed for scientists, Cameron Neylon explained the usefulness and speed of peer-filtered information-gathering:

> It also helps people to find you. As well as this items with comments or likes then get popped up to the top of the feed so items that are generating a conversation keep coming back to your attention.
>
> These conversations can happen very fast. Some conversations balloon within minutes, most take place at a more sedate pace over a couple of hours or days but it is important to be aware that many people are live most of the time. (Neylon, 2008)

Neylon highlights a particular FriendFeed discussion where he asked a question about how to produce a protein model:

> I was looking for some help in building a particular protein model for a proposal. I didn't really know how to go about this

> and didn't have the appropriate software to hand. Pawel Szczęsny offered to help and was able to quickly come up with what I wanted. In the future we hope to generate data which Pawel may be able to help us analyse. ... We are still a long way from the dream of just putting out a request and getting an answer but it is worth point[ing] out that the whole exchange here lasted about four hours. (Neylon, 2008)

Thus, although the communication process is informal, it is assisting with the production of formal scientific outcomes such as grant proposals. Informal discussion on FriendFeed also includes discussion about data—where to find it or how to manage it. Research participants in this case study generally download data for manipulation on their own computers rather than performing computational analysis online, although their informal discussion also includes chats about data, particularly regarding data standardisation, which is of concern to many in the group. This relates to their group aims of facilitating scientific "bursty work", in that data interoperability is key to allowing scientists to work on short projects, as now they can spend a lot of time normalising data before it is ready for analysis.

3.8.4 Gatekeepers

RQ: Does the use of online tools by academic life scientists circumvent gatekeepers in science and/or create new ones?

Informal scientific communication is not an area where the actions of gatekeepers such as institutions, publishers, editors and grant-awarding bodies have particularly influenced scientific behaviour. However, there is some evidence that the scientists using FriendFeed have begun to blur the

boundaries between the formal deliverables of science and informal communications by using online tools to communicate directly with their peers.

With regard to the organisation of conference outputs, normally the remit of the conference organisers and an associated journal publisher, Saunders, Beltrao, Jensen, Jurczak, Krause, Kuhn and Wu (2009) used FriendFeed to create a space to aggregate all information and note taking from the 2008 International Conference on Intelligent Systems for Molecular Biology (ISMB). They used features of FriendFeed to set up parallel informal reporting structures for the conference without the knowledge of the organisers; they then used this information to synthesise a conference report suitable for publication. Previously, the conference journal issue consisted solely of the formally produced papers, and now there is a living snapshot of data and thoughts about the conference as it unwound in real time, available on the Web, in addition to the more formal summary which they published in *PLoS Computational Biology*.

In the case of the Saunders et al. (2009) paper, there was also an interesting dynamic with the conference organisers. At first the scientists did not mention their efforts to document the conference online to the organisers, but once the conference organisers found out they embraced the concept and in the subsequent year made sure that an official FriendFeed group was set up for the conference (see Figure 16 below).

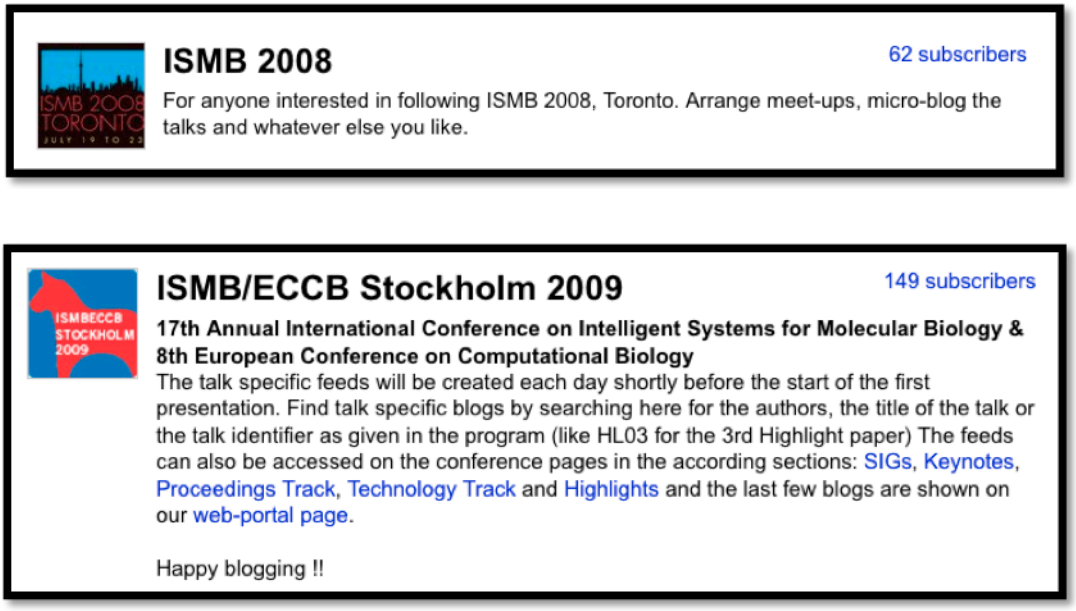


Figure 16. ISMB Conference 2008 and 2009 FriendFeed groups

Publishers and entrepreneurs are devoting some effort and resources towards building their own platforms to enable similar kinds of online communication, for example: *ResearchGate*, *Epernicus*, *SciLink* (discontinued), *Laboratree* (Indiana University, discontinued), *BioCrowd*, *2Collab* (Elsevier, discontinued April 2011), *Mendeley* and *BenchFly*.[40] Although those that are still functioning may have working scientific communities within them, they are different from the community described here for several reasons—most of them are not open—users must log in (and sometimes provide proof of institutional affiliation) before they can see any data, and several do not allow content to be indexed by Google or other search engines, which limits the findability of the sites. Others are focusing on one particular service, for example *BenchFly*, which is a site for

[40] http://www.researchgate.net/, http://www.epernicus.com/, http://www.biocrowd.com/, http://www.mendeley.com/, http://www.benchfly.com/

sharing open-access video protocols, and *Mendeley*, which focuses bibliographic management and sharing data and annotations on research papers.

It seems that most interaction happens in open and accessible online sites which other scientists are drawn to by a core founding group, such as the Biogang discussed above, which shows characteristics of a SIM in its stated challenges to current ways of doing science.

3.8.5 Scientists and public engagement

RQ: How do academic life scientists see that their engagement with the public has changed since they adopted online tools?

Scientific interaction with the public is not a significant aspect of the activities of the subjects interviewed for this case study. Although *The Life Scientists* is open to the public, it is difficult to judge what impact, if any, it has over and above the popular scientific blogs where the public engage with individual scientists via comments (an activity which will be discussed in detail in Chapter Four). There is little evidence in the discussions that non-scientists are participating in *The Life Scientists*. Most people belonging to the group self-identify as scientists and/or are employed in some scientific capacity. Although the participants are aware that their activities on FriendFeed are performed *in* public, they make a distinction between that and interacting with the general public.

However, another aspect of engaging with the public is this group's success at promoting their open science goals to the "scientific public" in

popular scientific journals and media such as *Nature*, *Science*, *Scientific American* and other highly read journals. Over the past two years, various members of the Biogang have succeeded in publishing articles, commentary, letters and opinion in a range of these publications.

Some worth highlighting include:

> *Stitching science together*, Cameron Neylon, *Nature Opinions,* 15 October 2009: Google Wave is the kind of open-source online collaboration tool that should drive scientists to wire their research and publications into an interactive data web.
>
> *Information awakening*, Michael Nielsen, *Nature Physics*, vol 5 2009: Online tools for collaboration and sharing information have changed the routine of scientists. But the revolution that will turn scientific information from a collection of files into an active system has just begun. Mentions blogging and references a FriendFeed room.
>
> *Science 2.0*, Mitchell Waldrop, *Scientific American* 298, 2008: Interviews Bill Hooker, Cameron Neylon, Jean-Claude Bradley and discusses Open Notebook Science and Open WetWare.
>
> *Communicating chemistry*, Theresa Velden & Carl Lagoze, *Nature Chemistry*, 2009: New web-based models of scholarly communication have made a significant impact in some scientific disciplines, but chemistry is not one of them. What has prevented the widespread adoption of these developments by chemists—and what are the prospects for adoption over time? Article mentions Open Notebook Science.

A full list of these publications is available in Appendix G. This success in publicising their goals in the mainstream scientific media again indicates that members of this group are participating in collective action which enrols others in their goal of transforming scientific research processes.

3.8.6 Status and Rewards

RQ: Do the various forms of online dissemination being used by academic life scientists yield rewards and status comparable to traditional outputs, or are there new mechanisms developing for reward and recognition of these activities?

Use of FriendFeed is not consciously associated with furthering career status and rewards by those who use it. However, they are able to leverage the FriendFeed network of contacts to find collaborations, publish, source and win research funding via the FriendFeed network. Although most FriendFeed users interviewed are bloggers and would list their blog on their curriculum vitae, they tend not to list their membership of an online social network. They acknowledge it has value to them but may be concerned employers would see FriendFeed as something like Facebook rather than a valid science communications network.

Finding funding is an omnipresent concern for scientists. On August 13, 2008, Bill Hooker posted a link about SubMeta, a source of funding. It received 16 likes and generated a 20-comment discussion about who might apply for it. Included in this discussion was Jean-Claude Bradley (a chemist at Drexel University who runs his laboratory using Open Notebook Science based on wikis, a system which will be examined in detail in Chapter Four).

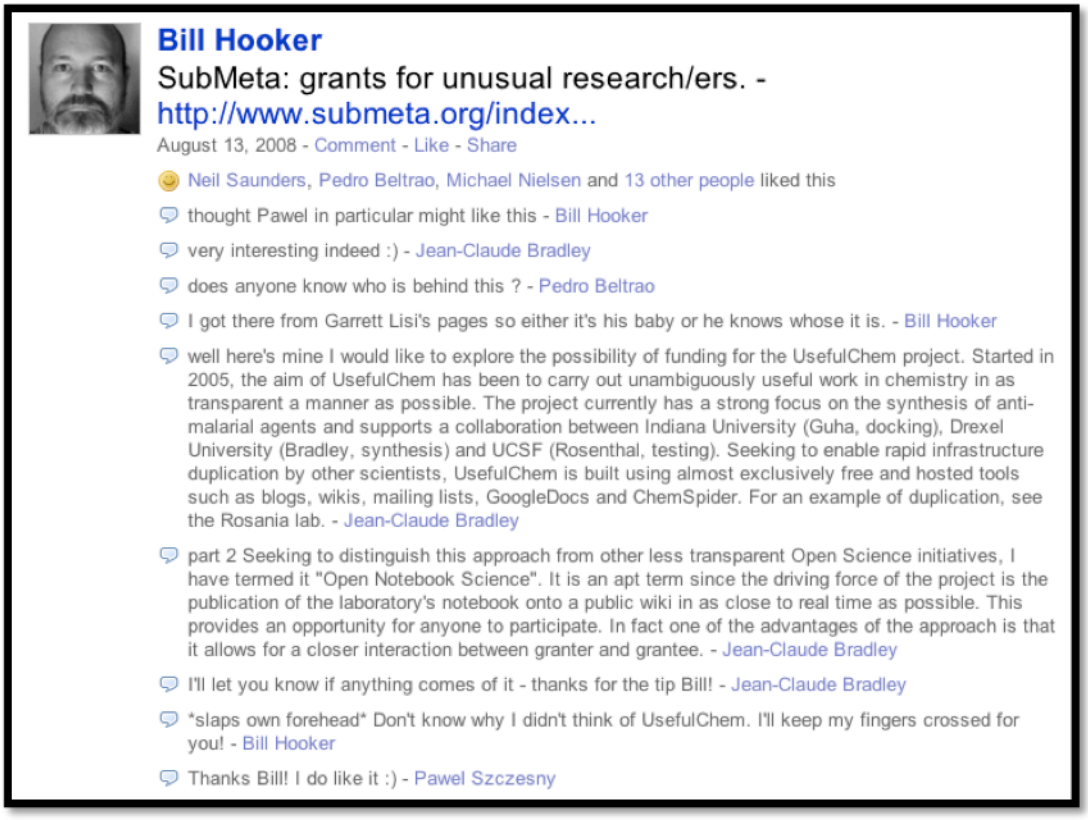


Figure 17. FriendFeed discussion about SubMeta funding (August 13, 2008).

On November 4, 2008, Jean-Claude Bradley announced on his blog (and FriendFeed) that he had received SubMeta funding for his Open Notebook Science Competition. Without the prompting and discussion on FriendFeed, the project's second round would not have gone ahead. This is an example of the coherent group of FriendFeed participants using it to drive forward the Open Notebook Science awards. Bill Hooker (the originator of the post) is also now one of the judges of Jean-Claude Bradley's Open Notebook Science competitions (Hooker interview).

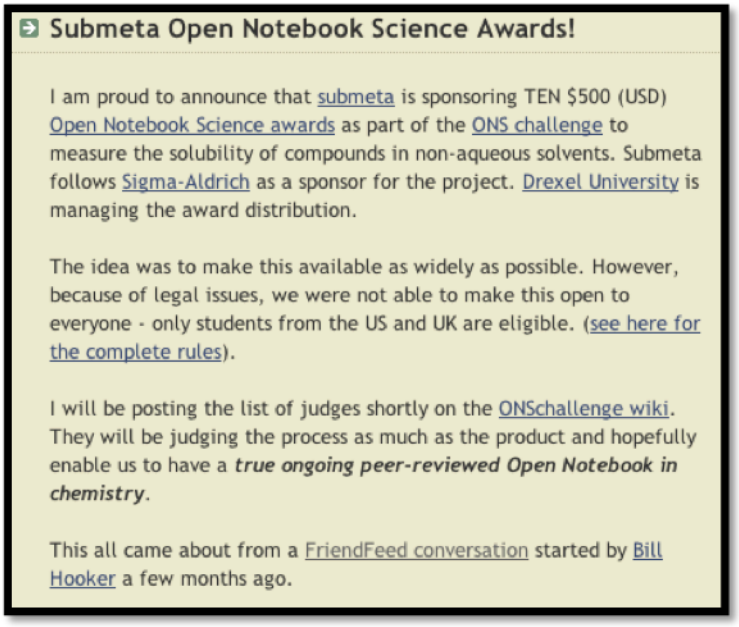


Figure 18. Jean-Claude Bradley announces SubMeta funding award on his blog.

Long-term job stability in academia as measured by availability of tenure track positions is, in general, declining (Kaplan, 2010). Several countries no longer offer tenured positions, and the competition for tenure in those countries that have it is intense. This tends to foster a disinclination to experiment with new forms of communication or status-making among younger academics, although they do express a wish for the status quo to change (Harley et al., 2010). However, for scientists who participate in online networks like FriendFeed, they consider the new connections they make and collaborations they enter into of great value to their future careers. There is not currently an established role for online social networks in science, but those participating now may well benefit in future from the networking opportunities offered by this communication style.

3.8.7 Openness of scientific data and discourse

RQ: Where in the chain of collecting and disseminating research materials are the incentives and barriers to open practices in the use of online tools by academic life scientists?

FriendFeed provides an online space which facilitates the collection and dissemination of research materials. As demonstrated in this chapter, FriendFeed use is mostly about informal scientific communication, rather than data-sharing. There are posts about particular data sources and notifications when a new data source of interest comes online or launches new features, but much of the general discussion is about the concept of openness in science rather than actually openly sharing data. There are debates about what should and should not be open, and reactions to news announcements concerning publishing and openness. These discussions indicate the presence of an invisible college of scientists who share common interests in questioning and exploring the boundaries of the current science communications system.

The group members are conscious that their use of FriendFeed is unusual among scientists. They frequently reflect on their use of the group and even conduct small experiments or analysis of the group's activities themselves. Some examples include:

- Pedro Beltrao experimenting with collaborative work based on sharing data via FriendFeed (tracking of actions) and Google documents.

- Neil Saunders' quantitative analysis (on his blog) of FriendFeed Life Scientist group activities, including some network analysis.
- Cameron Neylon led collaborative work on a paper about the use of scientific social networks, detailed above in Section 3.8.2, Online Collaboration, as well as several blog posts for scientists on how to use FriendFeed and discussions about what a standalone FriendFeed for scientists would look like (discussions about this were prompted by Facebook's purchase of FriendFeed).

The examples above are a small sample; there are many further discussions about the scientists own communication practices on FriendFeed and their blogs. Participants in *The Life Scientists* group seem to be self-aware in relation to all of the research aspects discussed in the research questions above.

FriendFeed seems to be a communications hub for many who are involved in promoting open science, and it is in the discussions about open science that take place that there is evidence of a Scientific Intellectual Movement (SIM), as the group members enunciate their visions of what open science should be through discussion:

> To me open science means having the information available so that you can reproduce an experiment, you can take it further ... You should be able to practice science ... and you should be able to find the scientific data that you need and follow through on it without being restricted by subscription, restricted by licensing, all kinds of other restrictions that people put in force—whether they're cultural or legal. To me that is open science. Because to me science is something that belongs to

> everybody. Especially the natural sciences. They're not things that we're making, they're things that exist. And being able to study them—everybody should be able to, without restriction. (Singh interview)

> Going back to open science, if everyone is sharing everything then yes, you perhaps have more to lose by not participating. But at the stage it is now, some people actually think they've got a lot to lose. They may be right, they may not. It's still an ongoing experiment. (Grant interview)

As highlighted above in the section on the Biogang, Frickel and Gross (2005) hypothesise that the emergence of a SIM is facilitated by several key conditions, including: high status individuals within the field harbouring complaints about certain conditions or dominant schools of thought; structural conditions providing access to key resources; access to micromobilisation contexts (that is, an ability to recruit); and members of the SIM frame the intellectual arguments attractively. In this sense, the actions of life scientists on FriendFeed as it permeates their everyday worklife do appear to be "mundane actions directed at contentious ends", namely the task of transforming scientific communications (Frickel & Gross, 2005, p. 226).

3.9 Conclusion

The main research question of this thesis asks: "Is the use of online tools revolutionising, reinforcing or having unintended consequences for the research practices of academic life scientists?" As we have seen in this chapter, the changes wrought by the use of online tools are piecemeal, emergent, and still incomplete. What we have also seen, however, is the formation of a group which espouses shared norms and the use of a new

tool which seeks to challenge current ways of doing science. Frickel & Gross (2005) and others (Griffith & Mullins, 1972; Parker & Hackett, 2009) have argued that new SIMs come about when a number of people set themselves up in opposition to existing groups or practices and mobilise a variety of resources in support of their own goals. Similarly, the theory of Computerisation Movements (Kling & Iacono, 1994; Elliott & Kraemer, 2008) has argued that the way in which new tools are introduced into not just research, but a variety of other areas, is that a number of groups who are dissatisfied with existing tools form a group behind which they enrol various resources to create new tools.

As far as new online tools and practices are concerned, both of these theories have one part of the picture, but they also have some gaps. In their general theory of SIMs, Frickel & Gross talk about SIM emergence whereby groups of challengers, in the manner of social movements, set themselves up in opposition to the established current of thought, mobilising the resources (publications, positions, creating specialisms) with which they seek to promote a new intellectual current. This current, as discussed in the Introduction to this thesis, extends open practices across all activities in the research process, including, as shown within this chapter, using social scientific networking to filter information, and get answers to questions emerging at all stages of the research cycle. Clearly, the FriendFeed group presented in this chapter adopts productive strategies for doing research which renew and extend previously established practices: openness,

sharing, and online collaboration outside existing channels are being promoted as part of a new agenda in the life sciences. The General Theory of SIMs, however, applies to intellectual or scientific change without reference to the role of technology, which can be a means for dissemination of and strengthening the SIM. Computerisation Movement theory, on the other hand, considers the ways in which new technologies are brought into play and stabilised, again marshalling resources to do so, but deals with socio-technical change rather than addressing intellectual or scientific change. A combination of these two theories provides the best fit for the practices shown in this case.

My contribution in this FriendFeed case study is to present evidence which marries these two approaches and thereby fills a gap in the research literature which shows how online tools are being harnessed to spearhead new ways of doing life sciences. In the case of FriendFeed—and (though in different ways, as we shall see) in the cases of open laboratory notebooks and science blogging (Chapters 4 and 5)—there is a proliferation of more and less successful attempts to use new tools, in this case for online scientific networking.

The amount of attention garnered by the Biogang and *The Life Scientists* group, particularly their success in disseminating ideas in the mainstream scientific media, shows that these new ideas and practices around these tools are becoming recognised, and are shifting research and communication practices by means of these tools in the direction of the

norms espoused by the Biogang and *The Life Scientists* (openness, sharing resources, promoting distributed collaboration across boundaries). Though it is of course too early to compare these changes to science and knowledge to those ushered in by Gutenberg's introduction of the printing press, the self-image of these challengers is that they are introducing potentially far-reaching changes in scientific practice, and within the limited scope of a small group and their initial successes with the use of new tools, they are becoming recognised as an innovative group.

I have demonstrated that across the seven research aspects which characterise the organisation and procedures of scientific work, there is a reconfiguration of scientific boundaries taking place, which some in science may regard as revolutionary, but in reality are a mixture of reinforcing current research practices (by making the invisible colleges mainly visible in high status groups), having unintended consequences for others (by creating new niches among a small newly visible group with an agenda to change life science practices), and transforming (even if not revolutionising) yet other practices such as putting together research proposals with distributed collaborators, collaborative filtering of information, and sharing information sources via informal channels. As we shall see, similarly mixed changes, albeit in different settings, can also be found in the cases of open laboratory notebooks and science blogging.

Chapter 4
Open Notebook Science: Data in the Public Sphere

4.1 Introduction

Scientists have long used paper notebooks to record what goes on in their laboratories. Isaac Newton recorded his seventeenth-century experiments in handwritten and illustrated detail. In the present day, this recording continues almost unaltered, as academic scientists still document their research in paper notebooks, hand-writing the details of failed and successful experiments and pasting-in data and images. Most of these contemporary records may not be of future historical interest like Newton's laboratory notebooks, but they have the same function, which is to serve as a record of scientific work, and so they are an essential tool for the production of scientific knowledge and establishing provenance over ideas and data.

As old laboratory notebooks linger in dusty departmental storerooms or in the bottom drawers of graduate students' and professors' filing cabinets, the data they contain is preserved only in the loosest sense. These data are not indexed or searchable, and cannot easily be found again. However, the introduction of computerisation in laboratory science has had an impact on the way scientists think about, store and use data. The way experiments are recorded in laboratories is changing, and some scientists are challenging the ways in which this data is recorded, and most particularly, shared. This sharing is in turn having an impact on other established formal scientific

processes such as collaboration and publication, as well as informal communication, scientific roles and scientific interactions and connections with the public.

The previous chapter examined life scientists' use of FriendFeed, an online social networking and information filtering tool, and demonstrated that their use of it is akin to the concept of "invisible colleges" of online colleagues and that such emergent online behaviours are beginning to challenge the predominant way scientific research is produced and disseminated. It argued that a Scientific/Intellectual Movement (SIM) has formed around the concept of changing scientific research practices and its activities are facilitated and disseminated by the use of online tools.

This chapter presents the case of another emergent online activity, the use of Open Laboratory Notebooks (OLNs), a use which is also known as Open Notebook Science (ONS).[41] OLNs are online tools used to publish experimental scientific research and data, straight from the laboratory bench to the public Internet. Scientists using OLNs are also members of the aforementioned SIM seeking to change research practices using online tools. In addition this chapter will demonstrate that these scientists are involved in the related CM as they attempt to move the records of laboratory science entirely online. Also within this demonstration is an examination of the properties and outcomes of changing text on paper to text on the Web—does notebook digitisation transform the laboratory into a

[41] http://en.wikipedia.org/wiki/Open_Notebook_Science

data machine, making data available for use and re-use due to the inherent qualities of hypertext linking?

4.2 Chapter structure and scope

Laboratory notebooks are central to scientific knowledge production, and for this reason it is necessary to initially explain the types of scientific tasks that are performed in a laboratory setting, how data is produced within the laboratory, and how science practices relate to prevalent forms of formal knowledge dissemination via print publication. This scene-setting of common practices later provides a contrast to the descriptions of OLN use. The bulk of this chapter describes how OLNs are used within the academic laboratory, and details the impacts that their use is having on users, the wider scientific community, the public, and science.

Prior to this description, a review of relevant literature canvassing current research on laboratory life and practices, the collection and retention of electronic data, and the small body of research on "Science 2.0" (using Web technologies to share, store and/or disseminate scientific research, with differing degrees of openness), and open science practices is required. It should be noted that this literature is itself largely descriptive in nature, as until the advent of OLNs the mysteries of laboratory practices could only be examined via painstaking reconstruction of processes from historical notebooks (for example, Gooding's examination of Faraday's notebooks); commitment to a detailed ethnography of a particular laboratory (for example, Latour and Woolgar's *Laboratory Life)*; or being

involved in the development of laboratory software with its concomitant explorations of how scientists document their work in the laboratory (examples below in Section 4.4.1).

This chapter also examines whether the data committed to Open Laboratory Notebooks changes not just scientific communications but also science itself. The data is digitised, which makes it easily transferable, but also, unlike the bulk of scientific data today, its openness is its key feature, as it allows unprecedented access to and sharing of raw experimental data. This enables accurate replication of experiments when required, and access to the building blocks of the scientific journal article, released from the strictures of that format while still presented in a style familiar to scientists.

OLN usage is examined in six laboratories located in the USA, the UK, and Australia. The few laboratories that are actively practicing open notebook science in biology, biophysics and organic chemistry dictate the boundaries of this case. Other scientists are also practicing Open Notebook Science but they are either in other fields (for example, mathematics) or work as solo researchers. Eleven interviews have been conducted with OLN users, ranging from professors and principal investigators to graduate students. The scientists were asked about their use of OLNs, how they got started using them, what the impacts of their use have been, and where relevant, what drove their technology selection decisions.

Some of the academics involved in the practice and/or promotion of Open Laboratory Notebooks are chemists and mathematicians. Extension

beyond the boundaries of life science is justified because these scientists are an active part of the open science SIM and collaborate formally and informally with the life scientists and members of other scientific disciplines.

4.3 The role of laboratory notebooks in science

Laboratory notebooks are the canonical record of experimental plans, protocols and results. Experimenters are obliged to maintain one, and document their experiments to the standards expected by their laboratory. These standards vary by institution and departments or laboratories within institutions, and/or have different levels of enforcement. However, institutions or departments usually produce guidelines for laboratory notebook use. For example, the University of Oxford Research Services Office provides the following recommendations:

> A well-maintained and properly documented laboratory notebook establishes a permanent record of research protocols and results which can be referred to in the future, most commonly in the preparation of scientific papers and reports. Additionally, it can be an invaluable source of information for a variety of purposes, including determination of claims of discovery where new inventions are concerned; demonstration of adherence to standards of good practice, and of academic and ethical integrity; and compliance with contractual provisions permitting sponsors to audit work carried out in pursuit of sponsored research. (University of Oxford Research Services webmaster, ¶2, 2010)

The University of Oxford also recommends that paper notebooks be retained by a department for a minimum of six years after completion of a research project or after a researcher leaves, and in particular mentions that:

> Electronic notes should be avoided. However, if such notes are taken hard copies of the entries should be printed out regularly, signed, dated and affixed to consecutive pages of a bound notebook. (University of Oxford Research Services webmaster, ¶2, 2010)

Real practice in the laboratory can contrast greatly with the recommended approaches, as illustrated below in the quote from a senior scientist speaking to researchers of the Harley et al. (2010) report on biology in their study of the "future landscape" of scholarly communications:

> I retain none of my primary data. When my students leave, they take all of their notebooks with them with my blessing, unless there happens to be a notebook that's critical to an ongoing project. I don't retain any primary data, because I don't have the capacity to. I wouldn't be able to walk in my office if I had every notebook from every graduate student who's ever worked in my lab here. So the university does not provide the archiving capability for doing that, and the data just disappear. And I'm living with some element of risk there because if I'm asked, in the case of a fraud investigation or whatever, I have to depend upon the goodwill of my students to maintain their notebooks. (Harley et al., 2010, p. 281)

Depending on the level of seniority or supervision, the laboratory notebook is very much an individual scientist's record of their work. Laboratory notebooks should record the experimental processes and outcomes in order to facilitate writing scientific papers, with their formal structure of Abstract, Introduction, Materials and Methods, Results, and Discussion/Conclusion. In many fields, experiments are not easily represented in a paper notebook, as references must be made to data or tables or movies that are too large or unsuitable for the paper format. Despite this, most academic laboratories still use paper notebooks, which in many cases remain the property of the institution employing the researcher

after their departure, although this may vary between institutions and disciplines. The reasons cited for retention of this data are varied, and include: protection against future intellectual property litigation (they are an essential part of establishing the provenance of specific ideas); ease of continuing particular projects after a graduate student or researcher leaves the laboratory; and aids for necessary reproduction of research or prevention of unnecessary repetition.

The variety of approaches to laboratory notebooks can be illustrated by highlighting some findings from a workshop on blog-based open laboratory notebooks I attended in July 2008. Cameron Neylon, a contributor to this research in the FriendFeed, blogging and OLN cases, hosted the one-day workshop. Several senior scientists, some social scientists and graduate students discussed the use of laboratory notebooks. Answers to questions posed at the workshop revealed a great deal about current usage of laboratory notebooks. In this group of thirteen people, there were wide variations in how relatively standard data was recorded, including time and date, experiment number or name, material batch numbers, supervisor's signature, references to other data, printouts of data, level of detail in protocols and safety data.[42] Scientists at the workshop also provided a wide range of answers as to where raw data was stored,

---

[42] Blog summaries of the workshop: http://lucypower.com/126-blogging-the-lab-book/, http://cameronneylon.net/blog/placeholder-post-for-workshop-comments/, http://duncan.hull.name/2008/02/29/lablogs/

including the individual's own computer, various laboratory computers, a laboratory server, supervisors' computer, and/or in a file of computer printouts or as printouts pasted into a laboratory notebook. This experience served to reinforce the findings of the research detailed in the literature review below, and has highlighted the variety of laboratory notebook practices in science.

Modern science produces much data in digital format. In the life sciences, the problem of the "data deluge" is often cited (Hey & Trefethen, 2003b). In many life science laboratories, the data is not only a deluge, it resides in different places without a unifying naming standard that renders each piece of data attached to an experiment or protocol. For commercial laboratories, the resolution of this floating data (or "dark data") issue is being addressed by the use of electronic laboratory notebooks (ELNs).

Electronic laboratory notebooks are widely used in commercial biosciences, drug discovery and research and development laboratories, particularly since electronic records became admissible court evidence in the USA in 2006 (Federal Rules of Civil Procedure, Rule 34 (b) (2), 2010).[43] Before that, a hybrid notebook was used, where electronic records were printed out, witnessed and signed to comply with regulatory and intellectual property policies in private laboratories. However, ELNs are not in widespread use in academic settings, partly because of institutional recommendations as noted above, as well as implementation costs and less

[43] http://www.law.cornell.edu/rules/frcp/Rule34.htm

tangible reasons (Taylor, 2006). For example, schraefel et al., in their study which attempts to elicit the essential qualities of "notebooks as artefacts", note that scientists they interviewed had a negative "visceral" response to the idea of commercial electronic laboratory notebooks, saying "We wouldn't touch those with a barge pole. They're horrible... Not for us" — although, as the authors note, the commercial ELNs do appear to capture the essential data required. (schraefel, 2009, p. 1).[44]

The summary above shows the diversity of approaches to paper laboratory notebooks, which is supported by prior research on laboratory life and practices detailed in the literature review below.

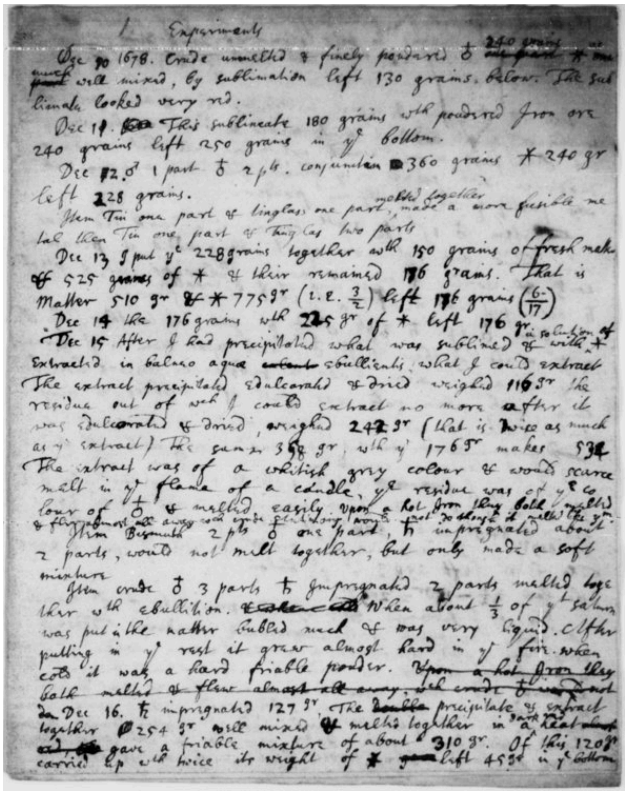
4.4 Relevant literature

This section reviews the relevant literature in research on laboratory life and practices, knowledge production, and the collection, retention and dissemination of electronic data. Research literature about laboratories and laboratory work practices falls into three categories: description in sociological examinations of knowledge production practices, for example the laboratory studies of Knorr-Cetina (1981), Latour & Woolgar (1986), Lynch (1985), and Traweek (1992); accounts of the history of science or development of specific scientific technologies, for example Rabinow's 1997 account of the development of the polymerase chain reaction (PCR) technology; and descriptions that are part of projects seeking to implement

[44] Note: author uses lower-case surname in publications. See http://eprints.ecs.soton.ac.uk/8780/

or prototype electronic laboratory notebooks in previously paper-based laboratories, that is, part of the computer-supported cooperative work field; for example Polonsky et al. (2006) and Tabard et al. (2008). There is also a limited amount of research literature on connecting scientific work practices and use of the Internet, sometimes called "Science 2.0". Within this thesis, Science 2.0 is defined as the use of Web 2.0 technologies to facilitate open science practices. A formal and widely agreed definition has not yet solidified.[45]

4.4.1 Laboratory Life & Practices: Constructing the Scientific Record



[45] Waldrop's definition from his 2008 Scientific American article is "how researchers are beginning to harness wikis, blogs and other Web 2.0 technologies as a potentially transformative way of doing science" (Waldrop, 2008). The Science 2.0 Wikipedia page also draws on Waldrop, although this page's neutrality is being disputed (February 18, 2009). Under these circumstances I thought it best to use my own definition within this thesis.

Figure 19. A page from one of Isaac Newton's (1678) experimental notebooks (Newton, 2006).

Laboratory notebooks have not historically attracted much scholarly attention. Holmes, Renn and Rheinberger (2003) report that reconstruction of past processes of scientific work through the notebooks of scientists such as Galileo, Newton and Darwin began in the post-war period, and until relatively recently even this study was overshadowed by the scholarly attention given to their correspondence with others rather than their experimental notebooks.

Bazerman (1988) writes about reconstructing scientific practice using historical laboratory notebooks, and features accounts of how notable past scientists such as Isaac Newton have shaped their academic writing from their notebooks through to formal publications. Bazerman notes that scientists keep their messy, unpolished thoughts in their laboratory notebooks. For example, after highlighting the praise garnered by Newton's *Opticks* (1704), both at publication and subsequently, Bazerman comments that "the texts that are closer to the spontaneous outpourings of Newton's mind, such as his student notebook, have hardly the compelling presence [of *Opticks*]" (Bazerman, 1988, p. 124). Journal articles are then the construction of a public version of the laboratory event, and before now the messy, unpolished thought constructed in the laboratory was not easily accessible—although OLNs, as we shall see, make this thought accessible, alongside the raw experimental data.

This view is supported by Gooding (1990), who sets out the case for using original material such as experimental notebooks to reconstruct knowledge practices, using the notebooks of Michael Faraday as an example (Gooding, 1990). Gooding reproduces Faraday's experiments as workflow diagrams, which allows comparison of the "exploratory process with the published narrative ... [which] should help us to understand how experimental narratives bring together cognitive and social elements of experience to generate arguments that make the best evidential and rhetorical use of experiments" (Gooding, 1990, p. 192). OLNs can allow for a further opening up of this process by facilitating direct comparison of research notes with final publication.

Holmes, Renn and Rheinberger (2003) describe fourteen cases of laboratory notebook examination by historians, which can reveal the thought processes behind the construction and testing of scientific theories. As they note, the beginning of the scientific process is contained in the laboratory notebooks, and the end is in the formal publications they produce, which often bear little resemblance to the data contained in the notebooks:

> Research notes – or scribblings, as some forms of them might aptly be called – represent a special genre of scientific writing. They are literary activities in their own right, circumscribing a space that lies between the materialities of experimental arrangement, or the unexplored potentials of theoretical formalisms, and the structured formats of printed communication that are released eventually to the scientific community. (Holmes et al., 2003, p. viii)

Most recently, continuing the lines of thought established by Bazerman, Gooding, and Holmes et al., Wickman (2010), in his summary of a year-long ethnography of the laboratory practices of a single chemist, characterises the laboratory notebook as a "constructive resource", in that it allows physical, conceptual and rhetorical exploration of the actualities of the experiments, the wider theoretical concepts behind them, and the linkages to specific characteristics of the academic research field (Wickman, 2010, p. 264).

In recent years a strand of STS has engaged in undertaking ethnographies of the laboratory, with the research goal of understanding underlying epistemological processes and outcomes in science. There are several well-known ethnographic sociological studies which detail work in life science laboratories. Relevant contributions to this discussion from these works include Knorr-Cetina, 1981; Woolgar, 1982; Lynch, 1985; Latour and Woolgar, 1986; Latour, 1988; Traweek, 1992; Rabinow, 1997; Knorr-Cetina, 1999; and Lynch, 2002.

The focus of these studies is an anthropological approach to laboratory studies characterised by long-term fieldwork, and theoretically informed by various approaches to the social construction of scientific knowledge. This thesis is not questioning the origins of knowledge construction in science, so their theoretical discussions are less relevant to my immediate purposes, but this literature has made valuable contributions to the body of knowledge about day-to-day work in scientific laboratories. Lynch's 1985

and 2002 studies point out the discrepancies between the written-down protocols of the laboratory and the actual bench practices, again highlighting the "messy" practices of actual science. Lynch questioned the veracity of descriptions of method that appear in journal articles by saying that this they do not square with what was in the original notebooks. The use of OLNs could deflate this issue, as original methods and descriptions of activities in the laboratory are directly comparable with the finished journal product, and can even be referenced within the journal article.

In his study of the Handelsman plant sciences laboratory at the University of Wisconsin, Kleinman (1998) highlights the above studies as examples of laboratories that are relatively insulated from commercial contact and pressures, in contrast to the Handelsman laboratory, which is in close contact with commercial entities. Kleinman argues that studies of "insulated" laboratories "must surely have affected what previous scholars have seen", suggesting that there are historical and structural features of the institutions around the laboratory that also shape laboratory practices (Kleinman, 1998, p. 294). Kleinman adds to the argument that the scientific production of knowledge is multifaceted, consisting not just of social relations but also of the equipment and processes behind the creation and maintenance of the equipment of the laboratory. In essence, this equates to the concept of Kling et al.'s STIN models cited in the Introduction to this research (Kling et al., 2003).

These studies also build on the theoretical perspective put forward by Fuchs (1992), who again states that "laboratories are especially powerful nontextual agents supporting scientific statements", powerful, he argues, partly because the data produced by machines is difficult to deconstruct and question (Fuchs, 1992, p. 66). OLNs potentially open up the "black box" of scientific laboratories. Does the fact that the finished product (the journal paper) can be directly compared to the messy thought processes and machine-made data of the OLN make science different?

Another group of research articles about laboratory life are studies done by researchers in Human-Computer Interaction (HCI) or Computer-Supported Cooperative Work (CSCW), often about prototype implementations of electronic laboratory notebooks. Studies seeking to implement different electronic notebook prototypes or projects reveal much about contemporary ways of working with paper notebooks in biology laboratories. Several studies in this vein attempt to characterise the information needs of biologists as well as examining use of their own ELN prototype.

Mackay, Pothier, Letondal, Boegh and Sorensen (2002) prototyped three kinds of "augmented laboratory notebook" via participatory design, and as part of this process they videotaped biologists at the Institut Pasteur in Paris using paper laboratory notebooks, and commented on their ways of working. Comments they made on their observations show that laboratory practices are similar in general, but non-standardised across laboratories, as

they echo the comments made by the small group of scientists attending the Abingdon workshop on electronic laboratory notebooks described in this chapter's introduction. Their observations of paper notebook users say that notebooks are assembled as multi-media documents, with writing, drawings, pasted images of X-rays or microscopy, pasted computer printouts of raw data and results, and sometimes even physical specimens. In addition, they observe that:

> [Paper] notebooks are also used to point to external objects, such as refrigerated or frozen specimens, surgical results stored on slides and even live laboratory animals. Notebooks also provide links to on-line information, including important research articles and web sites. (Mackay et al., 2002, p. 42)

Mackay et al. also asked scientists about the disadvantages of paper notebooks, and noted that the inability to search them online or through any other method was considered an issue, particularly in cases where minor experimental variations needed tracking. In addition, creating manual tables of contents or indexes of their work was time-consuming.

Polonsky at al. (2006) analysed 40 randomly-selected papers from academia as well as what they term analyst reports and the professional press (they do not expand on this nor provide a list of the papers) and derived a classification of information needs in the biomedical sciences. From this they decided that there were seven "top-level needs" represented in these reports, with several sub-sections each, including: collective data management (search quality, sustainable across different media); data storage (electronic); data integration across different media; personal data

management (a subset of collective data management); project management, and communication (automatic reporting and publishing) (Polonsky et al., 2006).

This shows that although a wide range of needs are perceived by biomedical researchers, very few of them can be achieved using paper laboratory notebooks (for example, search, which is named by OLN users as one of the advantages of their use), and neither can electronic notebooks fulfil all of the information needs on this list. As will be discussed further below, although storage is now plentiful and cheap, objectives such as data integration across different media and automatic reporting and publishing are still future aspirations for OLN users.

Polonsky et al. (2006) argue that electronic laboratory notebooks force users to choose between flexible but unstructured data entry (which limits semantic linking opportunities) or rigid data organisation (which can limit creativity) and maintain that this dichotomy is the reason why many have not adopted ELNs (p. 2). There is some evidence to support this assertion, but as this chapter will show, the choices made by users of OLNs are not dictated or constrained by specific technical issues or features, but rather the ability to quickly publish and store many forms of media generated in the course of experimentation. OLN use is marked by expediency rather than perfectionism. However, some users express a desire for semantic linking features (links with extra information, or metadata, which supply information about the content and context of the link), and the

establishment of firm standards around data publication, but the lack of these features does not dissuade OLN users from getting started, and is thus not a prerequisite for OLN use.

Klokmose and Zander (2010) acknowledge that the use of paper for laboratory notebooks is still far more widespread than online or computer notebooks, but approach this issue in a different way – they examine whether "there is something inherent in ICT infrastructure that invites resistance from the users" (Klokmose and Zander, 2010, p. 119). They use interviews and design workshops to elicit information from their small research sample of five PhD students from the Physics and Astronomy department at Aarhus University, Denmark. Their main finding was that "integration between the different aspects of their activities in the laboratories is one of the key aspects missing in their current use of laboratory notebooks"; that is, that experimental work in physics has high task variability, and paper notebooks are not sufficient for recording all aspects of these tasks. Given this observation, it appears that there is a space waiting to be filled by electronic notebooks, since if the paper notebooks are not sufficient, why aren't more scientists using electronic laboratory notebooks, which can go some way towards fulfilling this integration? This question will be examined in the Discussion section of this chapter.

The common thread running through these reports and analyses of academic laboratory life and ways of recording information in the laboratory is that experimental information is widely distributed even

within the laboratory, non-standardised and loosely governed; for example, the data created is scattered around various non-public storage media; only sporadically archived; occasionally governed by institutional rules, which may or may not be adhered to; and generally not recorded in a way that reduces unnecessary repetition of experiments or enables close reproduction of experiments if required. Much of the information required to tie these threads of experimental information together is kept in the researcher's personal information store – their brain – and is thus lost when researchers move out of the laboratory. The research literature on these areas leads to the expectation that open online laboratory notebook users may be deriving solutions to the kinds of problems experienced with paper notebooks, such as searchability, storability, shareability, and experimental consistency. This will be examined in the Results and Discussion section of this chapter.

4.4.2 Electronic Data: Collection, Retention and Dissemination

Electronic data management practices are relevant to this discussion because a large part of the activity of maintaining an open laboratory notebook is concerned with data—transferring it from the laboratory bench to the notebook, and once it is online, handling and presenting the data in a way that makes it human-readable, citeable in journal articles, able to be archived and made machine-readable for search engine indexing, ideally in a manner that makes it interoperable with data from many other machine

sources, although as shown below, this is difficult to achieve and is not yet happening.

Life sciences research is increasingly a data-driven science, as shown by the emergence of bioinformatics as a scientific discipline and the rapid increase in available genomics data as more organisms are fully sequenced (den Besten, 2003; Welsh, Jirotka & Gavaghan, 2006). This data can be mined for biologically relevant patterns, used to generate hypotheses for experimentation, and as an aid in building models of biological function (Blake & Bult, 2006). However, unlike particle physics, for example, biologists don't tend to work in huge collaborations, with the result that the standards around biological data are not consistent, which means that much data is not interoperable.

Preprint archives are not generally used in the biosciences, and although data sharing is common in certain fields, it is not shared until after publication, and sharing is often mandated by funding or publication requirements (Harley et al., 2010). Harley et al. describe three main data sharing methods: data published as a supplement to a journal article, deposited in an institutional or other repository, or shared ad-hoc on a personal level. Harley et al.'s interviews with 25 biological scientists reveal a variety of attitudes towards data sharing and archiving, including concern over a lack of "good" archives for data and absence of a shared data presentation standard, differing journal standards for submission of supplementary data, under- or over-curated data (under-curated, or raw

data is difficult to work with, and over-curated processed data may not contain enough information to be useful to other researchers).

Internal data-sharing works well within some large projects and in particular fields (development of the atomic bomb, high-energy physics, radio astronomy) but biological data (a) lends itself less well to multiple projects—often the smaller scale means that results may only be relevant to very specific problems being tackled with experiments by one laboratory, and (b) until recently there has not been a lot of data sharing in the life sciences, apart from genomic data (Lathe, Williams, Mangan, & Karolchik, 2008). This means, in general, that much work needs to be done in the life sciences to catalogue and curate existing and future data in order to reduce repetition and provide technical and theoretical insights (Mead, Shadforth & Bessant, 2007).

Does the use of OLNs reduce any of these stresses? Perhaps, as it makes it easy to see the development of ideas, establishes the provenance of work and work timelines due to consistency and interaction with others (not just reliance on server timestamps). However, the issues of archiving and citation are just as relevant and concerning for OLN users.

4.4.3 Secrecy, scooping and experimental repetition

Unnecessary repetition of research, the inability to reproduce others' experiments due to a lack of information, and slow production of results due to maintaining secrecy in scientific work are issues in current scientific practice that open scientists seek to overcome. Evidence that these issues

exist is provided below, and the activities of the open scientists examined later in this chapter will be evaluated in this context. This section will demonstrate that the life sciences are secretive in many fields, and that researchers perceive that secrecy is increasing in some fields. Outside of commercial interests it is a time-limited secrecy, in that the goal is still peer-reviewed publication, but delaying both publication and promotion via research community gatherings such as conferences is relatively common as a method of ensuring priority or protecting intellectual property.

Secrecy in science has a long history, landmarked by such notables as Leonardo da Vinci, writing backwards and impenetrably in his Codex Leicester (1406-1407); the obfuscation and codes employed by the alchemical explorations of Newton, Hooke and Boyle in the 17th and 18th centuries, and the secret laboratories harnessing the atom and turning it into a weapon during the 1940's Manhattan Project. As journal publications developed into the preeminent mode of allocating academic rewards, scientists shared knowledge cautiously, assessing the perceived risk that their generosity and belief in the open processes of science could lead to them being "scooped".

Secrecy in academic laboratory research has markedly increased over the last thirty years, due to competition for limited funding sources and increasing institutional interest in commercialisation due to the Bayh-Dole

Act (1980)[46] and similar legislation since enacted in other countries. Academic laboratories are in competition with each other for limited funding, and this can manifest itself in secretive behaviour. Most scientists can relate anecdotes of laboratories they have worked in where secrecy of data and results was the norm, and materials sharing was rare or unknown (McCain, 1991). This anecdotal evidence is supported by studies which use survey and bibliometric data to assess the effects of commercialisation and competition, both of which have increased in recent years (Grushcow, 2004; Hong and Walsh, 2009; Rodriguez, 2009). Hong and Walsh compared the results of two sets of survey data - Hagstrom, (1974), 1976 subjects and Walsh et al., (2000), 399 subjects. Their analysis showed that secrecy has increased in this time period across all of the fields studied (experimental biology, mathematics and physics), especially in experimental biology, and furthermore "that this change in secrecy is associated with the growth in scientific competition during this period" (Hong & Walsh, 2009, p. 158).

In his study of the *Arabidopsis thaliana* (a plant model organism)[47] community, Evans (2010) found that industry sponsorship reduces sharing of methods and materials, but sponsored scientists are more likely to have shared pre-publication manuscripts as a mitigation of this fact for the rest of

---

[46] The Bayh-Dole Act "allowed universities and small businesses to retain title to federally funded inventions and to retain the revenue generated by licensing such inventions to private corporations" (Grushcow, 2004). Mirroring legislation has been enacted elsewhere, which has encouraged the establishment of university Technology Transfer Offices to capitalise on research outputs (Hong & Walsh, 2009; Rodriguez, 2009).

[47] "A model organism is a [non-human] species that has been widely studied, usually because it is easy to maintain and breed in a laboratory setting and has particular experimental advantages" From: http://genome.wellcome.ac.uk/doc_WTD020803.html.

the community. However, industrial involvement still has a chilling effect on sharing, as Evans concludes that "scientists find it difficult to enrol other academics in their research findings as they become enrolled in an industry project of avoiding and mitigating competing science" (Evans, 2010, p. 757). Evans analysed citation data from *Arabadopsis thaliana* papers to evaluate sharing in the community, and interviewed scientists in academia and industry to confirm his findings.

These findings are further supported by bibliometric research data. Grushcow's (2004) analysis of approximately two hundred meeting abstracts and subsequent matching publications (about 15% of the meeting presentations) from the 1980 and 1990 meetings of the American Association of Cancer Research measured secrecy by recording the gap between cancer researchers' presentation of research at a meeting and its publication in a peer-reviewed journal. A short time gap showed that data had been withheld, rather than exposed to informal discussion within the community, whereas a long time gap between meeting presentation and publication meant that even early data had been readily shared with the research community, as meeting presentations allow for discussion and refinement of ideas, formation of new collaborations, or similar research to be redirected or abandoned if a group discovers they are doing parallel work. Grushcow found that "scientists seeking to patent their work do in fact withhold presentation of their data at scientific meetings" (p. 1).

Although every scientist can relate an anecdote about scooping, there appears to be little evidence to support the idea that scooping is widespread. Of course it may be that reporting of the practice is purposely low, as it is difficult to prove scooping, and a wrongful accusation could be damaging to the accuser. There are many instances of similar discoveries being made at the same time with no knowledge of the other; some discoverers have even shared a Nobel Prize for their unconnected discoveries.[48]

Campbell et al., (2002) investigating data withholding in academic genetics, surveyed 1849 scientists receiving NIH funding, of whom 1240 self-identified as geneticists. Their survey confirmed the findings of previous studies which identified growth in secretive behaviour (12% of geneticists reported denying information or materials requests), and also unearthed a few of its negative effects. Respondents reported that they felt this withholding inhibited necessary reproduction of others' experiments, the timely production of their own publications, and the research directions they chose to follow (Campbell et al., 2002). Furthermore, as Evans (2010) points out, the "increased ambiguity" of decreased sharing could make it more difficult to replicate experiments, and the lack of information

[48] For example: 1975, Howard Temin and David Baltimore, simultaneous but separate discovery of reverse transcriptase; 1976, Burton Richter and Samuel Ting, pioneering work in the discovery of a heavy elementary particle of a new kind; 2007, Oliver Smithies, Mario Capecchi and Martin Evans, for their discoveries of principles for introducing specific gene modifications in mice by the use of embryonic stem cells (http://nobelprize.org/).

available may even make attempts to do so produce different results (Evans, 2010, p. 761).

Scientists' concerns about increasing secrecy have inspired several recent initiatives which aim to counter this trend, which will be briefly explained here and examined in more detail in the Results section below. Science Commons was started in 2005 and aims to encourage data and biological materials sharing via licensing systems similar to Creative Commons, including standard materials sharing contracts and a clear system of rights. The Panton Principles (2010) are a statement about open data sharing in science, and aim to encourage a voluntary standard of data openness.[49]

Further evidence is contained in the growth of Web video, which has enabled such initiatives as the *Journal of Visualized Experiments* (JOVE), *SciTeeVee*, *SciTube* (YouTube for scientists), as well as the growth of method-sharing websites such as *Nature Protocols*, *Springer Protocols* and *Cold Spring Harbor Protocols*. Lastly, the use of Open Laboratory Notebooks is another attempt to reduce secrecy in science. However, as we shall see in the Results and Discussion section below, the growth in competition and commercialisation and attendant growth in secrecy has had an impact on the choice of research topics that OLN-using scientists study. These initiatives also involve scientists who take an active interest in openly sharing the tacit knowledge of the laboratory (Collins, 1974).

[49] "For science to effectively function, and for society to reap the full benefits from scientific endeavours, it is crucial that science data be made open" http://pantonprinciples.org/

4.4.4 Research on Science 2.0 and OLNs

As mentioned in Chapter Three (Section 3.4.5), the scientists who practise Open Notebook Science have achieved success at promoting their actions in popular scientific media, such as the journals *Nature* and *Science*. However, as a research phenomenon, the use of Web 2.0 and open laboratory notebooks has so far attracted limited attention in the sociology of science.

Several articles in field-specific journals have discussed the use of electronic notebooks, and open laboratory notebooks in particular, over the past two years. However, these should be considered part of this chapter's research data rather than specific research on the use of OLNs, as they are written by those internal to those fields and OLN use is discussed within the context of short examinations of the work being conducted by the research participants of this chapter, or are authored by the research participants of this chapter. Examples of these publications include: Nielsen's 2009 "Doing Science in the Open", in *Physics World*, which references the physics preprint server arXiv, GenBank, the *Journal of Visualized Experiments* (JOVE) and the OLNs of Jean-Claude Bradley and Cameron Neylon; and Sanderson's 2008 article "Data on Display" in *Nature*, which discusses Bradley and Neylon's OLN use.[50] Further papers of this nature will be cited where appropriate in the course of the Discussion section below.

---

[50] Nielsen: http://physicsworldarchive.iop.org/full/pwa-pdf/22/05/phwv22i05a38.pdf
Sanderson: http://www.nature.com/news/2008/080915/full/455273a.html

Most recently, Lievrouw (2010) enquired whether the use of social media technologies in science (defined as "wikis, blogs, tagging and bookmarking, conferencing, etc.") are signalling a return to Derek de Solla Price's "invisible colleges" (aka "little science"), as opposed to 20th century science dominated by wealthy institutions and professionalism (aka "big science") (Lievrouw, 2010, p. 219). Lievrouw suggests that the communication of science, whichever form it takes, relies on three distinct phases – conceptualisation, documentation and popularisation. These three phases are characterised in different ways for different types of scientific practice, with the shift to "telescience" changing relationships between the phases, particularly in accelerating the popularisation phase.

Lievrouw refers to open notebook science in her discussion of online collaboration and group work, listing both OpenWetWare and Jean-Claude Bradley's UsefulChem as prominent open notebook science sites. My research agrees that "informality, reciprocation, and interaction have become defining modes of scientific communication online" (Lievrouw, 2010, p. 231). However, this communication generally takes place within the adjunct tools of the OLNs, like FriendFeed and blogs, and not within the OLNs themselves, which are still strictly research notebooks, as will be shown below. The open notebooks still retain their status as the original records of experiment. The discussion that Lievrouw refers to, however, does take place, albeit in other online locations. One exception to this is that

there is sometimes discussion between the graduate student and their advisor within the notebook itself.[51]

The UK's Research Information Network (RIN)[52] conducted several case studies on Open Science in 2010. One case they investigated was Cameron Neylon's open laboratory notebook, LaBlog. They interviewed Cameron Neylon and Jeremy Frey as well as some staff from the Rutherford Appleton Laboratory where Neylon works. The case report is a description of how LaBlog works at the Rutherford Appleton Laboratory (of which more below in Section 4.7.7), as well as some commentary.[53]

The RIN report notes several benefits and disadvantages of open working, which include the benefits of perceived positive impacts on research efficiency (shortened research cycle, avoiding duplication of effort), improved research quality and scholarly rigour (input and criticism received from a wider range of peers), improved visibility inside and outside their fields, new collaborations and community-building (aided by

---

[51] A 2008 Scientific American article titled "Science 2.0: Great New Tool, or Great Risk?" is an experiment in what the author calls "networked journalism", in which a draft version of the article was published online and comment invited on any aspect of its content (Waldrop, 2008). Waldrop defines Science 2.0 as "how researchers are beginning to harness wikis, blogs and other Web 2.0 technologies as a potentially transformative way of doing science" (Waldrop, 2008, p. 1). This article highlights activities which will be further discussed in this and other thesis chapters, and also contains contributions from many of the scientists participating in this thesis research. For this reason it will be referenced further in the discussion section where necessary.

[52] http://www.rin.ac.uk/system/files/attachments/NESTA-RIN_Open_Science_V01_0.pdf

[53] Of interest is the fact that only one group from the six cases in the report chose to be anonymised, a group in the medical sciences who work with human subjects and therefore have many restrictions on data-sharing. This group was also the most sceptical about open ways of working. This finding echoes my own experiences as a researcher in this area, as none of the scientists I interviewed have elected to be anonymised, which is congruent with their strongly expressed opinions about Open Science and their Open Notebooks.

visibility and access to data for re-use) and a sense of increasing return on investment for research by enabling data re-use and protocol-sharing (NESTA and RIN, 2009).

Disadvantages cited by participants include the lack of solid evidence of benefits and rewards (which may discourage further uptake of open science), a lack of incentives, rewards and support for engaging in open research, restrictions on the time and skills available for building or enhancing their online research environments, concerns about the normative scientific culture of secrecy, and concerns about data misinterpretation (as normally data is not left open to interpretation but is filtered and made sense of first by the researcher and later, peer reviewers) and institutional, legal or ethical restraints on the release of data and materials (NESTA and RIN, 2009). These viewpoints will be further considered in the Discussion section of this chapter.

This literature review has considered the research on scientific work practices in the laboratory and their connection to laboratory notebooks. The review leads to the expectation that findings about open laboratory notebook scientists will include insights about technology adoption – that scientists do not expect their tools to be perfect, but will use them if they are a current best fit for their needs and/or have a low technical learning curve; that OLN users hope that their openness will improve the ability to replicate experiments where necessary, reduce unnecessary repetition and increase data sharing; and that they do not anticipate being scooped, either

because their openness establishes provenance or their research is unique. OLN users are publishing their live laboratory data in tandem with other scientific social networking tools to disseminate widely and form new collegial networks.

4.5 Open Laboratory Notebooks

Researchers using OLNs place all of the data from their laboratory experiments, protocols and results publicly online, thus revealing not only successful results but also the "dark data" of the laboratory (failed experiments or unpublished experimental data) (Goetz, 2007). The scientists involved usually follow a procedure based upon Bradley's definition of an Open Laboratory Notebook:

> There is a URL to a laboratory notebook […] that is freely available and indexed on common search engines. It does not necessarily have to look like a paper notebook but it is essential that all of the information available to the researchers to make their conclusions is equally available to the rest of the world. Basically, no insider information. (Bradley, 2006)[54]

The practice of publishing experimental data directly to the public Internet is unusual across the sciences, and the scientists practising it are few in number, most of whom are a subset of the FriendFeed users studied in the previous chapter, and active advocates of open science.

The next section introduces the research participants for this case and provides an outline of their involvement in open laboratory notebooks and open science. It also discusses why there is some overlap between subjects

[54] A shortened version of this quotation is the current definition on the Open Notebook Science Wikipedia page.

in the case studies of this thesis, as many OLN users are also active bloggers and FriendFeed users.

4.6 Research participants

Eleven scientists have been interviewed for this Open Laboratory Notebook case study. Although not exhaustive, this sample represents a significant proportion of life scientists currently practicing open notebook science. Members of this community, who find each other and establish communications using the Internet, maintain the Wikipedia page for this topic, "Open Notebook Science".[55] These contacts may be initially established via Internet searches, as scientists considering using one speculate whether anyone else is doing ONS, or become attracted to the idea after finding relevant open research or discussions within their discipline.

In 2005, when he first opened up his online laboratory notebook, Cameron Neylon knew that there were science blogs sustaining discussion on the Internet, but he was not sure if anyone else had open notebooks, so he went on to Google to check, which in turn led to his involvement with the open science community:

> One Google search later, a lot of people had thought of it [open notebooks]! But that search, I guess, in a sense, is what then led me into seeing various other people doing similar things, or talking about similar things online. (Neylon interview)

[55] http://en.wikipedia.org/wiki/Open_Notebook_Science

Contact is maintained via commenting on each other's blogs, and more recently, scientific social networks such as FriendFeed (detailed in Chapter Three). Personal email is less significant than public discussion in these networks; many conversations take place online, in FriendFeed, Twitter, or blog comments, as Lang remarks: "the group that I work with, if they have a thought, they'll write it down, or tweet it, or something. Even if it's silly, they'll just put it out there" (Lang interview). The openness of these discussions is relevant to the research participants, as it is through these casual comments and responses that they make new research connections or find new resources, as they feel confident using these connections: "if I come across a problem that I know they're [any online connections] an expert in - I know I could email them and they would reply" (Lang interview).

The scientists interviewed represent six laboratories. In each laboratory the principal investigator has been interviewed, and where available one or two graduate students or post-doctoral fellows who are using the system have also been interviewed. See Table 8, Open Notebook Science research participant details, below, for details of each research participant.

The Wikipedia page on Open Notebook Science helps with regard to assessing whether a representative sample of those practicing Open Notebook Science has been obtained. The page divides a list of individuals and groups into three categories: Active (with subcategories for Experimental and Theoretical notebooks), Archived and Partial/Pseudo (by

which they mean that there is openness, but not all data is immediately exposed, or necessarily ever exposed). Of the nine individuals named as Experimental ONS practitioners[56], I have interviewed five (Bradley, Neylon, Lang, Salvagno and Maloney). Three more from the Partial/Pseudo list of seven have also been interviewed (Koch, Johnson, Redfield) and three who are not on the list (Murray-Rust, Todd, Bulger). The three are either collaboration partners or students of those interviewed and are not considered to have their own open notebook, or they are in some way connected to this network and can offer insights into how some aspects of ONS have developed. The list is probably not fully comprehensive, as there may be others; but they are not connected to anyone in this network or have only expressed an intent to have an OLN, and have not yet started one.[57]

Research participants' experience ranges from undergraduate (1), current doctoral student (2), post-doctoral fellow (1) and mid-career or senior scientist (7). They represent a range of science fields and disciplines including biophysics, biochemistry, chemistry, and genomics.

Most of the OLN users are also current FriendFeed users, and some of them are central figures in the FriendFeed Life Scientists group (Cameron Neylon and Jean-Claude Bradley), and have therefore also been included in

---

[56] As of 15 February 2011, page last edited 8 February 2011. Snapshot from this date available at: http://en.wikipedia.org/w/index.php?title=Open_Notebook_Science&oldid=412733839

[57] For example, Joshua Shaevitz, head of an optical biophysics laboratory at Princeton, who has expressed an interest in using OLNs but wanted to get the full support of his laboratory members, as he says in a blog post from 2010: "Mostly, we haven't done it because not all the lab members are in favor of this move. We are still in active negotiations, but I hope to have the issue resolved soon" (Shaevitz, 2010).

the Chapter Three FriendFeed case study. Several OLN users are also active science bloggers (Redfield, Murray-Rust, Neylon, Bradley, Koch, Todd) and indeed blogging has been a key component of their open science practices, either stimulated by the feedback received from having an open notebook, or facilitating connections with others which eventually led them into contact with the open science movement. This chapter will not go into detail about their blogging activities, but will quote from blog and FriendFeed discussions relevant to the use of OLNs. The next case study chapter on scientific blogging will detail the motivations for blogging and impacts on research practices for scientific bloggers.

Table 8

Open Notebook Science research participant details (n = 11)

Name	Current Position	Laboratory	Location	Education
Cameron Neylon	Senior Scientist at Science and Technology Facilities Council	Neylon	Rutherford Labs, U.K.	PhD in biochemistry, Australian National University, 2000
Jean-Claude Bradley	Associate Professor, Department of Chemistry	Bradley	Drexel University, PA, USA	PhD University of Ottawa, 1993
Steve Koch	Assistant Professor of Biophysics	Koch	University of New Mexico, USA	PhD Cornell University, 2003
Anthony Salvagno	PhD Student	Koch	University of New Mexico, USA	Undergraduate degree in physics, Columbia University
Andy Maloney	PhD Student	Koch	University of New Mexico, USA	Undergraduate degree in physics, UT Austin
Peter Murray-Rust	Reader in Molecular Informatics	N/A	Cambridge University, U.K.	DPhil Chemistry, University of Oxford
Mat Todd	Senior Lecturer in Chemistry	Todd	Sydney University, Australia	PhD Organic Chemistry, Cambridge University
Andrew S.I.D. Lang	Professor of Mathematics	Bradley	Oral Roberts University, Oklahoma	PhD Mathematics, U Missouri, 1998
David Bulger	Undergraduate science student	Bradley	Oral Roberts University, Oklahoma	Undergraduate; accepted into medical school 2012
Martin Johnson	Postdoctoral researcher, marine and atmospheric chemistry	Johnson	University of East Anglia, UK	PhD Chemistry, University of East Anglia

Name	Current Position	Laboratory	Location	Education
Rosie Redfield	Professor of Cell and Developmental Biology	Redfield	University of British Columbia, Canada	PhD Stanford

4.7 Inside the Open Laboratory Notebooks

Electronic laboratory notebooks are used as tools by scientists for many reasons, including but not limited to creating a record of their laboratory processes as they are produced, being able to reproduce their research by having the level of recorded detail that this requires (often not all details necessary are recorded in a traditional written laboratory notebook due to time constraints or distributed nature of the data available), sharing access-controlled data with collaborators, creating records that are accessible from multiple places (for ease of work), and making data searchable and in some cases, partially machine-readable.[58]

Online and open electronic laboratory notebooks support and extend these features, and according to their users, extend the practices described above in ways that benefit their science. The "extension features" cited by participants who use OLNs include eliminating the possibility of being scooped—if data is being extracted directly from the bench to the public Internet, there is a clear precedence of the production of this work, particularly as not only the data but also the experimental design and progress are all publicly recorded. They also report that sharing data is easy

[58] For the purposes of this thesis, machine-readability is defined on two levels: firstly, text that is human-readable but can also be indexed by search engines for full-text search (most OLN data) and secondly, semantically-marked-up text which is not only human-readable and search-indexable, but also tagged with extra machine-readable data which will enable more sophisticated uses. Examples of this include automatic linking to gene sequences in specified databases, automatic data citation in papers, linking to 3D models in other software (such as Second Life), video markup and transcripts that enable specific parts of a video to be linked to from text. In general, even the non-semantically marked up text and data have a different kind of affordance (full-text searchability) to text printed on paper.

as there is no need to tunnel through firewalls or worry about data security, as they can "just Google it". OLN users also claim that they and their students gain in terms of exposure to more scientists in a wide range of disciplines, which enables new collaborations and useful contacts in the scientific network.

Although many scientists using traditional practices regard using an OLN as risky due to the perception that OLNs could allow rival scientists to "scoop" their results and publish more quickly, or that they devalue their own intellectual property by not patenting their work, those who practice Open Notebook Science are quick to rebut these arguments and point out the advantages they have experienced from making public their data and practices. Positive and negative claims from interviews with those using OLNs for information-sharing, collaboration and data dissemination will be assessed below.

As noted in the Research Participants section above, few scientists currently practice Open Notebook Science, although this model is receiving considerable attention in popular scientific media (See Section 4.4.4, Research on Science 2.0 and OLNs, and Appendix G). Practitioners argue that opening up their notebooks enables less unnecessary experimental repetition and speeds up scientific knowledge production and communication, and that it can enable further and better scientific interaction with the public, claims which will also be examined in this chapter. Those most active in promoting OLNs can in any case be seen as

members of a Scientific/Intellectual Movement, sharing a similar agenda of promoting openness in all aspects of their scientific work. These OLN users and promoters also partially overlap with the scientific social networkers and bloggers comprising the other two case studies of this thesis. The overlap between these activities has been described above in Section 4.6, Research Participants.

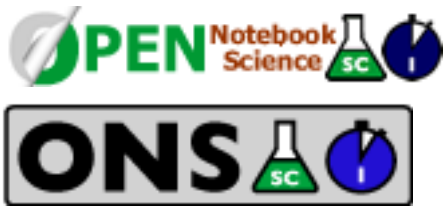
Electronic Laboratory Notebooks are designed to make it possible for bench scientists to record all of their experiments, protocols and materials in electronic format. This may or may not include some level of experimental material stocktaking or automated data entry from machines. Open Laboratory Notebooks also add the features of openness, that is, the notebook should be publicly viewable and findable via an Internet search engine. A third reason is semantic web integration—bringing together machine and human-readable data. This goal is of varying importance to participants, as will be seen in the descriptions below. The experimental recording should ideally be done in a way that facilitates reproduction of the experiments and encourages citation of the data.

Table 9

Open Notebook Science Claims Logos

Designed by: Anthony Salvagno, Shirley Wu

http://onsclaims.wikispaces.com/

Logo example	Open Notebook Science Claim
	ACI All Content - Immediate The entirety of the lab notebook and associated supporting raw data are available to the public in as close to real time as possible. If it isn't in the notebook others can assume that you haven't done it.
	ACD All Content - Delayed The entirety of the lab notebook and associated supporting raw data are available but after a significant delay - perhaps for patenting or publication reasons.
	SCI Selected Content - Immediate A portion of the lab notebook and associated supporting raw data are available in as close to real time as possible. Others cannot assume that if it isn't in the notebook you haven't done it.
	SCD Selected Content - Delayed A portion of the lab notebook and associated supporting raw data are available after some delay. Others cannot assume that if it isn't in the notebook you haven't done it.

There are varying levels of OLN openness, from fully open, with all data shared as close to real-time as possible, to some content shared, but not immediately, referenced above in Table 9. Levels of openness may vary

from project to project, depending upon the collaborations in place. For example, Cameron Neylon, who runs an OLN in his laboratory at the Rutherford Appleton STFC laboratories, posted in his blog that he had produced some experimental data which could not go in his OLN, describing the situation in these terms:

> We did one very cool and simple experiment, one that worked, and one that actually has some potentially really interesting implications. The only things [*sic*] is … I can't tell you about it. It's not my project and the result is potentially patentable. At one level this is great ... but as a person who believes in the value of sharing results and information as fast as is practical it is frustrating. (Neylon, 2010a)

Both the ability to record data and openness can be facilitated by various online platforms. Wikis and blogs are popular, either built from scratch or appropriated for this use. The sections below describe the types of platforms in use by several Open Notebook scientists. The platform is important because it has implications for methods of data archiving, retrieval, citation and publication.

As will be shown below, these OLN's, even when purpose-built, do not make the science *easy*, nor can they necessarily change the practices of those individuals who would have kept a sparse paper notebook. Other difficulties (some shared with paper notebooks) include such things as building data tables, which can be time-consuming; working out backup and archive strategies and figuring out how to embed data from other sources such as video and spreadsheets. However, the electronic notebooks are capable of displaying a far wider range of data than is the case with

paper-only notebooks, and allowing dissemination of this knowledge to many more people, although it should also be acknowledged that erroneous data could spread this way—but again, this problem is shared with paper publications, and errors can be more easily rectified online, before the data is rendered into a publication ready for peer review. For example, Bradley found old chemical measurement errors in publications and corrected them with data in his OLN, which then made the newer and more accurate data findable via Google.

4.7.1 UsefulChem - wiki-based OLN

Associate Professor Jean-Claude Bradley's group in the Department of Chemistry at Drexel University, Philadelphia, conducts research into synthesising novel antimalarial agents, for which the initial need was identified via a Google search (Bradley, 2007). Bradley's group use a free public wiki (Wikispaces) to publish their open laboratory notebook. Users must create an account and login to edit a wiki page, and pages are usually protected from editing by the creator. However, all pages of the wiki are publicly viewable. The notebook is called UsefulChem.[59]

Bradley's system has been in use since 2005.

> In 2005 ... I asked myself, "if you were going to be completely transparent in doing chemistry research, what would that look like?" So I started a blog ... where I chronicle how I found the various projects and how we met all our collaborators and all of this, but the idea was basically simple—to just do science transparently.

[59] UsefulChem is at: http://usefulchem.wikispaces.com

> I wasn't clear as to the tools I would use, necessarily, at that point, but I just knew that whatever it was it shouldn't require a username and password to access. So it just started like that and then it evolved to what it is today. (Bradley interview)

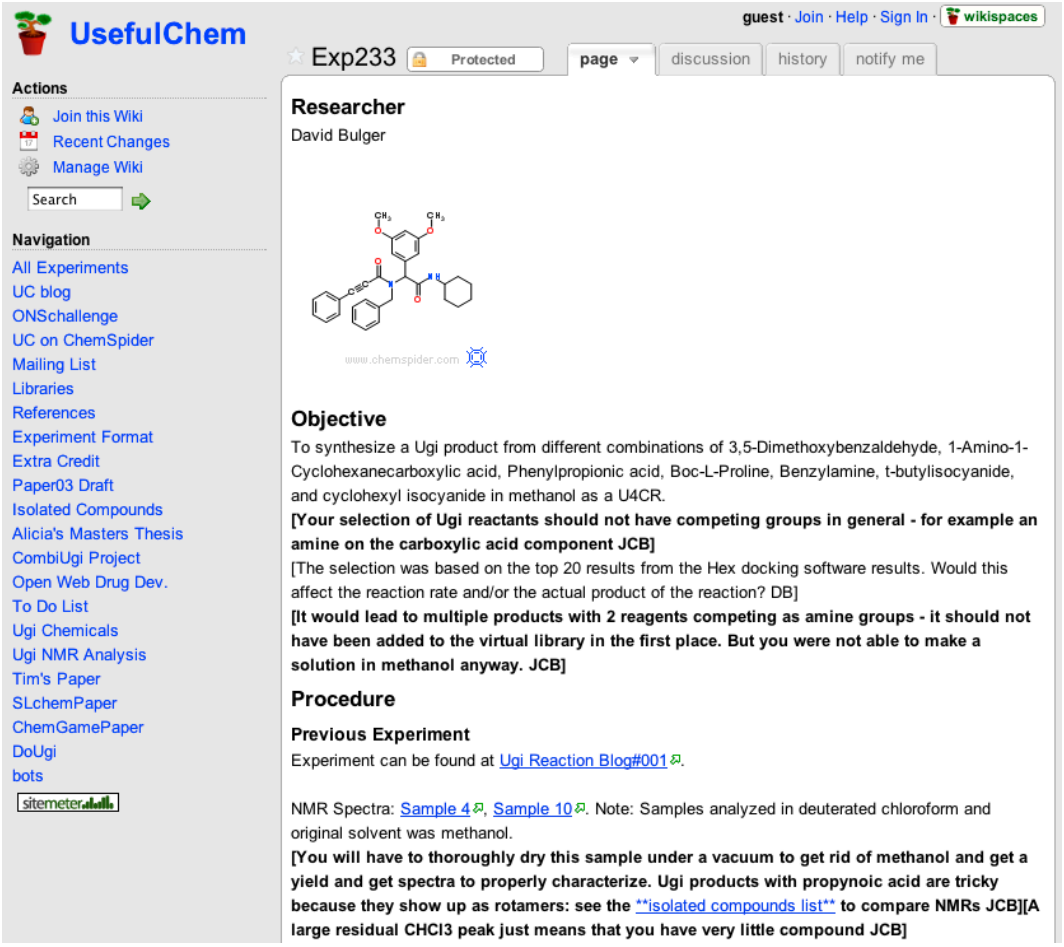


Figure 20. Screenshot of an experiment on the UsefulChem wiki (shows Bradley's comments in bold).[60]

Wikispaces has the features common to many wikis including edit version tracking, the ability to extensively hyperlink pages within and outside the wiki, the ability to embed data from other online sources (such as YouTube videos or Google Documents spreadsheets) and built-in alerting mechanisms that can provide RSS or email alerts to subscribers

[60] http://usefulchem.wikispaces.com/Exp233

when changes are made in pages, discussions or files on a page or across the whole wiki.

Bradley has an informal approach to communication and discussion within notebooks. He makes his comments in bold, and students can respond in the notebook or talk to him (since they are normally co-located) about the issues, then corrections can be made. Bradley's main objective is "to have a communication going and to have things be fixed and be improved over time", which in practice means that the system is constantly evolving (Bradley interview).

Bradley's OLN started as a blog, with one blog post per experiment, but moved to the wiki, which automatically tracks edits to a page, as he found that the blog format was not amenable to corrections and discussion:

> What I discovered fairly quickly is that really didn't work well, because there were so many revisions to make, there was so much error correction, and although you can edit a blog post, it doesn't keep track of all the various edits. (Bradley interview)

Editing seamlessly while recording all changes is a key feature for Bradley, because "there's always something that we find out later on, and so having that ability to constantly change the lab notebook, or add to it, or tend to it, is pretty important" (Bradley interview).

The ability to archive wiki data to online storage or a local server or network folder is not built in. To enable backup for archival purposes, Bradley and his collaborators have created a "home-grown" archiving solution which starts with a series of steps using Google spreadsheets' ability to export to Excel, uses a few scripts written by contributors to the

project, and ends with daily Excel backups of their data on a Windows server.[61]

Bradley has a history of being involved and interested in laboratory software development. In earlier years he patented the processes behind a Web-based (but private) software system he designed and implemented, called SMIRP - Standard Modular Integrated Research Protocols (Bradley & Samuel, 2000). "The idea is that you would make the components of the experiment into modules, and then you could basically connect the modules up together ... and put them together in the form of a human-recognisable paper" (Bradley interview). However, Bradley later realised that SMIRP was not open enough for the open science approach he envisioned: "there was so much transparency missing, because you had to go through so many steps to make information public" (Bradley interview). SMIRP, although more structured and focused on laboratory processes, did not facilitate the kinds of interactions Bradley wished to promote by being open. As will be seen below, activities such as the Open Notebook Science Challenge were more successful in connecting scientists to Bradley's scientific and open science work.

4.7.2 Open Notebook Science Challenge

An important development of Bradley's OLN has been the growth of the Open Notebook Science Challenge. The Challenge uses crowdsourced

[61] See http://usefulchem.blogspot.com/2009/04/automatic-back-up-of-ons-files-google.html

research—that is, any individual or group with access to the appropriate laboratory facilities can participate—to collect measures of the solubility of non-aqueous organic compounds and share the resulting data openly. This solubility data is useful as groundwork for synthesising the target drugs of the UsefulChem project.

Eligible students are motivated to participate with various prizes - cash prizes of $500 USD, or magazine subscriptions, sponsored in the past by SubMeta, the Royal Society of Chemistry and Nature (ONSChallenge wiki, 2011). Part of the focus of the Challenge is encouraging students to learn how to keep good open laboratory notebooks, and to extend OLN practices to a wider group. The rules of the challenge reveal this clearly, firstly by displaying a simple metric for a good OLN:

> We are looking to reward students who show promise of becoming good scientists. This includes maintaining a laboratory notebook (including links to raw data) in a manner that makes it easy for others to understand exactly what you did and what you observed. The log portion of the notebook must be recorded in a timely fashion, usually within 24 hours of performing the experiment;

secondly, by acknowledging that laboratory notebooks, and OLNs, are a learned skill;

> Nobody maintains a notebook perfectly from the start and that is why the organizers will be providing feedback in BOLD directly on your notebook pages. How quickly and thoroughly you address that feedback is an important criterion;

and lastly, by encouraging the students to become part of the scientific social network around open science:

> Creativity in the design of the solubility measurement technique in terms of efficiency, speed and reproducibility will also play a strong role in the evaluation. Participation in the scientific community via Web 2.0 networks would certainly be a plus. (ONSChallenge wiki, 2010)

Judges of the competition include recruits from Bradley's own network—Neylon, Hooker, Lang and Koch among them. Hooker spoke of the enjoyment of being a judge:

> Everything his [Bradley's] students do is online, in more or less real time, and if you have the interest you can dig down into it. And that's what you do when you're judging this contest, you look at the experimental log, and you try to ask questions and point the students in whatever direction you think they should be thinking, and it's absolutely just fascinating, and a lot of fun. (Hooker interview)

There is evidence from the wiki's weblogs that the solubility data from the challenge is being found, as Bradley says: "we have about 100 queries a day [on the solubility data], and they're not from our institution" (Bradley interview). This is the only evidence available that the solubility data is being used, although Bradley states that due to the small data points (a single figure would often be the only point used), he would not expect citations: "if they're trying to optimise a reaction, maybe they'll look up a specific solubility measurement, but they might not think that's significant enough to cite in a paper. So I'm not really looking for that, not really expecting that kind of outcome" (Bradley interview).

David Bulger was an undergraduate chemistry student at Oral Roberts University, Oklahoma. Andrew Lang, one of Bradley's collaborators on UsefulChem, noted that both Bradley and Neylon (whose OLN is detailed

below) contributed their time and expertise to assist Bulger's OLN research in their laboratories. Practicing ONS as an undergraduate, Bulger has made much of his open science connections—he has visited the Bradley and Neylon laboratories, and won the February 2009 ONS challenge. Bulger's remarks about his use of OLNs reflect the goals of the challenge: "with collaborators viewing the notebook before, during, and after the experiment, I'm motivated to be disciplined in recording everything in the open notebook" (Bulger interview). Bulger graduated with honours from ORU, and is now a medical student in Tennessee. As will be seen in the next section, Bradley and Lang continue to collaborate together.

4.7.3 Interdisciplinary collaboration via OLN

Several scientific collaborations have been sparked by connections made through FriendFeed and blogs, and the work itself has been done on the OLN. The following example is drawn from Jean-Claude Bradley's collaborators.

Andrew Lang is a Professor of Mathematics at Oral Roberts University in Oklahoma. Lang's initial connection with Bradley was via a common interest in teaching chemistry via Second Life (SL).[62] Bradley and Lang found themselves SL neighbours after acquiring plots of virtual land on *Nature's* island for their experiments around building 3D molecular and chemical models. They collaborated for eighteen months in SL, which culminated in a paper at a conference on teaching and learning chemistry in

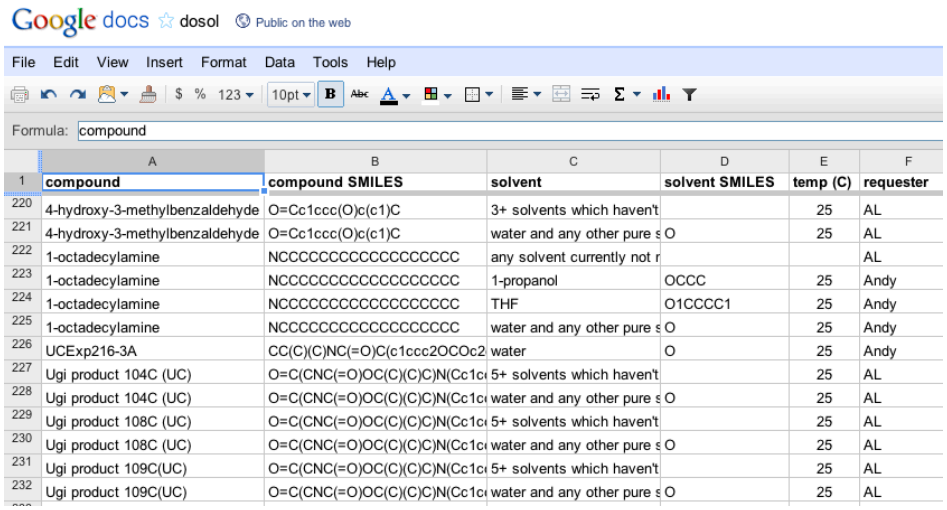
[62] Second Life is the largest commercial virtual world. http://secondlife.com/

SL, where they actually met in person for the first time, rather than as avatars in SL.

Most recently Lang has been a contributor to Bradley's solubility challenge competition by creating interfaces between data and Web and improving the search engine rankings of the OLN so its data can be found. "It's mainly on the computer and on paper and pencil ... I have my little section of the wiki where I post my stuff ... I put up what I do in real time too" (Lang interview).

In his interview, Lang elaborated on the public/private dichotomy of UsefulChem. One of the features of the Open Science Challenge is a form where anyone can request an experiment. This form feeds into a Google Docs spreadsheet, named *dosol* (for Do Solubility), which Bradley reviews when a new request is received, to check that the requested experiment will not be dangerous for the students to perform, and then assigns to a queue. Requests for experiments come from a variety of sources, many of them students, some in Bradley's laboratory, others from laboratories around the world. Some requests are received from chemists working in commercial concerns, and these are treated just like other requests, and as Lang states, the data produced will always be open:

> If we collaborate with somebody who is closed, our collaboration is still open. So you actually see [our] half of the [project]. And so our data that we produce is open, and anybody in the world can use it, because it's out there. It's just that we don't necessarily give the whole story about how our collaboration with that closed institution came about. (Lang interview)



	A	B	C	D	E	F
1	compound	compound SMILES	solvent	solvent SMILES	temp (C)	requester
220	4-hydroxy-3-methylbenzaldehyde	O=Cc1ccc(O)c(c1)C	3+ solvents which haven't		25	AL
221	4-hydroxy-3-methylbenzaldehyde	O=Cc1ccc(O)c(c1)C	water and any other pure s	O	25	AL
222	1-octadecylamine	NCCCCCCCCCCCCCCCCCC	any solvent currently not r			AL
223	1-octadecylamine	NCCCCCCCCCCCCCCCCCC	1-propanol	OCCC	25	Andy
224	1-octadecylamine	NCCCCCCCCCCCCCCCCCC	THF	O1CCCC1	25	Andy
225	1-octadecylamine	NCCCCCCCCCCCCCCCCCC	water and any other pure s	O	25	Andy
226	UCExp216-3A	CC(C)(C)NC(=O)C(c1ccc2OCOc2	water	O	25	Andy
227	Ugi product 104C (UC)	O=C(CNC(=O)OC(C)(C)C)N(Cc1c	5+ solvents which haven't		25	AL
228	Ugi product 104C (UC)	O=C(CNC(=O)OC(C)(C)C)N(Cc1c	water and any other pure s	O	25	AL
229	Ugi product 108C (UC)	O=C(CNC(=O)OC(C)(C)C)N(Cc1c	5+ solvents which haven't		25	AL
230	Ugi product 108C (UC)	O=C(CNC(=O)OC(C)(C)C)N(Cc1c	water and any other pure s	O	25	AL
231	Ugi product 109C(UC)	O=C(CNC(=O)OC(C)(C)C)N(Cc1c	5+ solvents which haven't		25	AL
232	Ugi product 109C(UC)	O=C(CNC(=O)OC(C)(C)C)N(Cc1c	water and any other pure s	O	25	AL

Figure 21. The *dosol* experiment request spreadsheet on Google Docs[63]

UsefulChem participants seem to have no issues with straddling this public/private divide. However, they do respect the privacy wishes of those they work with, as Lang states: "there are occasions when we get contacted by a pharmaceutical company ... that is heavily into secrecy ... and we don't really feel that we could make those conversations public. So it's kind of interesting because we're open but ... on the outside of our group we interact with people who are not so open"(Lang interview).

Lang also feels that another form of their data and communications, the emails which they exchange to formulate ideas and work out project details, should perhaps be part of the scientific record: "that's a possibility in the future, is actually not only publish the research, but actually publish the communication as well, I think that would be very valuable, for someone looking back. Where did this idea come from, they could actually trace it to the email conversation. Not that our emails are that interesting! Part of the

[63] *dosol* is at: https://docs.google.com/spreadsheet/ccc?key=0Ak1R8T6wt4YQckktTmYyYWlIbFktem9IRWRHd2ppVmc&hl=en#gid=0

scientific record, really I think" (Lang interview). This viewpoint is interesting to compare with the reactions from climate scientists at the University of East Anglia after the theft and release of thousands of their emails, known colloquially as "*Climategate*" and discussed in detail in the UK Parliament in a subsequent enquiry, which criticised the academics involved for not sharing their data, in particular noting that the Climate Research Unit at the UEA "could have been more open at that time in providing the detailed methodological working on its website. We recommend that all publicly funded research groups consider whether they are being as open as they can be, and ought to be, with the details of their methodologies" (Secretary of State for Energy and Climate Change, 2010, p. 5). The report also recommended that data funded by public research should be made available whenever possible. The resistance of the UEA scientists to share data even after receiving valid Freedom of Information requests for it contrasts sharply with Lang's opinions, and highlights why many scientists resist sharing publicly-funded datasets—sharing is still a contentious, non-normative practice in many areas of science.

4.7.4 Synaptic leap and open chemistry

Mat Todd is an organic chemist at the University of Sydney, using OLNs for several projects around open source drug discovery. Todd recently received funding for a project to synthesise synthetic praziquantel, a drug successfully used to treat the disease schistosomiasis. Better synthetic compounds would substantially lower the costs of the drug, enabling

treatment of many more people. What engaged him was the "serious price constraint on the whole situation ... if you gave the problem to an academic chemist, they would start sketching stuff on bits of paper and then they would come up with a hundred different solutions to this problem, but none of them will be any good, because the cost of the process will be prohibitive" (Todd interview). Todd's challenge is to develop a compound within those price constraints, which he decided to do openly.

Todd runs his projects through two interfaces—the Synaptic Leap blog, which contains most of the ideas and discussion around the praziquantel projects (Figure 23), and a LabTrove blog-based OLN for experimental data (Figure 24). The Synaptic Leap non-profit organisation and set of projects was started in 2006 by Ginger Taylor, a software manager, in order to "provide a network of online research communities that connect and enable open source biomedical research" (Taylor, 2006).

The Synaptic Leap has project communities built around schistosomiasis, malaria, toxoplasmosis, and tuberculosis. Todd was one of the early users of the site, and still uses it extensively, as well as serving as chairman of its board of directors.

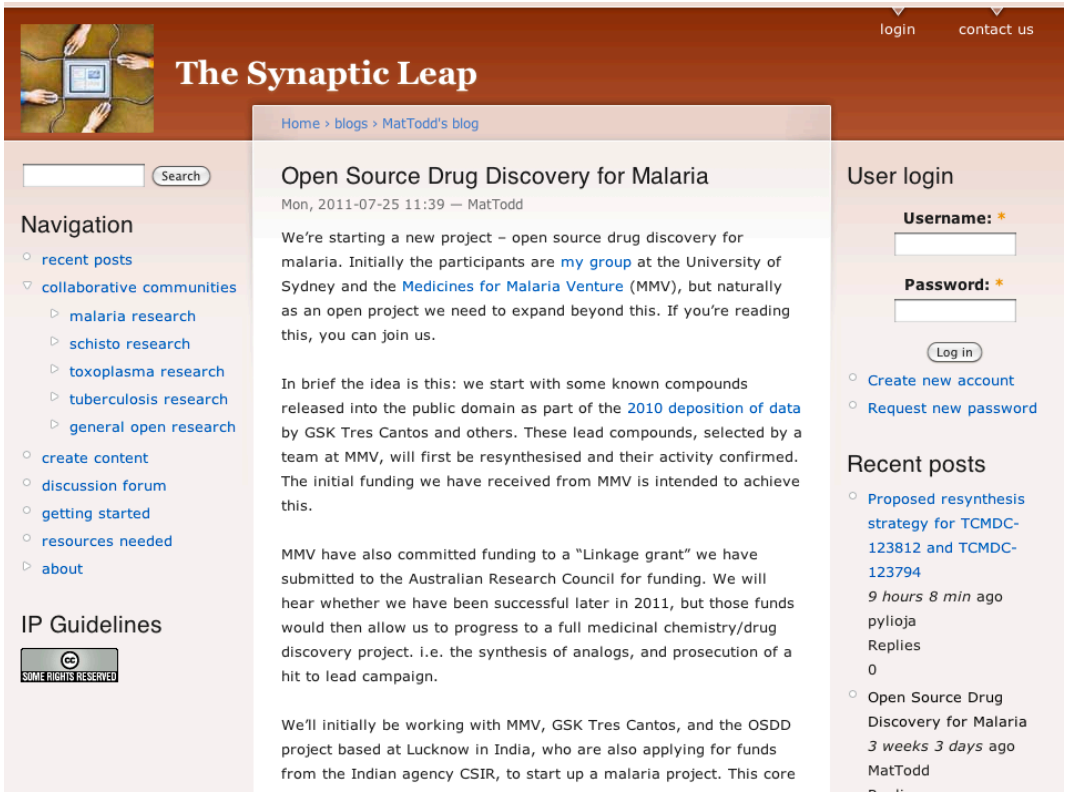


Figure 22. The Synaptic Leap OLN blog

The Synaptic Leap is "just a simple blogging site ... we posted the problem on the web and so we sort of thought things would then start happening, because the project was public and I started talking about it, though we discovered the obvious thing, which everyone who's ever done any computer software and open source would know, would tell you, which is nothing happens if you just post a problem and expect people to help you out" (Todd interview). Todd decided that funding was important to get the project going, and successfully applied for funding from the World Health Organisation (WHO) and the Australian Research Council (ARC), which was awarded. Various clauses in the contract being argued between the WHO and the ARC (which had nothing to do with the project being open, neither funder had any issues with the openness) meant that it

was 13 months before work could commence on the project, in January 2010.

Todd knows Cameron Neylon from his online open science network. He remembered Neylon's work at Southampton, and that there was a blog-based OLN system in development, now called LabTrove. For his projects he needed an open-source solution so that others could use it without incurring costs: "it looked pretty basic, but then I was just looking around at other possibilities and I realised that there was actually nothing else that we really could use. Mainly because we needed a system where anyone in the future who might be interested in being part of the project could interact with the data, but also maybe even join the project" (Todd interview).

Todd remarked of LabTrove that "it takes a while to get used to it and it takes a while to post the first few times... it just has the basic elements of loading text and some data and some pictures. It's massively better than anything else we could have used, because everything else was closed and expensive, but it is – it needs to be much more intuitive" (Todd interview). However, LabTrove is in active development, and its developers are responsive to Todd's project needs, as he notes "whenever something comes up that's problematic, we e-mail them and that might be once every couple of weeks or something and they respond very quickly" (Todd interview). Todd's instances of LabTrove will eventually be hosted on the University of Sydney's servers, and he is also actively assisting with its future development by securing some funding for software development

along with Jeremy Frey's development team at the University of Southampton.

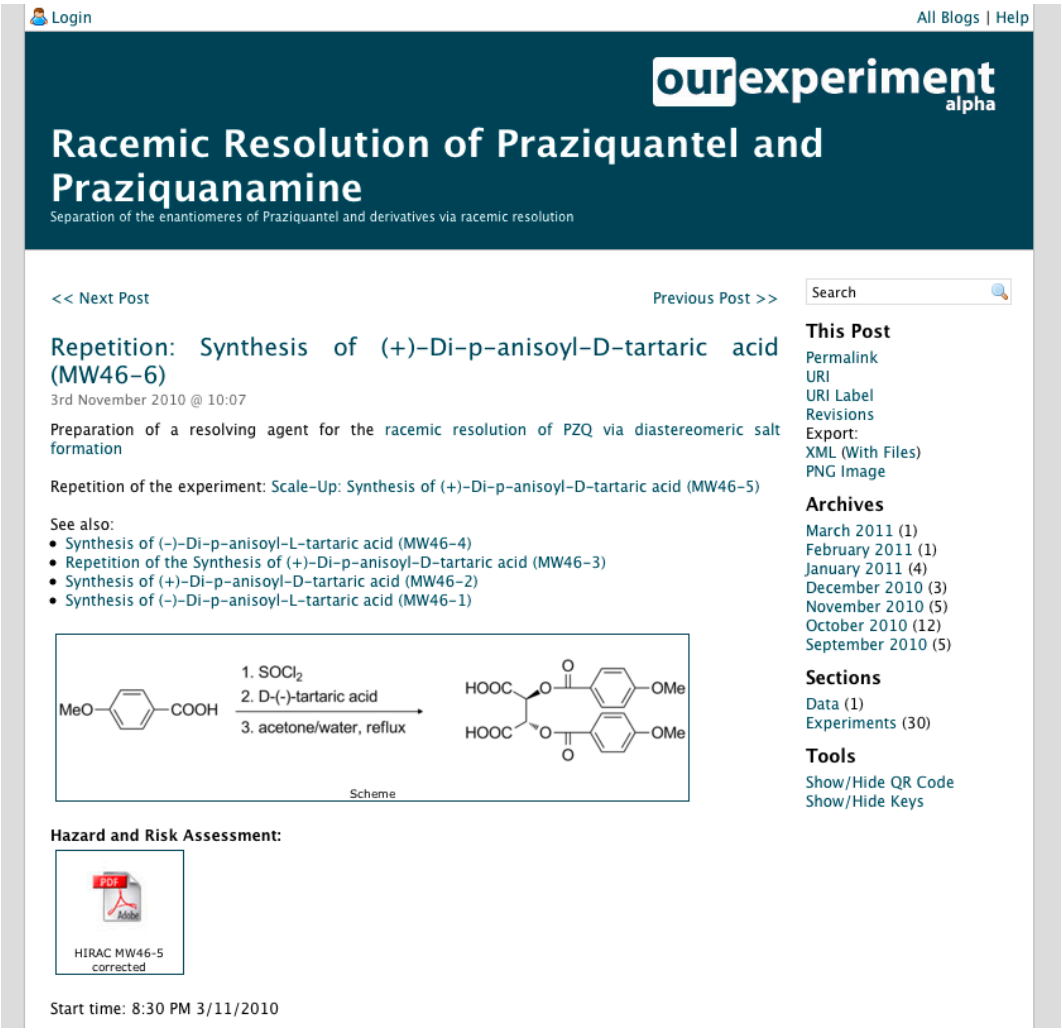


Figure 23. Mat Todd Blog-based OLN[64]

Within the Todd laboratory, students are given the option to go open or not: "I've basically given my students the options. So I explicitly said to all of them if you want at any point to take your project open, then let's do it. And so far, none of them have done that" (Todd interview). Todd

[64] http://www.ourexperiment.org/racres_pzq/1185/Repetition_Synthesis_of_DipanisoylDtartaric_acid_MW466.html
Based on the Southampton Labtrove project (http://www.labtrove.org/)

speculated that the students worried about the possibility of being scooped (discussed in detail above in Section 4.4.3), and that they were simply too busy to take the time to set up their notebooks in the face of as-yet-unproven benefits.

However, a post-doctoral fellow in Todd's laboratory has done some open work and it has shown clear benefits, and Todd is optimistic that more students will want to be open in the future: "having seen the acceleration of his research that happened when we went open with it, and having seen the productivity gains from it we got from an electronic lab notebook, then I think I can probably take that to my other grad students and say look, this actually really helps your research if you do this" (Todd interview). Most recently Todd has started a contest to involve students from around the world in his open research, similar to Bradley's ONS challenge described above. This part of the project was launched in June 2011, and at time of writing in September 2011 it had not yet been taken up by any students.

The reactions of Todd's colleagues and peers who are not involved in open science have been mixed: "people to my face, I think, would be quite optimistic about it, say that's a great idea and all this kind of stuff. But certainly no one ... would contribute ... I guess there's a distinction between people's opinions when they hear about it, when they talk about it versus people's responses in terms of active participation" (Todd interview). Todd quickly realised that people needed guidance and encouragement to contribute: "I think at the start we just assumed that people would

contribute ... out of the goodness of their hearts, but ... we've realised that you really have to go after it, hold people by the hand and help people a lot and show them how they can contribute" (Todd interview). The ability to contribute, for many academics, requires some training—as Todd remarked in interview, many academics are not particularly technology-savvy, and some require guidance to create blog posts or use a Wiki (although, once learned, these skills do not seem difficult).

Funding is crucial to Todd's vision of open science. In a comment on FriendFeed he explains how he thinks specific funding assists open science projects: "I have one open project recently underway on the synaptic leap that is fully government funded, and I have just applied with four collaborators for a larger project in a hot area of basic synthetic organic chemistry that would also be completely open ... Obtaining a grant to fund the kernel of activity that would come from your lab and the labs of your direct collaborators would mean you can develop some initial momentum" (Todd, 2010). Out of this kernel of activity, Todd and his collaborators have published two relevant papers: one in *PLoS Neglected Tropical Diseases*, detailing their Praziquantel synthesis project; and the other in *Nature Chemistry*, with a description of the open science approach employed on the paper, which also made the cover of the journal, shown below.

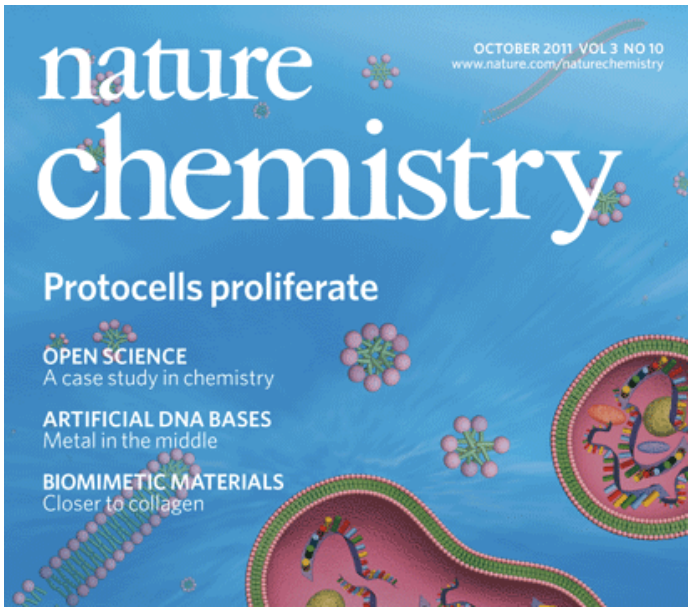


Figure 24. Cover of Nature Chemistry, October 2011, v3(10)

Todd strongly believes that open science gives scientists a competitive edge, in that working openly allows high-quality research to be produced faster. "The competitive advantage (rather than all the other obvious advantages one could list) is that open science is faster than closed lab science, which is good for everyone" (Todd, 2010), an opinion also shared and expressed by several other OLN-using scientists in this study, as we shall see below.

4.7.5 The Koch Laboratory wiki-based OLN

The Koch research group at the University of New Mexico are experimental biophysicists, examining various biophysics problems, with a particular interest in molecular motors. Koch and his group also develop a lot of testing and experimental equipment in-house, such as optical tweezers for corralling and observing bacterial molecular motors. The Koch group uses OpenWetWare (OWW), a free non-commercial electronic

notebook solution based on Mediawiki (which is the same software used by Wikipedia) which was created and launched in 2005 by MIT graduate students.[65] OpenWetWare hosts several hundred laboratories from more than one hundred institutions worldwide and many more personal pages for scientists and other academics.[66] However, as far as can be ascertained, the Kochlab (along with Todd's OWW paper-construction wiki pages mentioned above) has the only fully public OWW notebooks, as other groups and laboratories appear to keep most of their data behind logins, or not in OWW.

[65] http://openwetware.org/index.php?title=OpenWetWare:FAQ/Access&oldid=242622 Accessed 02 November 2010.

[66] http://openwetware.org/index.php?title=Labs&oldid=469423 Accessed 02 November 2010.



Figure 25. OpenWetWare home page.

Anyone may view content on OWW, but only registered users may edit content. Although any registered user may edit text on OWW, the site's users have not so far experienced any significant page vandalism, and judge the risk to be low at this time. The risk is reduced by the site's design, as users cannot comment anonymously but must register and show which research institution they belong to before commenting, which inhibits antisocial activities on OWW pages (Waldrop, 2008). OpenWetWare's FAQ on access to pages outlines why this approach has been adopted, and states

that the aim of these policies is to "achieve as large and active a user base as possible while maintaining a collaborative environment".[67]

The Koch laboratory has used OWW since Steve Koch joined the University of New Mexico in 2006. Initially his laboratory was hosted on a private version of the OWW wiki, but as his students became comfortable with the idea of being open, they gradually moved to completely open notebooks. Data is mostly input manually, and comments are made by Steve Koch using a comment template (see Figure 27 below). Koch laboratory members use several different applications to create and edit data (Google Docs and Spreadsheets, Evernote, graphs and other images (animated and still), FriendFeed pages, and YouTube videos), and tend to embed these other data sources into their pages while documenting experiments. Koch explains:

> We're still struggling with what's the best and easiest way to capture information in the lab, and do open notebooks. Because editing the wiki remains a little bit slow, ... the students are trying out Evernote or Google docs, a whole slew of other things, but we're still using OWW as a hub to embed those things, but with information coming in from different sources. (Koch interview)

[67] http://openwetware.org/wiki/OpenWetWare:FAQ/Access, Accessed 02 November 2010.

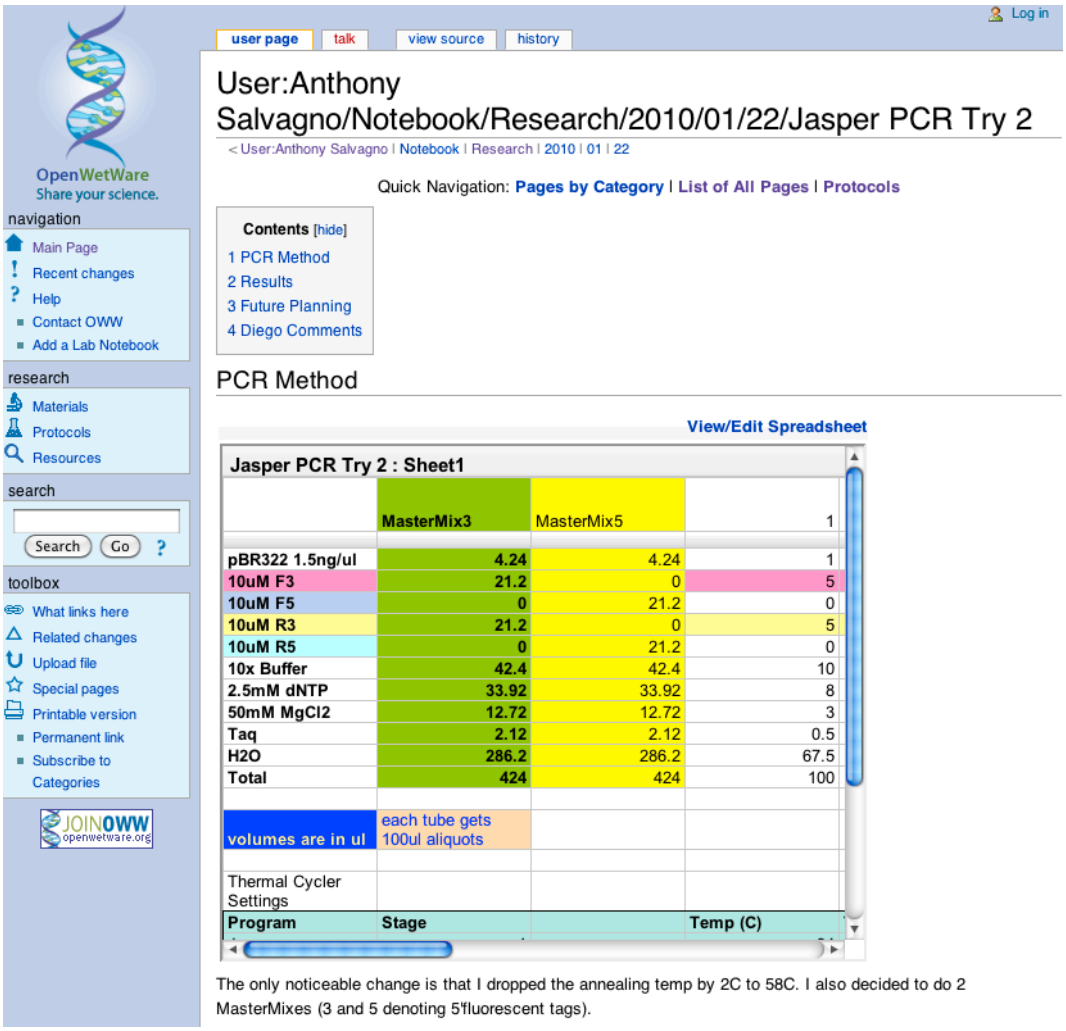


Figure 26. Kochlab Open WetWare OLN experiment page, showing embedded Google spreadsheet.

The Koch laboratory, as of late 2010, has three graduate students with open notebooks, as well as many undergraduate physics coursework students who create open notebooks for their laboratory classes, which are also commented on by Koch. These are all publicly viewable laboratory notebooks, some of which Koch has highlighted as good examples in FriendFeed discussions on open laboratory notebooks.[68]

[68] http://friendfeed.com/stevekoch/0b06a31a/another-excellent-primary-lab-notebook-from

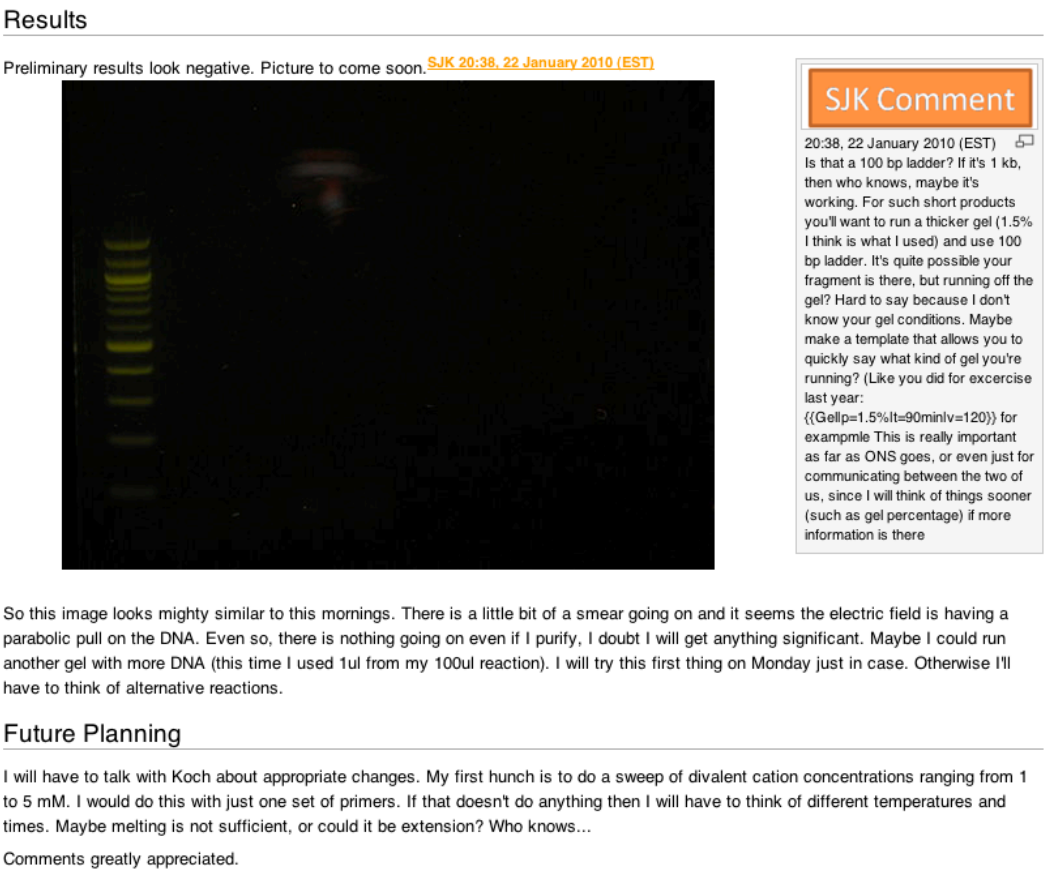


Figure 27. Results section of a sample Kochlab OLN page, showing discussion, a Koch comment and an embedded results gel image.

Two of Koch's graduate students, Andy Maloney and Anthony Salvagno, were interviewed for this research. Salvagno was the first to "go open" in the laboratory. Although Koch had discussed open notebooks with his students, he had decided that it would be their decision to open up their notebooks, and not a compulsory part of joining his laboratory. Salvagno says of this process:

> About a year and a half ago Koch was just saying that we should start to go open, and I said "OK I'll go open" and he just said, "Are you going to go fully open?", and I said "Yeah, I'm just going to do everything that I do now on the private wiki, but do it on the open, you know, public website". So I just went

> full-blown, I didn't even slowly convert. He [Koch] gave us a lot of freedom to do what we wanted and organise our notebooks how we wanted to. (Salvagno interview)

Maloney noted that he was then able to access Salvagno's notebook using his normal online searching behaviour:

> I didn't see how there was too much of a difference between the private notebook and a public notebook, except whenever I decided to Google search something in the public notebook, I realised that - wow - it's totally indexed, and I can search for things in my notebook very quickly, and find the answers very quickly, if I have ever actually written it down. (Maloney interview)

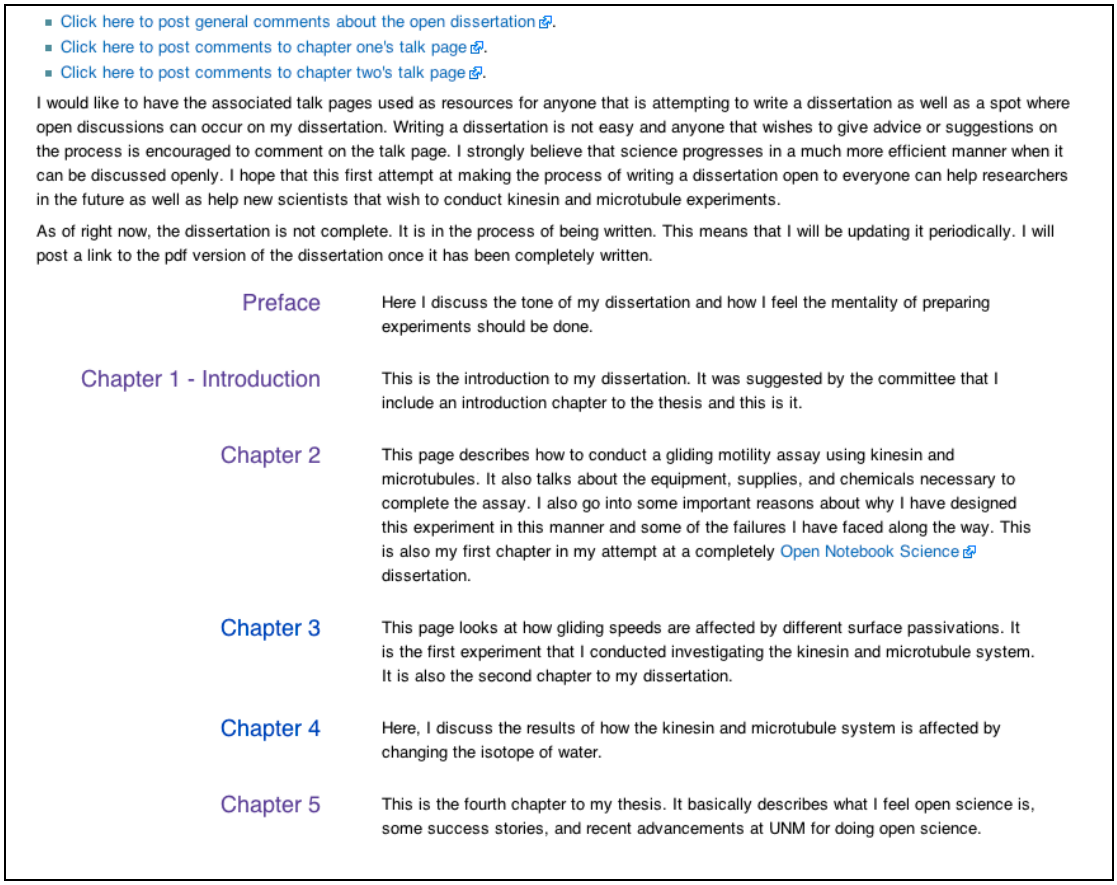
Maloney had used a paper notebook when working in a previous lab, and commented positively on the differences between electronic and paper: "I realised that the wiki was more advanced and advantageous than the paper notebook, because you can post links to papers, you can post videos, you can post pictures, very easily" (Maloney interview).

The laboratory group heads, Koch, Neylon and Bradley, all remarked on the overview of the group's work that everyone having an OLN affords, which Maloney also saw as an advantage: "Steve had access to the notebook 24/7, which means that he could be anywhere and look at the entries for the lab notebooks, and make comments on them" (Maloney interview).

4.7.6 An Open PhD Dissertation

Maloney, a graduate student in the Koch laboratory, not only kept an OLN, but also wrote his entire dissertation on his public OWW wiki, and in May 2011 submitted and successfully defended his thesis.

Maloney wrote the dissertation entirely online, but to comply with University of New Mexico regulations, he also had to submit a "standard" dissertation to his university and examiners, which he did by converting his wiki to PDF format. Maloney found this problematic. He sought advice on FriendFeed on thesis and scientific writing styles (Maloney, 2011). The examiners did not have any issues with the openness of the thesis, but more with Maloney's discursive and informal approach to telling the story of his research.


- Click here to post general comments about the open dissertation.
- Click here to post comments to chapter one's talk page.
- Click here to post comments to chapter two's talk page.

I would like to have the associated talk pages used as resources for anyone that is attempting to write a dissertation as well as a spot where open discussions can occur on my dissertation. Writing a dissertation is not easy and anyone that wishes to give advice or suggestions on the process is encouraged to comment on the talk page. I strongly believe that science progresses in a much more efficient manner when it can be discussed openly. I hope that this first attempt at making the process of writing a dissertation open to everyone can help researchers in the future as well as help new scientists that wish to conduct kinesin and microtubule experiments.

As of right now, the dissertation is not complete. It is in the process of being written. This means that I will be updating it periodically. I will post a link to the pdf version of the dissertation once it has been completely written.

Preface	Here I discuss the tone of my dissertation and how I feel the mentality of preparing experiments should be done.
Chapter 1 - Introduction	This is the introduction to my dissertation. It was suggested by the committee that I include an introduction chapter to the thesis and this is it.
Chapter 2	This page describes how to conduct a gliding motility assay using kinesin and microtubules. It also talks about the equipment, supplies, and chemicals necessary to complete the assay. I also go into some important reasons about why I have designed this experiment in this manner and some of the failures I have faced along the way. This is also my first chapter in my attempt at a completely Open Notebook Science dissertation.
Chapter 3	This page looks at how gliding speeds are affected by different surface passivations. It is the first experiment that I conducted investigating the kinesin and microtubule system. It is also the second chapter to my dissertation.
Chapter 4	Here, I discuss the results of how the kinesin and microtubule system is affected by changing the isotope of water.
Chapter 5	This is the fourth chapter to my thesis. It basically describes what I feel open science is, some success stories, and recent advancements at UNM for doing open science.

Figure 28. Screenshot of part of Maloney's open notebook dissertation, showing chapter outlines and welcoming comments.

Koch's recollections of Maloney's dissertation committee comments were posted on Maloney's dissertation comments page:

> The committee as a whole was uncomfortable with the style of the dissertation. Not the fact that it was open but the (for lack of better phrase) "less formal" style of writing. We agreed that the most likely reason this could matter was the risk that potential future professional connections (e.g. postdoc advisors) may be turned off by the unusual style and discount Andy as a scientist. I agree that's a risk. (Koch comment on Maloney, 2011a)

The committee also stressed that they had no issues with the open science aspect of the dissertation:

> The entire committee was very impressed by the open notebook science, the open data, the open dissertation and everything about that. One member besides me explicitly said that Andy should be commended for his efforts in changing the way science is carried out. (Koch comment on Maloney, 2011a)

The dissertation committee asked for some minor corrections and the addition of an explanatory chapter about a particular scientific aspect of the thesis.

Maloney notes that there have already been positive outcomes from his open thesis: the page has received over 6,000 visits, and he has evidence that the contents of one section of his open thesis have resulted in a much faster replication of a particular scientific protocol, as he stated on FriendFeed:

> Yep, open science works! My open dissertation outlines how to make a gliding motility assay using kinesin and microtubules. It took me nearly a year to perfect the procedure. It took a fellow graduate student at UCSC only two weeks to reproduce my procedures from my open dissertation and get the assay to work. Which, is not trivial and not easy. (Maloney, 2011b)

4.7.7 The LaBlog - blog-based OLN

Cameron Neylon is a biophysicist, and Senior Scientist in Biomolecular Sciences at the ISIS Neutron Scattering facility at the Science and Technology Facilities Council (STFC).[69] Neylon runs a blog-based open laboratory notebook, which has been purpose-built using portions of open-source code as well as custom programming. Neylon's LaBlog[70] had its origins at the University of Southampton, where Neylon was working in the chemistry department alongside Professor Jeremy Frey. Frey is the Project Director of the Combechem project, to which the *Smart Tea* project (and its descendants) belonged, which aimed to develop an ontology of standard procedures in experimental science, by breaking down experiments into their constituent parts, using the process of making tea as a proxy for scientific experimentation (Taylor, 2006). The vocabularies gleaned from this project and other activities within the Combechem project (funded by the UK's Engineering and Physical Sciences Research Council from 2001-2005) have resulted in working code which can be used in laboratories, such as Neylon's LaBLog.

[69] http://cameronneylon.net/about/biographies/

[70] http://biolab.isis.rl.ac.uk/camerons_labblog/

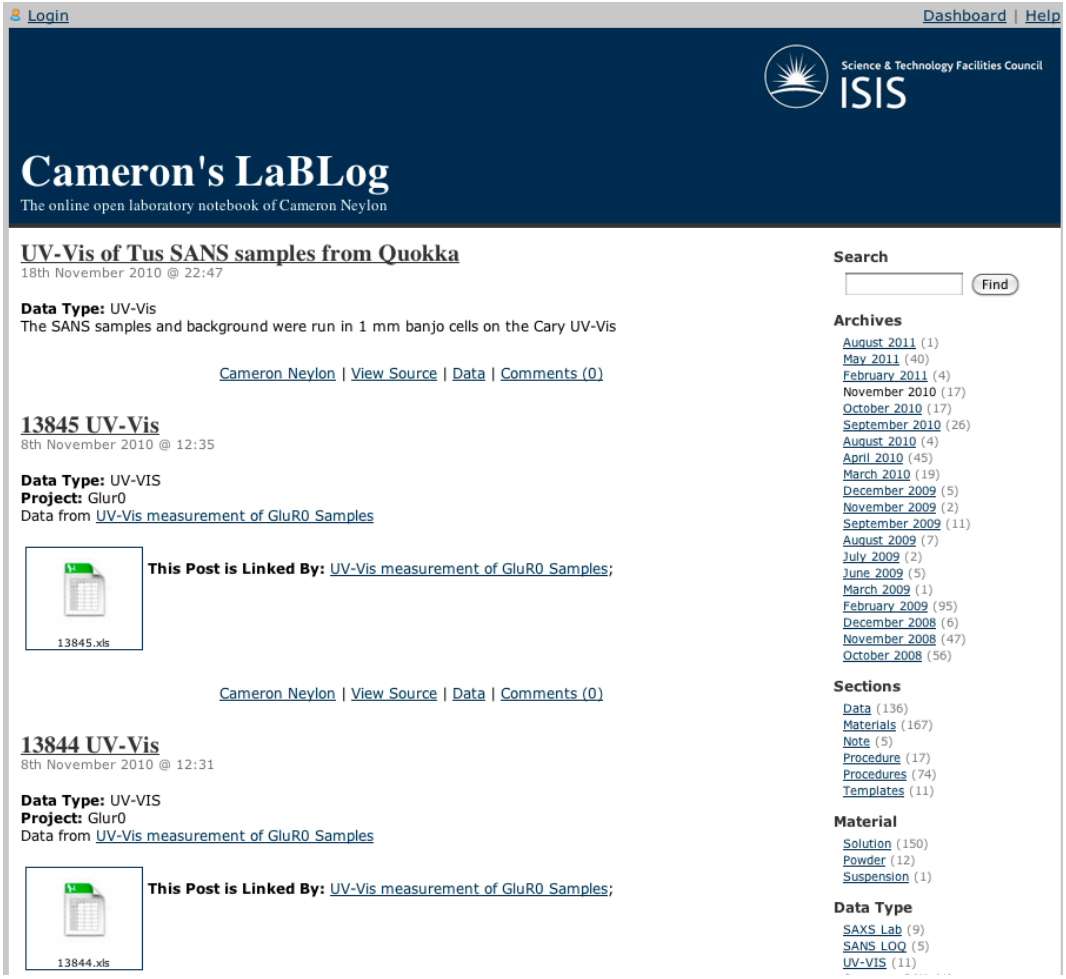


Figure 29. Cameron Neylon's LaBLog

The LaBLog is designed to be Semantic Web compatible. In practical terms, this means that it enforces collection of a wider range of data points about an experiment than the less structured wiki-based OLNs featured above. It supports object linking, tagging and has rudimentary stock management properties, and can record the status of some laboratory machinery (such as refrigerators, seen in Figure 30, below). Categories such as Data and Data Types, Materials, Procedures, Instruments and Projects also assist with structure and classification (although the blog is also full-text searchable and indexed by Google).

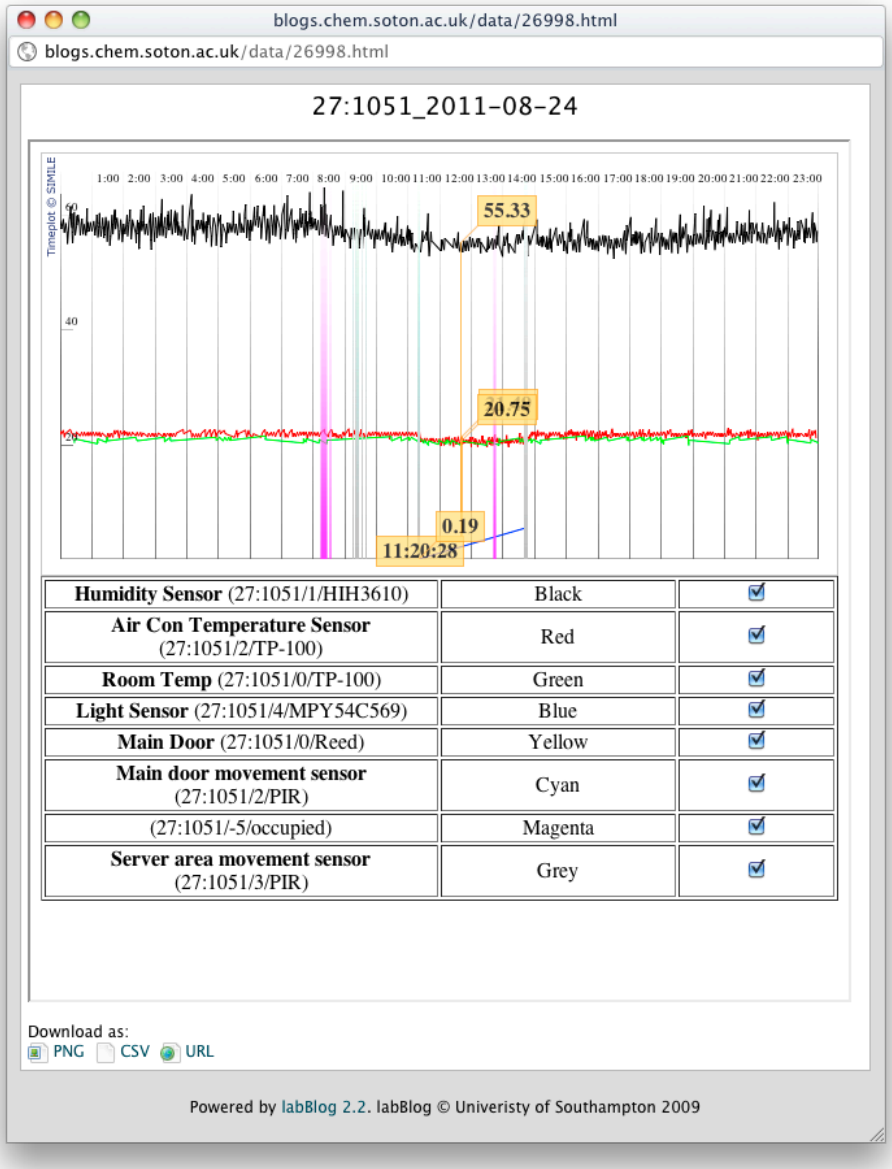


Figure 30. Refrigerator temperature tracking in a Southampton laboratory

The aim of the LaBLog's compliance with some web standards is that it is making steps towards becoming an end-to-end scientific publishing solution, ideally, as it develops, data recording will become more routine, certain categories and posts could be combined to produce, for example, the Methods section of a paper by using some scripting, and it could become a

laboratory stock management aid. However, it is still some way from these ideals.

Neylon's introduction to Open Laboratory Notebooks was unintended, dictated by technological issues outside his control. He was using the Frey group's blog-based notebook mentioned above, based and hosted in Southampton. However, Neylon was accessing it from his new workplace at the Rutherford Appleton Laboratory near Oxford. This involved tunnelling through various firewalls and logins, and the extra work to access it was becoming tedious, as he explains:

> I was having immense problems either configuring, I ended up with these strange settings where I'd have 2 different browsers, 3 different proxy settings to get through different firewalls to authenticate one thing or another ... Just to reach it, and certainly to add material to it. And I just got really tired of it, ... I said *well I've had enough of this, we're just going to make it completely public.* So we did and then I thought, *oh, that's a bit radical! Hadn't really thought about the implications of that* (Neylon interview) [italics mine].

Opening up the pre-existing but previously private laboratory notebook had a marked impact on both Neylon's scientific activities (such as collaborations) and his scientific networking activities. Neylon began blogging on Open Wetware to document the activities around his open laboratory notebook. Opening the laboratory notebook created a focus for Neylon's science networking and his engagement with the issues and discussion it created, an engagement which has led to him being one of the central figures of the SIM around open science.

4.7.8 RRResearch - blog-based OLN

Rosie Redfield is Professor of Cell and Developmental Biology in the Zoology department at the University of British Columbia, and her research examines the ability of bacteria to use DNA from their surroundings, primarily looking at *Haemophilus influenzae*, and using a combination of wet laboratory and bioinformatics microbiology tools to do so.

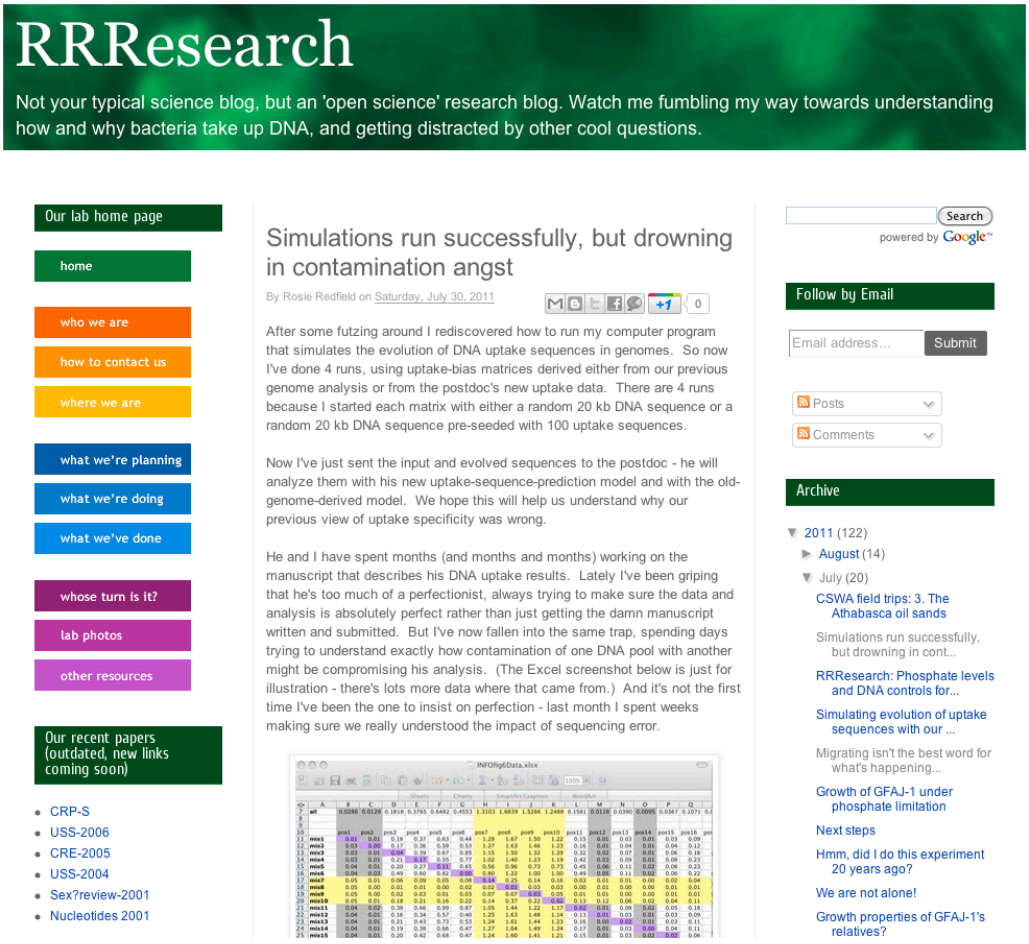


Figure 31. The RRResearch Blog

Redfield has been keeping a scientific blog since mid-2006, on average writing 11 posts per month, or 133 posts per year. Her prime motivation for doing this was to be open about her research. Redfield had been posting her grant proposals, successful and failed, on her laboratory web pages since

1996, and realised that blogging was a way to be even more open about her science: "when I saw this comment about how science was ... behind closed doors, in a science blog, because I'd started reading science blogs, I thought duh, you should have a blog about your research! And so I started a blog called RRResearch" (Redfield interview).

Redfield's laboratory website contains a lot of information about her group's work, contained under headings such as *What we're planning* - links to PDFs of full grant proposals (successful and failed), *What we're doing* - links to the laboratory's blogs, including RRResearch, and *What we've done* - a list of the laboratory's publications, including links to PDFs of all the files, kept on the laboratory's servers so interested readers do not have to go through publisher paywalls.

The laboratory's blogs are those of Redfield and a few of her graduate students and postdoctoral fellows. Redfield encourages her students to blog, but it is not mandatory for them, and Redfield is disappointed with the outcome: "my post doc was writing great blog posts, but they became less – longer and longer and less frequent and now he hasn't done anything since March. I loved reading their blogs, because it gave me insight into what they were thinking and I would go wow, these guys are thinking a lot more deeply than I've been giving them credit for! But then they just didn't really want to do it" (Redfield interview). Redfield partially credits this lack of interest to the prevailing competitive culture in microbiology—although she characterises her work as non-competitive, due to her field being small

and her theories being contentious. Redfield states: "there's a cultural paranoia about talking about what you're doing before it's finished ... not everybody buys into it, but it kind of creeps into grad students and there's always this sense of, you know, there could be people out there who will steal your ideas" (Redfield interview).

RRResearch, the blog, contains a mixture of blog posts about other people's research (of which more below in the science blogging chapter) and Redfield's posts about her own research. In her words, "very few science bloggers actually write about their experiments and their hypothesis and what they're trying to do and all the ways it's not working and the things that I do" (Redfield interview). The aim of RRResearch is to educate people about the realities of scientific life: "most scientists spend most of their time trying to figure out why their experiments won't work", and additionally to stimulate new collaborations and attract new graduate students. However, it has not had that effect—although she states it has been useful in terms of making new contacts, it has not resulted in new collaborations or a rush of new graduate students. However, as is discussed in the blogging chapter, below, Redfield has recently joined a blog network, Field of Science, and it remains to be seen whether the presumably bigger audience she receives as part of a network will be beneficial in this regard.

4.7.9 Marine chemistry - wiki-based OLN

Martin Johnson is a postdoctoral researcher at the University of East Anglia in the Department of Environmental Sciences, studying marine and atmospheric chemistry. Johnson started an open notebook in November 2010, hosted on his departmental web pages in a TiddlyWiki, a wiki written in Javascript and HTML, which does not require a database in the background, as it uses only HTML files.[71] Johnson's open notebook is not fully open—he states it is "Some content, Immediately" (see all of the ONS claim options in Table 9, above), although he would like it to be entirely open at some point in the future.

Like the other OLN users above, Johnson had to make choices about which technology to use for his OLN, as unlike some of those above (excepting Bradley), he has experience building web sites, as he spent five months in 2006 building a website called Research Pages, which was designed to be a content management and social networking site for scientists (it has basic networking features like groups, friending, and blogs). Research Pages still has a few hundred members but is not being actively maintained. Interestingly, despite having the ability to maintain or build a more complicated system, Johnson has opted for a fairly simple wiki, structured only in terms of naming conventions for posts and adding tags to posts. He also examined pre-built solutions such as Open WetWare, but found them quite domain-specific, and too biology-specific for his

[71] Johnson's notebook is at: http://www.uea.ac.uk/~e356/

needs (Johnson interview). This would seem to support the idea that since there is no one-size-fits-all-scientists solution for OLNs, standardisation of these technologies will take some time to develop.

At the same time, Johnson started a blog, and his first post explains why he went open: "I believe strongly in the philosophy of openness in science, as in software, and partly because I feel like airing my dirty laundry in public will force me to be a better and more rigorous scientist and ultimately do better work" (Johnson, 2011).

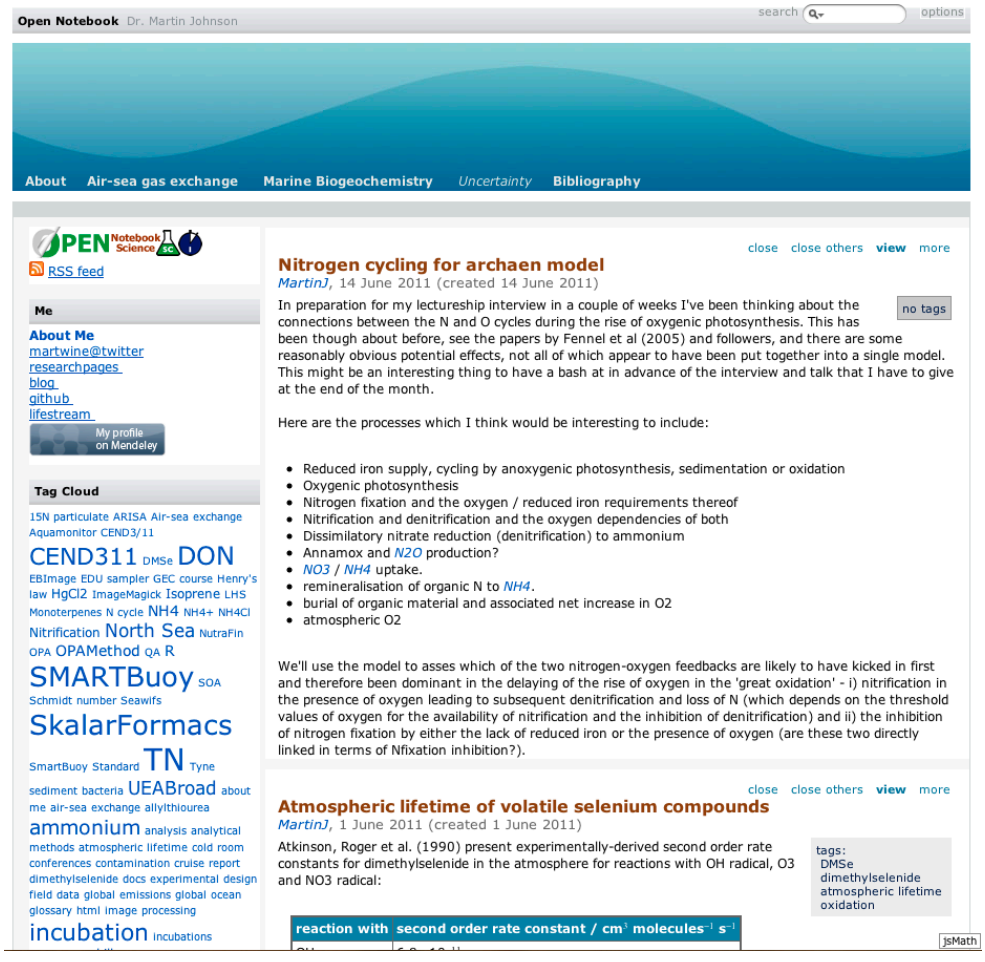


Figure 32. Martin Johnson's Open Notebook at UEA.

In terms of rigour, he has already found that his OLN helps his research process: "my notebook is now searchable and it's all tagged, so I can just

click on the tag I want and I get the answer to the thing that I can't remember. So it saves me loads of time actually. I don't have to recalculate or search back through to find ... the concentration of the standards I used the previous week ... I just go to the page where I last did it and copy it, so I can actually be pretty mindless in the lab" (Johnson interview).

Johnson waited until his notebook was established before talking to the principal investigator of his research project about it, until he was satisfied that the notebook was working for him, as he "didn't want to make a big thing out of it and then it not come to anything" (Johnson, 2011). His principal investigator thought it was a good idea, despite not being technically-minded himself. Johnson notes that his field is not "commercially sensitive or massively competitive" in any case (Johnson, 2011).

Because of the research niche he occupies, and the nature of the data he gathers, Johnson's data is unique, as he states, no one else can say they were at "51 degrees north, 3 degrees east on the 15th of February ... I was there on a ship and we know who was on the ship ... if somebody steals my data and I see it, I'll know it's my data" (Johnson interview). Johnson therefore, is not worried about his research being scooped or his data being stolen. In general, as we have seen, fear of scooping is not something that troubles the OLN-using researchers, partly because they often work in fields where their data is uniquely identifiable (Johnson, Neylon), or because they work in areas where there is little competition (Redfield, Todd), or some of the data

they produce is in and of itself not paper-worthy, but as a collection provides much-needed data for parts of science (Bradley). The discussion below examines further general aspects of the findings above in the context of the seven research aspects.

4.8 Results and Discussion

The examples above show how OLNs are being used by several life science and chemistry laboratories worldwide. Further discussion below expands on this by addressing the research questions of the thesis with specific reference to the use of open laboratory notebooks.

4.8.1 Openness of scientific data and discourse

RQ: Where in the chain of collecting and disseminating research materials are the incentives and barriers to open practices in the use of open notebook science by academic life scientists?

There is obvious impact of OLNs in promoting openness of scientific data and discourse. Despite some current technological limitations, OLN users cite several benefits of OLNs, including the findability of their data via public search, the re-use of their data by others, the transparency and reproducibility of their experimental methods (an essential condition of scientific reporting), and the conversations and collaborations with other scientists that OLN use has facilitated.

Incentives include connections with other scientists, better communications with one's peers and senior scientists, a more thorough record of work done (though this is still largely up to the individual), better

data recording, more ease in writing papers (copy/paste rather than transcribe), and the findability of past records (due to the data being machine-readable) aids experimental reproduction by providing quick access to past methods and recipes for basic chemical solutions.

Barriers include that this more thorough approach may actually take more time, the platforms and data formats may suffer from redundancy, and data can be lost from server failure or browser crashes. Researchers have to worry about archiving their data, the possibility of disinterest or negative responses from colleagues, and potentially, restrictions on the journals that will accept papers, as editors may regard the OLN as a form of prior publication. The issue of OLNs potentially being perceived as pre-published data is discussed below with regard to status and rewards.

With regard to the technology itself, the blogs and wikis used as OLNs are still not particularly well-developed for specialist scientific needs, as "the scientist is left with trying to coerce the available web based applications to do what one needs in order to publish data openly. This is just indicative how young the area of web based open science is and I hope it changes in the future" (Maloney, 2011a). However, the creation of a useful laboratory notebook, whether paper or electronic, is always challenging in any case, as Koch recounts: "when you're doing stuff on paper, or privately electronically, or publically, it's a challenge to get students to have a good notebook" (Koch interview). The use of an OLN does not lessen this challenge, but the public nature of the notebook does

cause them to reflect on the personas they present within the notebook, as Maloney remarks that in his OLN he feels he must "talk about it [a paper] in a scientific and respectful manner as opposed to ... if you have a private notebook you can blow off whatever it is that you find irritating" Thus, he is conscious of adopting a more professional tone than he would scribbling in a private notebook.

On balance, the OLN users described in this chapter are satisfied with the impacts of their data sharing so far, and have seen unexpected benefits in the form of co-authorship and credit, as well as a perception of being able to produce faster scientific results, and tellingly, use of the data produced by a range of other parties, including commercial laboratories. In terms of the open science movement, data-sharing is showing clear evidence of benefits for others as well, as shown by the use of Bradley's open notebook data and Maloney's open PhD methods.

4.8.2 Scientific Roles

RQ: Do online tools extend the professional roles of academic life scientists beyond their traditional ones?

Open Notebook scientists add new skills to their technology repertoire as they are required to evaluate, use and train others in the use of the technology they choose for their notebook. Some extend this further, and learn to use scripting languages or do some programming to further develop their OLN platforms, working on connecting these platforms to data retrieval, storage and archiving solutions (for example the LaBLog and

LabTrove) although as has been shown in this chapter, the wikis and blogs used as OLNs do not necessarily require this, and can be used as-is, provided that the scientists are willing to be flexible in their approach to data recording (for example, OpenWetWare and WikiSpaces).

With regard to changes in the roles of scientific researchers, we have seen that being an open notebook scientist entails greater than usual engagement with technology. OLN users make choices about which technology to use, deciding which constraints are acceptable, which technical features are necessary and sufficient for their current use. They figure out solutions for helping their work get found by colleagues, peers, and the public, deciding levels of access control (who can or should have editorial-level access). Lastly, they must also decide how to archive their data and how to cite it. Having to make archiving and citation decisions means that open notebook scientists also become curators of their data.

This engagement with technology makes the OLN-using scientists more visible in their online networks. As their data becomes findable and their practices more transparent, they are creating a new form of scientific work, one which is not yet measurable through the traditions of publication and citation. This role shift also changes the approach of OLN-using scientists to their research, as was illustrated in this chapter, with the examples of open-source drug design to combat schistosomiasis (Todd's group) and public competitions which allow anyone with access to the correct equipment to

contribute to scientific knowledge production (Bradley's Solubility Challenge).

There is evidence that this visibility is successful in the recruitment aspect of SIM emergence and growth posited by Frickel and Gross (2005). Johnson's OLN at the University of East Anglia is an example which emerged in the recent past, stimulated by witnessing other open scientists and knowing Cameron Neylon, and the publicity gained by Redfield's blog , detailed in the Scientific Blogging chapter, although for a different reason, also draws attention to her OLN practices, particularly if Redfield's reproduction of the GFAJ-1 experiment is eventually published.

4.8.3 Online Collaboration

RQ: Do online tools facilitate inter-institutional and interdisciplinary collaboration for academic life scientists, and if so, how?

There is evidence that scientific collaboration practices are affected by open notebook science, not always in the form of new or expanded formal collaborations, but through informal interactions and assistance precipitated by the discoverability of OLN data.

Examples include Maloney's open dissertation, when researchers at UCSC found his movies of microtubule motion on YouTube, and asked for the data behind them, which Maloney was happy to supply. They used this data in their own, unrelated research, as Maloney reports in his dissertation: "the data I took that was used in their study is well beyond anything I or Dr. Koch could have imagined since before our interactions with them, we

had no idea this area of research existed" (Maloney thesis, 2011). Furthermore, as reported above, the data in Maloney's thesis allowed the UCSC researchers to master the complex techniques involved very quickly, as "what took me nearly a year to perfect, Corey was able to do in a month" (Maloney, 2011a).

Another example from the Koch laboratory involves other researchers also finding their data via a Google search, and asking if they could use it. Koch and Maloney were acknowledged in the resulting paper, and Koch remarked: "I think it's a great success story because (A) they never would have known about our data if it weren't open. It didn't necessarily have to have an open license, but it needed to be discoverable. (B) we never would have thought to use our data for this purpose. So obviously value was created via openness" (Koch, personal communication, 4 Feb 2011).

There is also evidence that open notebook science can be practised in collaboration with non-open laboratory notebook groups, as evidenced by Redfield and Bradley. Redfield is collaborating with a group from Harvard University who will do the mass spectrometry of the results of her GFAJ-1 experiments, and she is documenting the whole process on her OLN blog. Bradley has collaborated with non-open groups and published the data from that work openly:

> We want to design a new anti-malarial drug, [but] we're a synthetic-organic lab, we're not proficient in computation, so we have an open collaboration with a group that will do docking for us, and then we make the compounds, we want to determine if they're active or not, and so we were able to ship the compounds to get assays, these people were not necessarily

> doing open notebook science, but they didn't have an objection to us making the results available in real-time. (Bradley interview)

Thus, UsefulChem has enabled several types of collaboration: the interdisciplinary collaboration between Bradley and Lang, the involvement of students in the ONS Challenge, and lastly, the "bursty work" informal small-task style of collaboration also found in FriendFeed. This mixture of formal and informal collaboration, with both formal publications and small data points as outputs, can be seen throughout the OLN users' experiences, as shown above.

4.8.4 Gatekeepers

RQ: Does the use of open laboratory notebooks by academic life scientists circumvent gatekeepers in science and/or create new ones?

Although open laboratory notebooks are not in widespread use, the employing institutions of the scientists involved have so far proved supportive, as Johnson stated above. This is echoed by Koch, who says: "the chair of the department gave me good feedback on the open science stuff. Probably in particular because I teach a course—a junior physics lab for undergraduate physics majors. I did that course completely open notebook science, open everything, and he thinks that's pretty cool" (Koch interview).

OLN activities do not formally count in terms of research and tenure assessment—after all, they are notebooks, records of the data and ideas generated in the laboratory. However, Todd attributes greatly accelerated

research outputs to the use of an OLN, something which does make a difference in terms of scientific career-building:

> The research went a lot faster than it would have done had we not [been open] ... there's no question about it. It was extraordinary getting the voluntary and the unexpected input of people who really knew what they were doing in terms of solving the problem that we needed to solve. The key was that the people who helped us, we didn't know them before we started the project. So they came on board because they saw our project, and came on board voluntarily. (Todd interview)

The free availability of OLN users' scientific data is not circumventing the traditional gatekeeper institutions in academia, that is, journal publishers, higher education employers and academic funding bodies. At this stage, OLN users are still publishing in journals, although they do tend to favour open-access journals when they have the funds available to use them. Bradley's group and collaborators are also embracing alternative publishing models—so far, they have self-published (using print-on-demand) books of the various chemical measurements they are making around solubility and melting points of various chemicals, aiming to begin correcting what they characterise as a vast amount of chemical misinformation already published as uncorrected errors in standard textbooks and highly-cited papers. Bradley comments on the value of their activities: "I make the case that open curation of melting point data is likely to be a high impact activity relative to the amount of time required to perform it. The problem is that errors such as these cascade through the scientific record and likely retard scientific progress by causing confusion and wasted effort" (Bradley, 2011). Bradley and his groups' effort to correct

the scientific record using crowdsourced open data bypasses the traditional gatekeeper of textbook publishing and makes the data available to more, and more easy to find via an online search.

Scientists in this study have speculated about, but not actually tested the possibility that the usage of OLNs may also constrain formal publication options, as some journals regard OLNs as a form of pre-publication, and may not accept submissions of papers based on open data (for example, American Chemical Society (ACS) journals).[72] However, none of the active OLN users interviewed have attempted to submit to an ACS journal. Bradley believes that his practice of drafting his papers online, alongside his OLN data, violates the ACS pre-publication rules, and therefore chooses to publish in other journals, often open-access journals. This limitation (although only noted in ACS journals to date) could potentially discourage graduate students from becoming OLN users.

---

[72] The ACS publishes policy summaries on prior publication for its journals. See for example the Journal of the American Chemical Society, reproduced below. The wording is sufficiently unclear that a prospective author would need to contact the editors to clarify the status of their ONS data with regard to prior publication: "The Journal of the American Chemical Society considers for publication only original work that has not been previously published and is not under consideration for publication elsewhere. When submitting a manuscript, an author should inform the editor of any prior dissemination of the content in print or electronic format. This includes electronic posting of conference presentations, posters, and preprints on institutional repositories and any other Web sites. Any content that has been made publicly available, either in print or electronic format, and that contains a significant amount of new information, if made part of a submitted manuscript, may jeopardize the originality of the submission and may preclude consideration for publication". http://pubs.acs.org/page/jacsat/submission/prior.html

Bradley also notes in an interview that he has not attempted to submit OLN-based work to ACS journals mainly because he also drafts papers online and openly, which he considers violates policies like the above.

4.8.5 Status and Rewards

RQ: Do the various forms of online dissemination being used by academic life scientists yield rewards and status comparable to traditional outputs, or are there new mechanisms developing for reward and recognition of these activities?

In terms of scientific career development and impact on scientific status and rewards, some graduate students using OLNs can demonstrate that their use has enabled them to connect with more senior scientists. Some of the laboratories featured in this case have won awards because of their open approach. However, these benefits may well only accrue at this early stage of ONS adoption – if ONS becomes widely used it is possible that these early-adopter benefits will disappear as the practice becomes common. If ONS use was widespread, it could be expected that academic credit and rewards might be forthcoming for those whose data sharing and re-use was most productive.

The students participating in this study cited several benefits from their OLN use, ranging from the connections and advice from more senior scientists as well as a taste of "Internet fame", as Salvagno put it: "I feel like I'm sort of famous. Even though I don't have millions of page views ... you search for my name and my notebook is the first thing that pops up ... Plus I feel like I have a lot of good connections to prominent scientists now that I probably wouldn't have if I hadn't done the open notebook" (Salvagno

interview). Johnson also relates to this, in terms of realising that just getting one's name out is useful:

> A lot of success in science is really about promotion, so the more my name is associated with things, the better, basically, so if people are searching for stuff and come across my notebook, even if they just read my name and the title of one of the posts, then, you know, kind of brand awareness ... career-wise, I feel like being at the cutting edge is probably pretty useful. Particularly in environmental science after climate gate. (Johnson interview)

There is good evidence that ONS data is being found, as Bradley reports up to 100,000 hits to date on his group's solubility database. The Bradley group have a citation format for the ONS which they use themselves, and in the past year have begun to find citations to their work. Bradley remarks that: "people have cited our wiki, or our blog, things like that, in articles that are generally about open science and related issues. So we are getting citations through the peer-reviewed literature that way" (Bradley interview). At time of writing, Bradley's ONS Challenge data has also been cited in one nanotechnology paper, although this does not reflect the true usage of the database, as Bradley notes: "one formal citation of our solubility dataset doesn't represent the 100,000 anonymous solubility queries" (Lau et al., 2011, Bradley, 2011). Maloney's experience with his open PhD dissertation and the reproduction of his research technique at UCSC (discussed above) was also beneficial—Maloney was offered co-authorship of the resulting paper but was satisfied with an acknowledgement. Opportunities like this demonstrate that the data in

OLNs is easily found by other specialists, and can result in paper and collaboration opportunities.

Several research participants commented on the fact that they did not think that being scooped by others was a possibility, citing reasons such as the impossibility of being able to reproduce the data (Johnson), others' disinterest in using the data (Bradley, Todd), the fact that the data was free to use, with no licensing restrictions at all. If another scientist used their data to produce a paper before them, they do not have a problem with that—although probably better science would be done by collaborating with the data producers (Bradley). Finally, the OLN users are sceptical that there is enough time for the average scientist to be stealing others' ideas: "all scientists think their ideas are like gold, and therefore they think that if anyone gets their idea they're going to rush off and do it—but in reality nobody has time to do all of their own ideas anyway" (Koch interview).

OLN use thus exemplifies one of the core features of a SIM, contained in technological form: that of participating in contentious activities relative to normal behaviour within their disciplines. The use of an OLN may not yet accrue significant academic rewards, but the existing evidence shows that it supports the normal processes of science while enhancing contact with other scientists, adding potential collaborators and subsequently more academic papers to a scientist's profile.

4.8.6 Science Communication

RQ: How do academic life scientists use open laboratory notebooks to publish different types of scientific data?

OLN practitioners publish all of their data, in as close to real time as is possible. Redfield and Johnson, featured in this chapter, keep a partial/pseudo OLN, which means they do not publish all data, but certainly still most data. (see Table 9, for example: "A portion of the lab notebook and associated supporting raw data are available after some delay. Others cannot assume that if it isn't in the notebook you haven't done it"). The consequences of full data disclosure are only just beginning to be felt by OLN users, but as demonstrated throughout this chapter, disclosure only seems to encourage communications with other scientists.

OLNs are a useful tool for stimulating and improving communications within a laboratory setting, and although they are not normally used for external communications, their form and content does invite feedback and discussion, usually via another site such as FriendFeed (Chapter Three) or a blog (Chapter Five). Internally, the laboratory group benefits from the searchability of the OLN how to prepare a particular solution correctly, how to build pieces of equipment, or get questions answered quickly:

> A big advantage of a wiki as opposed to a paper notebook is that I get to read everybody's notebooks that's in the lab. ... Which ... has actually come in handy a couple of times when either I, or someone else in the lab, has not had the specific knowledge to do something, of course we immediately collaborate from doing that, because we'll post a question, and someone will answer it pretty quickly. (Maloney interview)

Another benefit is the availability of students and group members' data to the boss: "I will check in [in the morning] with the lab notebooks that the students are doing, see if there's any questions that came up. This is all from home ... in the evening go through all the student's lab notebooks, the grad students and undergrads in the lab, just make sure I've commented on stuff" (Koch interview). This shows that full transparency among laboratory groups is achievable, and some evidence is amassing that shows it may be beneficial at a wider level.

4.8.7 Scientists and public engagement

RQ: How do academic life scientists see that their engagement with the public has changed since they adopted open laboratory notebooks?

Despite the openness of their research, those practicing open notebook science are not, in general, engaging in research that the general public can participate in or comment on without fairly extensive scientific training. Although Bradley's Open Science Challenge is indeed open, and public, it still requires access to a reasonably well-equipped chemistry laboratory and chemistry knowledge roughly equivalent to an undergraduate degree, or at least guidance from someone with that knowledge.

However, as will be further discussed in the next case study of science blogging, the scientists' OLN activity is often contextualised and explained through an accompanying blog, which can be more accessible to the public, or at least their peers, the scientific public.

The research participants' choice of research topics is perhaps more indicative of how OLN users may engage with the public, in that several of them choose to investigate the basic building blocks of drug discovery for neglected tropical diseases such as malaria (Bradley) and schistosomiasis (Todd). Although it can be argued that these are good targets for open science due to the fact that they do not attract the attention of profit-seeking pharmaceutical companies, and thus are not competitive areas, it is also true that this type of research is very much in the public interest, and it follows that the research data is of most use in the public domain.

4.9 Conclusion

In this chapter, I have demonstrated across the seven research questions that scientists' roles have been expanded as ONS practitioners attempt to find and improve the technology they interact with daily; for ONS users, the way they find collaborators is changing, as their ONS data is found and scientists from related or tangential disciplines make contact with them; that the practice of science via OLNs becomes part of the basis of scientific communications for users; that there is a (perhaps temporary) flux in who the gatekeepers are and how they influence the science being done; and that the OLN approach lends itself to potential involvement of the public in the raw science (depending to some extent on the subject matter being studied and what units it can be broken down into).

Status and rewards are still battles being fought. Career assessment is still largely measured in papers, although there is little institutional

resistance to the OLN approach in practice. Having the freedom to operate openly can also constrain to some extent, as some participants have mentioned, it influences what they research and which journals they can submit work to. With regard to funding, philanthropic and government funders are well-disposed towards open science, due to its wide dissemination.

The way in which people came to be open was via different routes (serendipity, getting around technology, planning) and using different platforms (wikis, blogs, already existing or build-yourself projects)—but the platform is not the main concern at this stage, although data standards and ease of adoption will become greater concerns in future. It does not seem to matter what precipitates the openness, whether it be necessity, serendipity or choice. Although Neylon went open originally by necessity (the technology to get to his notebook was a barrier to getting work done)—he then realised the advantages: "we made it open but talked about what we were doing, made that open, and people responded, and corresponded, and became part of a community, a wider community, and that community grew through mutual support and had a lot of the same interests which held it [together]" (Neylon interview).

At this early stage, with no established common software solution or set of operating procedures for this merger of bench and online science, some practitioners are spending as much or more time creating and tweaking the technology implementation than they would have spent using a paper

laboratory notebook, but they persist with this activity because they are experiencing other benefits such as a greatly increased scientific social network and, indeed, some measure of Internet "fame" among their peers in their online communities (an effect which may only be present for these early adopters). They have been able to exploit these contacts for new collaborations and significant peer and public outreach (see the list of papers in Appendix G). There is also still potential for online science practitioners to reap benefits later in their careers due to the discoverability of their prior work, or their successors in their present laboratories, when the data and details of methods and experimental models are easily accessible. Within the limited domain of their OLN activities, the OLN users have created a highly interconnected platform for their science, interweaving data, publication and discourse as they use their OLNs, FriendFeed, and blogs. Chapter Five will discuss the use of blogs within the scientific research process, and in Chapter Six, data from these three embedded cases will be used to illustrate how they are situated within the SIM/CM framework described in the Introduction.

Chapter 5
Scientific Blogging

5.1 Introduction

Currently available data suggests that less than 0.03% of scientists keep blogs, or that 0.001% of all blogs are written by scientists.[73] However, evidence in this chapter will illustrate that scientific blogs have a disproportionate impact on the outputs of mainstream media and popular scientific media. There are over a hundred and fifty million blogs in the world (2011, blogpulse.com). Of these, the vast majority are robot-generated spam blogs, personal diaries, and blog authors who comment on publicly-available news media.[74] A much smaller number concentrate on filtering the news available in specialist areas, such as politics and health. Academic bloggers are a minority among specialist bloggers. Very few studies exist which characterise the academic blogging world (or *blogosphere*), and the numbers of academic blogs are no more than educated guesses, and then generally only for the blogosphere writing in English.

One science blogger described the changing state of science blogs, saying that: "five years ago, I read every science blog in English. ... I could, as there were only dozens of us" (Zivkovic, 2010). This state of affairs has changed

---

[73] Estimated using Royal Society figures of 7.1 million scientists (see Footnote 15) 150 million total blogs (Blogpulse.com), and over 2000 registered scientific blogs (Munger, 2010)

[74] Reliable figures are difficult to ascertain, but estimates of spam blog figures range from 5% to 20% of the blogosphere. Retrieved from http://www.caslon.com.au/weblogprofile18.htm.

very quickly. In 2011, the number of science blogs is estimated at more than 2000 (Munger, personal communication, 19 September 2011). The fifth Science Online conference in 2011 sold out in one day, with more than 286 attendees and many more "virtual" followers online. Attendees and commentators generated over 15,000 Tweets and almost 500 blog posts.[75]

Ease of setup and use allow blogs to stimulate conversation and debate among the online population. The explosive growth of this phenomenon has led to the formation of blog aggregations that "aggregate, curate, and amplify the global conversation online" the writings of selected blog authors (GlobalVoicesOnline.org, 2010). The emergence of publisher networks of scientific blog authors, (including ScienceBlogs.com, Nature Network Blogs, PLoS Blogs, Wired Science, Scientific American Blogs) and independent author networks (such as Field of Science, Occam's Typewriter, Scientopia) as well as aggregations of science blogs run by commercial entities (Researchblogging.org and .com[76], Nature.com blogs) and groups of volunteers on Scienceblogging.org and Scienceseeker.org attest to the impact that science blogging is having on publishers and blog readers.

The two previous chapters presented the cases of scientists using online tools to record and share laboratory data on the public Internet via open

---

[75] http://scienceonline2011.com/look-whos-coming.htm , http://archivist.visitmix.com/pgroth/12

[76] Researchblogging was started by Dave Munger in 2008 and later acquired by Seed Publishing.

laboratory notebooks, and to connect with colleagues online using an informal networking tool, FriendFeed. The FriendFeed case argued that the networking establishes an invisible college of scientists working towards specific goals in changing the way science is created and communicated; that is, a specific vision of open science. The OLN case argued that open data in digital form allows unprecedented access to and sharing of raw experimental data, which can change the way scientific knowledge is created and extended.

This chapter presents the last of three interconnected online activities in science—examining the actions, motivations and outcomes of selected scientists who keep a public blog. Blogging is the bridge between the informal networking found on FriendFeed, and the official records of knowledge production found in the open laboratory notebooks. Blogging serves as a personal discussion space individuals use to reflect on their scientific activities, to discuss specific and general aspects of science with their peers, and to engage in making their science public, which can also involve explaining and engaging with the general public.

Scientific conversations begun on blogs and extended via comments on online tools such as FriendFeed and Twitter can be likened to the types of interactions scientists have had for many years in face-to-face forums such as disciplinary conferences, where scientists deliver their evidence during the formal rituals of talks and posters, and discuss their ideas, strengthen personal connections and make new connections between sessions, in coffee

breaks and at conference dinners—much as scientific bloggers deliver blog posts, then disseminate and discuss them on FriendFeed and Twitter. Blogs can also be seen as more informal versions of publications, or extensions of their online precursors—postings of information in email lists, or of personal websites.

However, blogging also extends this interaction by extending audiences beyond the immediate discipline of the scientist. Discussions on scientific blogs have the potential to reach a wider audience than the networks involved in FriendFeed and the specific types of science being done in OLNs. This wider audience, although often still scientists, allows for more contentious discussion on blogs, which can make scientific controversies visible in public—for example, arguments about blog commentary versus formal peer review (as we shall see in the "Arsenic Life"[77] paper incident related below) and the introduction of commercial advertising content into a science blogging network by its owners (Seed Publishing and "PepsiGate"). These examples illustrate current tensions between formal and informal scientific communications.

The scientists participating in this case study are using blogs for a variety of reasons: to talk about their everyday scientific work, to publicise their views on open science and related topics such as publishing and peer review, to develop and get to know their audiences via blog discussions,

[77] So-called after its Twitter hashtag, #arseniclife

and to explore via blog posts their own positions and arguments in relation to the scientific work they care about.

This study distinguishes between OLNs that are kept in blog format, and blogs kept as discussions and diaries—a distinction that the users also make, keeping a blog separate from their scientific work with OLNs. It also shows how blogs contribute to the open science SIM, by connecting the authors' long or short-form blog posts with commentary and discussion in other online tools such as FriendFeed and Twitter.

5.1.1 Background

As has been shown in both of the previous case studies, users of scientific social networking tools such as FriendFeed and OLNs are blurring the boundaries between formal and informal scholarly communication as they make open and public their creation of the raw materials of scientific knowledge. It will be shown in this chapter that scientific blogging also overcomes this boundary between formal and informal (see below) as blog discussions bypass established (and slow) systems of peer review to air their scientific opinions and criticisms quickly and publicly.

At the formal level, which traditional peer-reviewed print journals continue, many academic publishers now also have online open-access journals or articles within journals. These journals and articles are frequently available in advance of print publication, as if to forestall less formal methods of publication such as self-archiving (for example, Oxford University Press *Journal of Nucleic Acids Research*). Online-only open-access

journals (such as in Public Library of Science (PLoS) and BioMed Central—which are available online for free and provide a paper version only through print-on-demand) have also emerged, not least because the information-sharing capabilities of the Internet provide competition within and for traditional scientific publishing by being faster and more efficient at disseminating scientific knowledge. At the informal level, what used to be confined to "water-cooler chat" and the conference circuit is now also discussed on mailing lists, forums and blogs (Borgman, 2007). As these online discussions may remain findable for years alongside or ahead of the research literature, they also have potential to continue contributing to scientific knowledge acquisition and sharing practices, whether or not their writers are engaging in post-publication peer review analysis of formally published research.

In their study of Web links to open access journals Kousha and Thelwall (2007), asked "are there new Web types [of online citation links], and if so, are these trivial or potentially useful for studying or evaluating research communication?" (p. 1631). We know that scientific scholars are linking from blogs and home pages to ISI-indexed journals and discussing papers from them in their blogs (Kousha & Thelwall, 2007a; Vaughan & Shaw, 2005; Munger, 2011). What this evidence does not reveal is the context of these linking practices within discussions of peer reviewed research, which is discussed in detail within this chapter.

A great deal of the discussion around open science and the meta-examination of the tools the participants use to do it takes place in the scientific blogosphere. There are three main types of science blogs, which I characterise in this study as:

(1) *individual blogs* - kept by an individual, either self-hosted on a server or using a public blogging service such as Blogger or Wordpress;

(2) *blog networks* - collections of blogs hosted and maintained by an organisation, five active examples in July 2011 being Scienceblogs.com, maintained by Seed publishing (60 blogs); Nature network blogs, maintained by the Nature publishing group (hundreds of blogs, anyone who signs up to Nature network can blog); Guardian Science blogs, maintained by the Guardian newspaper in the UK (5 blogs); Discover Science Blogs, maintained by Discover science magazine (10 blogs); and PLoS Blogs, maintained by Public Library of Science, open-access publishers (20 blogs).

Table 10

Scientific Blog Types and examples (non-exhaustive)

Scientific blog type	Examples
Individual	Cameronneylon.net RRResearch.blogspot.com (Rosie Redfield, University of British Columbia) Intermolecular.wordpress.com

Scientific blog type	Examples
	(Mat Todd, University of Sydney)
Non-Academic Publisher Blog Networks	Scienceblogs.com Scientific American blogs Discover Magazine blogs Guardian Science blogs
Journal Publisher Blog Networks	PLoS blogs Nature Networks
Independent Blog Networks	Occam's Typewriter Field of Science Scientopia
Blog Aggregations	ScienceSeeker.org (individual blogs) ScienceBlogging.org/com (feeds from blog networks) ResearchBlogging.org/com (blog posts about peer-reviewed research)

Scientists who blog within networks are normally invited to do so by the hosting organisation, and they may be subject to some editorial rules (for

example, a certain frequency of posting may be expected, or there may only be simple community guidelines (for example PLoS blogs—"don't post anything illegal, don't plagiarise, have fun").[78] More recently, self-governing blog networks have emerged, created by groups of scientists who share a blog space but blog individually within it, such as Occam's Typewriter (13 blogs), Field of Science (20 blogs) and Scientopia (35 blogs). Science blogging platforms are volatile at the time of writing (mid-2011). New blog networks are being started, and abandoned, at a high rate. Some of these new networks were created in response to controversies within existing blog networks, such as the exodus of approximately 30 bloggers from ScienceBlogs.com in 2010 in response to advertorial content misrepresented, in the eyes of many of their blog authors, as scientific research blogging. This controversy, also known as "PepsiGate", is fully explored in Section 5.6 of this chapter.

(3) *Blog aggregations* - groupings of science blogs based upon certain criteria. For example blog posts about peer-reviewed research are aggregated on Researchblogging.org and .com. Scientists blogging about peer-reviewed research articles will add an image and link back to the Researchblogging site which allows it to locate and aggregate these posts. This system allows a blogger to write about a variety of topics, but only have their posts about peer-reviewed research aggregated by the site.

[78] http://blogs.plos.org/about/guidelines/

Blogging differs from other forms of online scholarship discussed in this thesis because blog discussion can be argumentative, controversial, and less collegial in several ways, which is less often observed in other online science networking forums such as FriendFeed; although this seems to depend on the blogger to some extent, and also the fact that science blog readers are more diverse—they may consist of scientists and non-scientists, and are not necessarily working towards the same goals as the bloggers (unlike, for example, the FriendFeed users collaborating on small projects together, or working on actions to promote open science).

Blogging is more concerned with and connected to the relationship between scientists and formal scientific publication than the informal water cooler chats and data production that were observed in the case studies of FriendFeed and OLNs. Longer-form reflective or analytic writing is more common in blogging than in FriendFeed. However, many bloggers are flexible (or possibly mercurial) in their attitudes to their blog and the blog writing process. They leave networks, start new ones, move to independent self-hosted blogging and back to a network, and/or take a hiatus from their blog (announced or unannounced). These changes may be in response to perceived editorial infringements as independence of voice is important to them; stopping can be because of burnout or because the pressures of daily work intervene (for example, comments like: "I'm writing grant proposals this month so I won't be blogging").

As it is a format which allows much longer-form writing than FriendFeed or the 140-character limit of Twitter, blogging is more closely aligned to scientific writing and analysis, and as such this case relates more to the research aspects of scientific connections with the public and the gatekeepers of scientific production, particularly publishers, but also institutions. The science bloggers studied in this chapter are a mixture of independent bloggers and network-affiliated bloggers because the scientific blogosphere is a mixture of these, and the examples in this chapter show the connections and disconnects between the different forms of online communication examined in this thesis.

5.1.2 Chapter structure and scope

This chapter will examine the changes that have been happening in the science blogging space over the time span of this research (2007-2011), which has been a period of growth and maturation of the medium as a scientific communication tool. This maturation has not been without growing pains for its authors and readers.

Two examples of disruption of the boundaries of traditional scientific discussion and dissemination will be used to demonstrate these changes. The "Arsenic Life" controversy illustrates the tension between current journal publishing practices and scientific rights of reply to the published research literature. The Scienceblogs.com "PepsiGate" controversy shows the will of science bloggers to be editorially independent and the current

fluidity of the scientific blogosphere, as well as some indications that science blogging can impact the output of mainstream media.

The literature about scientific blogging is better developed than it is for scientific social networking and open laboratory notebooks. However, as shown also for OLNs and FriendFeed in previous chapters, most research in this area does not focus on scholarly blogging activities, but rather on topics such as diary-style personal blogging, health blogging, and the large and popular area of political blogging. The literature review section of this chapter will examine past research in categorising blog types, the related area of analysing blog content, and finally will look at the small body of research on scholarly communications and blogs.

5.1.3 Definitions

In 2007 Technorati.com reported tracking 70 million blogs, of which 55% are regularly updated, while 120,000 new blogs were created per day (Sifry, 2007). Almost 40% of American and 22% of British Internet users read blogs, and blog audiences are steadily increasing (Dutton, Helsper, & Gerber, 2009; Lenhart & Fox, 2006). Qualitative studies show that these audiences are often loyal followers of particular blogs rather than browsers (Tremayne, 2006).

Blogs are Web journals or web-logs which have particular technical features differentiating them from other types of self-publishing on the Web (for example, personal Web sites or social networking sites such as Facebook, FriendFeed, and Twitter). Blogging software allows the user to

quickly create diary-style dated entries including text, video or audio, and then to tag, publish and promote them through a variety of services. Content is arranged chronologically, with the most recent content items (known as posts) most prominent on the blog site. Blog posts usually allow readers to leave comments and thus engage in discussion with the blog author. Blogs are particularly suited to quickly disseminating information in the current World Wide Web infrastructure due to a number of features:

- **Blogrolls** are the lists of links to other blogs or sites of interest that most blog authors compile.
- **Trackbacks** and **pingbacks** are features that "ping", or tell the blog-tracking search engines that a new post has been published, or that a comment has been made on a particular post, or when another blogger has referenced a post on their blog.
- **Tagging** provides authors with an option for classifying posts by adding keywords. Tagging allows for greater findability of the post content in search engines and grouping of related posts in informal categories (or tag clouds).
- **RSS or newsfeeds** of blog post content are a feature of most blog platforms. RSS feeds allow other software known as readers (such as Google Reader) and desktop applications (for example FeedDemon) to read multiple authors' blog entries in one place, without having to visit each blog individually. These feeds can be either a full feed, where the entire blog post is supplied in the feed, or truncated with a "Read more…" link to the full text.
- **Types of blogs** include personal diaries, filter blogs (filters, or links with commentary on information from around the Web), notebooks (long-form essays) or K-logs (which are often written around technological issues or projects, and may resemble project

logbooks) (Blood, 2002; Herring, Scheidt, Kouper and Wright, 2006; Herring & Paolillo, 2006).

- **Free blogging platforms** are numerous; it only takes a couple of minutes to set up a blog, with no overhead costs for the author. It is not difficult to install and use blogging software on a personal domain. The most popular free blog platforms are Blogger.com and Wordpress.com.
- The **blogosphere** is the collective term for all blogs. Due to the networked nature of blogs, groups of blogs on a particular topic tend to form dense clusters when viewed as hyperlink networks. For example, the "political blogosphere" or the "scientific blogosphere".

5.2 Relevant literature

Most research about blogging to date does not examine online scholarship or scholarly blogging; instead, it focuses on identifying and creating typologies of blog content by topic, or alternatively by approaches to content. Other research examines author motivations for blogging. The hyperlinked nature of blogs also attracts significant research on how they are networked and whether communities form around different blogging topics or networks, which will be examined below. There is a growing body of literature on other types of blogging, such as the influence of blogging on politics, social change, protest and marginality, and its potential influence on mainstream media reports. For example, Hindman found that techno-utopian ideas of political power-sharing on the internet are not supported by data—the most popular voices are still the traditional political elite (Hindman, 2009).

However, this research, while not directly about science blogging, does provide several insights which are also evident in the science blogging patterns studied in this chapter. For example, many science blogs provide a filter of already-published content, while others are personal journals with few external links. This section will examine the existing research on uses of blogs in scholarly communications, and the literature on science blogging. Much of the published research about science blogging is written by scientists who already blog, and wish to promote the practice within their own communities. It is worth examining because even these small-scale studies and personal reflections echo the experiences of the science bloggers interviewed for this chapter, and indicate that there are commonalities in the experience of blogging.

5.2.1 Blog Typologies

In the English language blogosphere, several studies have attempted to create an overview of existing blog genres. Herring, Scheidt, Kouper and Wright (2006) conducted a longitudinal content analysis of blogs in 2003 and 2004 which showed that over time, blog authors tend to write more and include fewer links (that is, become less like a filter blog and more like a notebook). They also found that the numbers of pseudonymous blogs are decreasing over time, as authors are more willing to reveal their real names and occupations. Both Herring et al. (2006) and Papacharissi (2006) found that the largest proportion of bloggers in their samples were students.

Bar-Ilan (2005) found that topic-oriented blogs tend to "concentrate and filter" information and hyperlinks on their topics from around the Web (Bar-Ilan, 2005 p. 305). Herring et al. (2006), found that roughly 70% of blogs were personal journals, 15% filter blogs[79], 3% K-logs, and 12% were mixed. Trammell and Keshelashvili found in their study of the "A-list" (most popular authors across all blogs) that "mixed type blogs were the most frequently used (37.8%), followed by diary/journal type (28.7%), and notebook (26.8%)"(Trammell & Keshelashvili, 2005, p. 974). Trammel, Tarkowski, Hofmokl and Sapp (2006) classified Polish blogs, and found that Herring et al.'s categorisations were consistent with the Polish blogging population, implying that there is a fairly stable broad typology of blog genres across languages and countries.

The science blogs in this study are mostly K-logs and notebooks, with some filter-blogging depending on the author—although filtering is active on FriendFeed, and increasingly transferring to Twitter, many of the scientists in this study are reserving their blogs for longer-form writing or extended commentary on particular articles.

[79] Relevant blogging terms are defined above in Section 5.1.3.

5.2.2 Blog Authorship

Although attempts have been made to quantify what blog authors talk about and what motivates them, this has largely been in the area of personal blogging, in the sense of a journal, or of filter blogs (more outward-looking blogs that direct readers to other websites, the most popular subgenres are politics and technology), or A-list (popular) blog authors in general, rather than in a professional or academic subgenre (Tremayne, 2006). Carter (2005) interviewed eight topic-based bloggers and analysed their blog posts to discover the role of authors in filter style blogs. The results suggested that authors of topical blogs often do not create new content although they strive to, and over time they may become more successful (Carter, 2005; Herring et al., 2006). They also often follow the journalistic convention of clearly indicating changes and errata if they edit a blog post, use the content of their blogs as an archival and reminder tool, and are connected as a community by a particular set of sources—in this study's case, links to legal judgements (Carter, 2005).

Hsu and Lin (2008) surveyed 212 Taiwanese bloggers to test their hypotheses that people are positively motivated to blog by six potential factors: perceived usefulness to the author (in terms of perception of enhanced performance), ease of use of the blog platform, enjoyment of the activity, positive attitudes towards blogging, acceptance of social norms and sense of identification with a community. Their respondents were mainly personal bloggers, and they found that the strongest motivations were enjoyment, ease of use and a sense of identification with an online

community. The authors cautioned that their sample was self-selecting (Hsu & Lin, 2008).

In a study of the factors influencing 75 medical bloggers' knowledge sharing activities, Karimi and Poo (2009) also found reputation (in the sense of perception that blogging will enhance their image and status) and enjoyment (of the activity and feedback from others) were major factors motivating the authors blogging activities. Trevino (2005), in a small interview study of college student bloggers, also found that audience feedback was a strong motivator for authors to continue blogging, although like many their initial impetus was to blog for themselves, to record ideas, thoughts and feelings about their lives. Findings from these three studies (Hsu & Lin, 2005; Karimi & Poo, 2009; Trevino, 2005) stress the enjoyment and sense of community that authors gain from writing a blog; this is echoed both in studies of scholarly blogging (below) and in the stories of the science blog authors interviewed for this chapter, whose motivations for blogging are examined in detail below. However, these motivations are also more diverse than simply enjoyment, as Nardi, Schiano and Gumbrecht (2004) found, below.

Nardi et al. (2004) interviewed 23 bloggers; they started with Stanford students and snowball sampled from there. They found very diverse approaches to topics and style even within a small middle-class sample. Their participants also reported that blogging was a social activity for them. The authors characterised blog discussion as "a studied minuet between

blogger and audience" (p. 5), and that although "the blog creates the audience, the audience [also] creates the blog" (Nardi et. al, 2004, p. 4). They found that blogs are a form of social communication, and elicited five motivations for blogging: to update others on activities and whereabouts, to express opinions to influence others, to seek others' opinions and feedback, to "think by writing", and to release emotional tension (Nardi et al., 2004, p. 4). Again, this diversity is found in the science bloggers studied in this chapter, particularly the concept that the blog and its audience are intertwined; several of this chapter's participants expressed their unexpected delight at finding they had some kind of audience, and welcomed the feedback they received in comments and emails.

The PEW Internet and American Life project conducted a telephone survey of a nationally-representative selection of the USA's 12 million bloggers (8% of Internet users) (n=233). Most bloggers in their sample were personal diarists (37%); none in the "other" category (<1%) said they blogged about science (Lenhart & Fox, 2006). Thus in terms of the whole blogosphere, science bloggers are likely to be a very small subset, as estimated at the beginning of this chapter, although they attract a disproportionate amount of mainstream media attention, as will be shown below. Although the PEW study mainly captures the views of those who blog about their personal lives, in terms of general findings, they found that the blogging community is young (54% are 18-29 year olds), evenly split among men and women, racially diverse, and that bloggers are heavy users

of the Internet relative to non-bloggers. Over half (52%) of bloggers blog mostly for themselves, and about a third (32%) blog with an audience in mind. Most regular bloggers spend between 2-9 hours per week updating their blogs (Lenhart & Fox, 2006). Not much is known about the general makeup of the science blogging community, although there is good evidence that women are in the minority (Munger, 2011).

The picture that emerges from the literature about blog authors and their motivations is similar to the findings of this research: authors blog mainly for themselves, at least initially, and then become aware of and interact with their audiences; they utilise the community features of their blogs (comments, RSS and blogrolls) to interact; and that regular blogging requires some time commitment (several hours a week) from its authors. Another major finding of the studies above, also found in this study's participants, is the sense that people find blogging enjoyable—they enjoy the audience interaction and feedback, and are encouraged to write more when they have more interaction (Nardi et al., 2004; Hsu & Lin, 2008; Karimi & Poo, 2009). Evidence from the research above suggests that blog authors tend to write on particular topics—this is certainly the case with the science bloggers studied in this chapter, who stick closely to the topics of their interest and expertise and rarely deviate into personal diarising (unless they can relate it to a particular scientific topic).

5.2.3 Blog Network Analysis and Communities

We turn now to the evidence of community formation via writers and commenters on blogs. Most of the research in this area uses Social Network Analysis (SNA) to examine various permutations of hyperlinking in blog data collections (links in posts to other blogs, links inward to particular blogs, links found in blog comments, and blogrolls).

Chin and Chignell (2006) define blog communities as distinct sets of blogs which exchange hyperlinks while discussing similar topics. They used SNA to analyse blog communications in 604 blogs, seeding their investigation with blogs that linked to a popular music blog. Their findings were triangulated with a survey of bloggers in the group, and a case study of the music blog. Blog authors surveyed within the group identified positively with Chavis' (1986) *Sense of Community Index* (SCI) adapted for blog SNA, which identified features such as feelings of membership/belonging, influence, reinforcement of needs and a shared emotional connection. Chin and Chignell found that it was possible to map the SCI findings (membership, influence, reinforcement of needs and shared emotional connections) onto specific social network analysis indicators (reciprocal links, star networks, triangular connections) in their blog network, thus identifying potential communities (Chin & Chignell, 2006). Their study could in theory allow other SNA researchers to quickly identify blog communities by these indicators, on which further research could be performed. Similar feelings were identified among the research participants in this thesis, as detailed below, although the SCI could not be

applied as it is a quantitative measure and this research uses qualitative methods (Fisher, Sonn, & Bishop, 2002).

Tang, Wang, and Wang (2009) also investigate community formation in blog networks alongside information diffusion, that is whether blog networks disseminate information in particular ways. Their corpus consists of 21,066 posts from 1327 blogs from SOHU blogs (SOHU is a Chinese search engine company). They also identified community structures in the networks studied, which they validated via searches, manually checking blogs for mentions of other blogs within the community. They found that blog community members identified by the SNA were four times more likely to talk about each other than non-community bloggers. Tracking information diffusion through these communities proved more difficult, although they were able to show that particular topics could be associated more strongly with identified communities. For example, manual inspection of the blogs in a community blog network around the topic "China Tennis Open 2007" proved it to be made up of, as one might expect, tennis fans, athletes, and reporters (Tang, Wang, & Wang, 2009).

Gruhl, Liben-Nowell, Guha and Tomkins (2004) performed a network analysis on a large blog corpus (RSS feeds from 11,000 blogs, a collection of 401,000 posts), to discover patterns of the spread of ideas and topics through blogspace, based on an infectious disease model of information diffusion. Gruhl et al.'s analysis found that blog topic discussion is divided into "chatter" (ongoing discussion whose subtopic flow is largely

determined by decisions of the authors) and "spikes" ("short-term, high-intensity discussion of real-world events that are relevant to the topic") (Gruhl et. al, 2004, p. 491). They also detected what they termed "resonance", a large spike in postings which is prompted by a small event in the real world.

Of interest are their findings around the behaviour of individuals within blogspace, as they found evidence that some authors "provide a boost to every topic they post about—over time, these are the users who can have significant impact on a community, or even on blogspace overall" (Gruhl et. al, 2004, p. 500). In other words, some users are better than others at propagating information across a community.

These network analysis research findings echo and bolster qualitative studies of blogs and blogging, in particular demonstrating that bloggers are motivated by the social connections made by blogging, and that they tend to find like-minded communities with which to interact online via their blogs. This interaction takes many forms, one of which is the promotion of open science, which ties in to this thesis's theory that online tools are helping their users drive a SIM. This kind of connected community was described in the FriendFeed chapter—the Biogang all know and reference each other's blogs and other activities on sites like FriendFeed and Twitter, and over the past few years have interacted with each other almost daily.

Although this study does not cover a large enough corpus of blogs to reproduce Gruhl et. al.'s 2004 findings, the findings from this chapter

indicate that some writers are firmly embedded in the scientific community and referenced or known by all of those who identify as members of that community, while others tend not to have as many interactions, but may occasionally experience "spikes", not only in their own traffic, but as their post becomes a topic within the scientific blogosphere—such as Redfield's experience with the Arsenic story, detailed below.

5.2.4 Scholarly communications and blogs

There are variety of studies of scholarly blogging, several reflecting on their own or a small sample of their interviewees' blogging practices; (Ewins, 2005; Walker, 2006; Halavais, 2006; Efimova, 2009; Gregg, 2009a, 2009b; Kirkup, 2010; Price, 2010); a few attempting to provide models for categorising these practices (Efimova 2009; Kjellberg, 2009); and some using webmetric and bibliometric analysis to look at linking (Luzon, 2009; Thelwall & Kousha, 2008; Barjak & Thelwall, 2008).

Investigating the motivations and output of professional and academic blog authors, Ewins (2005) offered some personal reflections on the role of academic blogging and online identity, suggesting that hyperlinking can be an intellectual as well as electronic connection between ideas, that blogs can preserve continuity of research thinking as a venue for disparate thoughts which may one day expand into a new article or book, and that the archives of a blog can contain the provenance of an academic's intellectual reputation, by allowing others access to their "back catalogue" (Ewins, 2005, p. 369). Ewins also raised a note of caution with regard to the online

academic identity, as he states that by reading a researcher's blog, search committees can get to know what they believe to be "the real you"—which is only the online "you"—that is, they may be misunderstanding the blog author's online tone and representation of self. Similar cautions are raised by several of the authors, and are discussed further below.

Within this case study, research participants were unconcerned about these cautions, partly because many of them started a blog to preserve the continuity of their research thinking. As Beltrao, who has been blogging for eight years, states: "you can actually go back and remember what you were thinking about, and all my academic progress so far is actually online" (Beltrao interview). Others dismiss these concerns because for them, blogging is an integral part of their scientific career and their online discussions form an important part of shaping their research profiles as well as their open science activism profile. Murray-Rust uses his blog for experimenting with ideas as well as activism: "the reason I write a blog is (a) because I want to try out ideas, so it's a way of exposing ideas and seeing what happens (b) it's a means of advocacy" (Murray-Rust interview). Murray-Rust is a senior scientist and Reader in Chemistry at the University of Cambridge, who has found that blogging focused his attention on his open science advocacy, as he kept finding or hearing about events and initiatives and thinking "I should blog about that". Murray-Rust did not report any negative repercussions of his blogging activities, unlike Walker, below.

Walker (2006) reported anecdotally that blogging became more difficult for her personally the higher she rose in the academic hierarchy. In an earlier paper, Mortensen and Walker (2002) found an inversion of typical academic hierarchy as a professor's article about blogging was criticised by bloggers, and the professor, not having a blog of their own, chose to respond with a post written on one of their graduate student's blogs (Mortensen & Walker, 1999). Walker speculates, now that she has advanced higher in the academic system, that "perhaps blogging disrupts hierarchies that serve me well, and that makes me ambivalent about [blogging]" (Walker, 2006, p. 3). However, Walker also believes that blogging was beneficial in helping her reach her current status (tenured scholar and chair of a department), as she says:

> I know that my blogging helped me gain a foothold among researchers in my field, that the regular writing and discussions with readers and other bloggers helped me become a confident writer, and that I had more opportunities to give talks and write in other genres than most of my non-blogging peers. So quite probably, blogging helped me succeed in earning a PhD and getting my first academic job. (Walker, 2006, p. 7)

Walker is unsure about the value of blogging as an academic practice alongside the formal rubrics of publishing and conference papers, but hopes that it will be sustained and gain more value across time. This ambivalence is echoed a minority of comments posted on Redfield's blog in the "Arsenic Life" episode detailed below, which in many ways closely resembled the normally wholly private discussions about papers which take place in the hallways, lunch rooms and laboratories of scientists

worldwide. In general, the science bloggers in this study have a wholly positive attitude towards their blogging activities, and any ambivalence they felt about blogging mostly centred around finding the time to blog in a busy life schedule. Most of the research participants felt that the majority of their non-blogging colleagues and peers were indifferent to their blog activities, even if they read them.

Halavais argues that blogging is making "invisible colleges"—the social ties that "help drive the more visible institutions of research" discussed in Chapter Three, FriendFeed—more visible, by bringing what were formerly private discussions into the public sphere of the Internet" (Halavais, 2006, p. 117). Public in this sense means that the blog can be easily found by interested peers, for although blogs are essentially a public medium, they are still most often only found by those looking for particular topics, given that there are hundreds of millions of blogs.

Price (2010) reviews selected blogs within the field of anthropology, and makes three key points: firstly, that blogs provide "opportunities to explore ideas and arguments, share knowledge, and develop writing voices in a public arena without the controls of consensus and the gatekeeping of traditional publishing outlets"; secondly, that the time costs of blogging effectively are high, and lastly, a related point, that the "political economy of academia is not structured to reward individuals building things for a common good outside of the peer-review process", a point which articulates Walker's concerns expressed above (Price, 2010, p. 210).

Kjellberg's (2010) findings were similar. Kjellberg interviewed twelve Scandinavian academic bloggers, asking them about their blogging practices and their motivations for blogging. Reasons included "disseminating content, expressing opinions, keeping up–to–date and remembering, writing, interacting and creating relationships" (Kjellberg, 2010). From this research it seems clear that some definitions of academic blogging are emerging: it is used for sharing knowledge and ideas at the pre-publication level, and to develop writing skills. This usage also emerges in the findings of this research. However, other authors have provided further definition of the kinds of writing found in academic blogging.

Efimova (2009) uses an autoethnographical approach to deconstruct her own blogging practices over the length of her doctoral studies, which developed into a study of the blogging practices of academics, specifically the use of blogs as research diaries and idea-generators, a use which is mentioned by several of the participants in this chapter's study. Efimova found that a weblog can serve as both a personal information management tool and a space for developing ideas, which she divides into three stages, and illustrates with examples of blog activity which fit these stages. The stages are: *Awareness and Articulation* (short blog posts to capture ideas, being motivated to write better and contextualise one's writing); *Sense-making* (feedback from others to establish connections, looking for patterns in previous posts); and *Turning into a product* (producing longer posts to get into writing flow or create early drafts of parts of papers, welcoming

feedback from readers or asking for review of early drafts) (Efimova, 2009, p. 296). Several examples of these stages will be given when examining the science blogging practices of the research participants below. Efimova's autoethnographically-derived categories of blogging as research diary can be generalised to the science bloggers studied in this case, as is done in Section 5.4, below.

Kjellberg (2009) derived an analytical framework for examining the role of scholarly blogging in research activities, and tested it with an analysis of three physics blogs. Kjellberg identifies four main areas of analysis: aims (they vary, and several can co-exist within one blog), form (blog technology, as described above in Section 5.1.3, Definitions) content (which was highly heterogeneous) and context (the community being addressed). Kjellberg also acknowledges the weaknesses of this framework, which mainly fall around the finding that scholarly blogs (even in a sample of three) display a wide variety of aims, content and contexts, even as the form stays familiar. Other authors can fill the gaps here, as Efimova does above for types of blog content, as does Kirkup, below, and Luzón also found data to support varying aims of blogging via an analysis of hyperlinking in academic blogs. This thesis also considers the contexts in which the science bloggers operate, inspired by the STIN framework which accounts for the social and technological actors, explained above in the Introduction to this research.

Kirkup (2010) discusses a type of scholarly blogging which focuses on academic practice rather than content. Kirkup interviewed six bloggers

from his own institution, estimating that one in ten academics were actively blogging, either within institutionally-hosted blogs (his university hosts a blogging platform for staff and students) or externally-hosted. Kirkup's research found further support for the kinds of writing identified by Efimova's autoethnography, in that Kirkup found his interviewees writing to articulate ideas that were not quite developed enough for publication, or were too marginal for publication: to "get stuff out of one's system" (Kirkup, 2010). One of Kirkup's participants was a senior professor and active blogger, who found that his scholarly publication output had slightly declined since he began blogging actively, as the "online 'conversations' he was involved in with a large numbers of 'followers' and other bloggers, satisfied his need to engage with others about new ideas" (Kirkup, 2010, p. 81). The security of the professor's senior position meant that he felt comfortable using blogging as a central part of his academic knowledge production, and like many of the bloggers in this study, he enjoyed the directness and immediacy of the feedback from his online audience.

Luzón (2008), via an analysis of links on 15 academic blogs, determined that links are strategic for academic bloggers and serve to distribute information, contribute to identity construction and also to collaborative knowledge construction in that within blog posts, approximately a third each of the links were self-referential (23.12% to other blog posts within the blog, the author's work elsewhere; a third were to within the weblog community (30%, including 13% to other scholars' papers); and the

remaining links were external (46.87%, 30% to online news articles, institutions or other websites of interest. The examples Luzón uses to illustrate the links between bloggers are similar to Herring et al.'s findings about blogging topics, and to my own unpublished study of 335 science blog posts, which also examined topics instead of links, but found similar numbers of blog posts using external data such as external news articles or multimedia (48% to Luzón's 46.87%) (Power, 2008).

Several of the authors of the papers above made reference to the negative effects of blogging, either anecdotally or directly reported by their interviewees. Gregg (2006), Halavais (2006), Dennen (2008) and Ewins (2005) mention a Chronicle of Higher Education column[80] in which the pseudonymous author, Ivan Tribble, maintained that blogging could threaten an academic's chances of being hired, as there was a perception that bloggers might "air a department's dirty laundry" (Tribble, 2005).

Kirkup (2010) provided specific examples of bloggers being asked to retract particular posts or parts of posts by their departments, negative comments by colleagues or bosses, and self-reflection that it was not a positive practice: "'Mr A' remarked of his blog: "There is no chance of a Chair once you start this kind of writing" (Kirkup, 2010, p. 82). In this chapter's study, each research participant in this case was asked questions which probed this issue, and the participants reported mostly positive or

[80] Tribble, I. (pseudonym). Bloggers need not apply. *Chronicle of Higher Education*, (2005) 8 Jul., Available at: http://chronicle.com/article/Bloggers-Need-Not-Apply/45022/(accessed 12 July 2011).

indifferent responses from peers and superiors, although this was not universal. Their positive and negative experiences are detailed below.

Collectively this research demonstrates that the Internet is having an effect on scholarly communications, inasmuch as scholars who maintain blogs and personal websites show a positive correlation with research productivity (Barjak, 2008; Vakkari, 2008). Additionally, the qualitative studies discussed above demonstrate that academic bloggers produce an identifiable genre of academic blogging, and write about academic practice as much as their research. Similar to the general blogosphere's personal diaries, but focused on research, some scholars blog mainly to keep a research diary for their own records, and some (once they sense they have an audience) value their online interactions as an extension or addendum to their offline scholarly activities such as formal publications and attending and presenting at conferences. Several scholars (Halavais, 2006; Kirkup, 2010) have commented that scholarly activity in the blogosphere also constitutes a thriving online invisible college through the active discussions taking place online. However, the small scale and autobiographical nature of many of these studies means that their findings are not generalisable to all scholarly bloggers, although many of their findings are similar with regards to motivations to blog and topics found in academic blogs. A few authors commented on the perception that blogging can be detrimental to an academic career, although there is little hard evidence to suggest this within the present study. Scholarly bloggers are, however, making

scholarly knowledge public and freely accessible, and enabling discussion around that knowledge while bypassing traditional gatekeepers of knowledge.

5.2.5 Science Blogs

Scientists who blog about their own and others' research are increasing in number, although it is difficult to estimate the exact numbers of "science blogs", as firstly one has to define what that means—for example: what a scientist is, what topics constitute science blogging, what constitutes an active blog in terms of updates. An estimate from 2007 put the number of science blogs at 1200 with growth being particularly facilitated by free aggregations of science blogs supported by various publishers (Bonetta, 2007). The site ResearchBlogging.org, which allows bloggers to self-register if it can be confirmed that they are blogging about peer-reviewed research, has 2000 registered blogs (Munger, personal communication, September 19, 2011). Scienceseeker.org, established in 2010, allows any blog author who considers their blog a science blog to register. They currently have 810 registered blogs.[81]

However, research specifically on science blogging is sparse. Bonetta (2007) interviewed ten scientific bloggers and provided a high-level overview of the scientific blogosphere, and found the primary motivations for blogging among her sample to be public education and outreach. A more recent overview of the state of scientific blogging interviewed other

[81] http://scienceseeker.org/blogs/, Accessed 14 July 2011

prominent scientific bloggers and called for more institutional support of scientific blogging as a method of public outreach and education (Batts, Anthis, & Smith, 2008). Both Batts et al. and Bonetta cited the same example of a blog as a useful public peer review mechanism (Cartwright and Comai's 2005 joint publication in *The Plant Cell*), indicating that there is scope for more in-depth research of the impacts of science blogging. Other examples of post-publication peer review have emerged since then—the Arsenic Life story is told within this chapter, and examples have appeared in mathematics, genetics and chemistry (Neylon, 2010b). Both reports were written for a scientific audience (they appeared in *The Plant Cell* and *PloS Biology*), but are reportage rather than in-depth analytical or empirical research.

This reportage style is the most common form of literature on scientific blogging. Wilkins (2008) provides an overview of science blogging for scientific readers in the journal *Trends in Ecology & Evolution*. Wilkins sees blogging as a "more intimate and responsive" way to communicate science than the popular scientific media, and also thinks it can play a "demythologizing" role by "correcting or avoiding the myths and pigeon-holes of science journalism" (Wilkins, 2008, p. 411). Wilkins cites the benefits of blogging as: potential to widen one's social network by making new contacts; gaining jobs via these new contacts; overcoming the "two-cultures divide" between the sciences and the humanities identified by CP Snow in the 1950s; to allow a platform for advocacy in areas such as open

access; and, if preserved, to provide a record of debates and discussion for future historians of science (Wilkins, 2008). Regarding the costs of keeping a blog, he cites the variable quality of blogs as information sources, particularly the proliferation of pseudoscience blogs, as well as the economic costs to scientists of keeping a blog, in terms of there not being any obvious income-generation stream (Wilkins, 2008). Wilkins does not mention the main cost cited by bloggers interviewed for this case study, that of the time cost of writing blog posts and keeping up with relevant discussions on one's blog and elsewhere on the Web.

Goldstein (2009) describes the features of blogs for an academic audience in the *Evolution: Education and Outreach* journal, focusing on descriptions of several blogs about evolutionary biology. Goldstein identifies four blog categories within this topic area, which he characterises as "professional" (written by academic writers and teachers, lacks a personal voice and anecdotes, and does not write in the first person), "amateur" (still professional scientists, often discussing the research literature, but more personal, writing about their passions about the topic), "apostolic" (strong first-person voice, likely to engage in written battles with foes—opponents of their stances or anti-evolutionists), and "imaginative" (looking at evolutionary science from the point of view of the visual arts, music, or fiction) (Goldstein, 2009, p. 548). Goldstein's conclusion is that blogs are of limited value in providing structured and comprehensive coverage of the science of evolution, but they are useful for identifying the current topics of

interest to the field and stimulating ideas about new research directions. Goldstein's four categories can also be found in the blogs studied for this chapter, mainly falling into the professional and amateur categories, although for some an extension of the "apostolic" category could be considered—for although they do not tend to engage in battles with opponents, they do write strongly about open science topics, with a view to converting readers to their points of view.

In an empirical study of science blog content, Groth and Gurney (2010) used bibliometric methods to discover how blogs were being used to discuss peer-reviewed research. Using a sample of 296 blog posts from the Chemistry section of Researchblogging.org, they took measurements of the age of journal articles being discussed, the impact factors of the journals they came from, and the topics being discussed via blog author-assigned keywords. Their main findings were that scientific discussion on blogs is immediate—very few of the posts (<18%) discussed research that was more than a year from publication. They also found that authors tended to discuss research from high-impact journals, both in chemistry and science overall (over 90% of publications). Lastly, their keyword and title analysis indicated that the authors were interested in conveying the broader implications of the research, in that "blog post word choice [was] more oriented to current social, political and medicinal topics, whilst explaining the underlying sciences behind each topic in more detail in the blog" (Groth & Gurney, 2010, p. 6). Researchblogging.org is discussed in more detail in

Section 5.5 of this chapter, Blog Networks and Aggregations, and the evidence from research participants in this study is that they are also inclined to discuss recently published research within a framework of current social or political topics.

A small-scale study by Shema and Bar-Ilan (2011) looks at the demographics of bloggers posting to Researchblogging.org, selecting 133 bloggers across 126 blogs who had a minimum of 25 posts about peer-reviewed research indexed by Researchblogging.org across 2010 (Jan 1 2010 to Jan 15 2011). Their findings included: a gender disparity of over two-thirds male bloggers to less than one-third female bloggers (5.56% were unknown), which mirrors Munger's findings (Munger, 2010); a small proportion of anonymous bloggers (18% of the sample); 69.05% independent bloggers in contrast to 30.95% bloggers belonging to networks such as Scienceblogs.com; and, like Groth and Gurney (2010), they found that most posts about peer-reviewed research referred to recent publications in journals with high impact factors, many of them multidisciplinary journals like *Nature* and *Science*.

Kouper (2010) analysed 174 posts and 1409 comments from eleven science blogs selected at random from a Google search, and examined them for what was called "modes of participation"—that is, several categories of information dissemination, such as informing readers of scientific news, explaining science to a lay audience, evaluating published research and explaining the author's position on an issue they deem controversial.

Kouper's conclusion was that "science blogs are too heterogeneous to be understood as an emerging genre of science communication" (Kouper, 2010, p. 3).

The response of parts of the scientific blogging community to Kouper's article was vociferous and critical. In summary, they argued that the sampling strategy was not articulated or justified, the sample sizes were not suitable to make generalisable conclusions from, and that several statements made about the bloggers and blogs were not empirically supported. This episode is interesting in itself, as it demonstrates the point argued later in Chapter Six, Synthesis, that this community is self-aware and keeps track of, and indeed participates in, scholarly research about its own activities. Kouper's study, although correctly identifying several types of scientific blog information dissemination, drew general conclusions about science blogging from a small sample of the science blogosphere, and I would argue that her conclusion does not take account of the complex interactions within, and the socio-technical context of, the current science blogosphere. de Moor and Efimova (2004) also argue that blogging is strongly contextual and situated, in that drawing conclusions about whole genres of blogging from blog data should take into account the blog's history, specific hyperlinking practices of each blog, and conversational practices of each blog (de Moor & Efimova, 2004).

The data presented in the remainder of this chapter use specific and contextualised incidents and highlights from scientific blogging practice to

illustrate the complexity and variety of the research participants' online social interactions. This chapter also expands on what is already known about science bloggers—we already roughly know how they are constituted (mostly men, mostly not anonymous, mostly independent of blogging networks)—to create a richer portrait of current science blogging. It describes the place of blogging within scientists' quotidian routines, details their motivations for writing and their interactions and feelings about their audiences, and captures incidents that illustrate the current volatility of science blogging ("PepsiGate") and the position of blogging as a form of post-publication peer review ("Arsenic Life").

5.3 Research Participants

Twenty scientists have been interviewed for this blogging case study. 91% (20/22) of the total participants across all three case studies also maintained a blog in addition to their FriendFeed or OLN activities, and every participant was asked the same set of questions about their blogging practices (See Appendix E, Interview Questions).

Nineteen participants are male, and one is female. Their work experience ranges from current doctoral student (3), recent postdoctoral fellow (4), mid-career scientist or professor (10) and former scientist (3). They all hold doctorates in a scientific field, or are working towards their doctorates. Four are professors within their fields. Nine participants use a free blogging platform, Blogger.com or Wordpress.com. Six host their own blogs on their own domains, and two are current members of a blog

network. Bora Zivkovic is employed full-time in science blogging, as editor of the Scientific American blog network. All of the participants blog using their real names.

On average, these bloggers create 2-4 posts per month each, although this can vary widely depending on their work, as those whose job is science blogging, or those for whom blogging is central to their online strategy (for example Mesko and Zivkovic) often post more than once a day. There is also "seasonal" variation, in the sense that some people are more likely to find time to post at conferences, or when the academic year is quieter, whereas some are more likely to post when their general work activity streams are higher.

Although most of the scientists interviewed across all cases of this thesis also have a blog, not all of them feature in this chapter, as some of them blog infrequently and use other online tools such as FriendFeed or Twitter for most of their online interactions. This chapter highlights practices which illustrate the variety of approaches to and uses of blogging. This variety is already evident in the wide range of practices found by the existing research outlined above. The sample is not designed to be representative of the experiences or practices of all science bloggers, but rather to complement the cases of FriendFeed users and open notebook scientists by exploring blogging, the third interlinked component of their use of online tools to facilitate scientific research and communications. That said, some scientific bloggers featured in this chapter are not FriendFeed users or open

notebook scientists, but all have some involvement or interest in open science, and episodes in their blogging history serve to illustrate wider discussions and controversies taking place around open science, open data, and scientific communications.

Table 11

Research Participants (n = 20)

Name	Blogroll size[82]	Posts/Month (when active)	Blog Address
Nick Anthis	98	3	Scienceblogs.com The Scientific Activist (no longer active)
Pedro Beltrao	44	1-2	Blogger.com http://pbeltrao.blogspot.com/
Jean-Claude Bradley	20	7.6	Blogger.com http://usefulchem.blogspot.com/
Richard Grant	0	4-5	Network - Occam's Typewriter http://occamstypewriter.org/rpg/
Bill Hooker	0	3	Self-hosted http://sennoma.net/
Lars Juhl Jensen	0	3	Wordpress.com http://larsjuhljensen.wordpress.com/
Martin Johnson	4	4	Wordpress.com http://martwine.wordpress.com/
Steve Koch	0	2-4	Blogger.com http://kochlab.blogspot.com/ http://stevekochscience.blogspot.com/
Andrew Lang	0	1	Institutionally-hosted http://webapps.oru.edu/new_php/ academics/faculty_blog/?user_id=236
Pierre Lindenbaum	0	6.4	Blogger.com http://plindenbaum.blogspot.com/
Berci Mesko	92	30.8	Self-hosted http://scienceroll.com/about/
Cameron Neylon	0	4	Self-hosted http://cameronneylon.net/category/blog/
Rosie Redfield	3	27.2	Blogger.com http://rrresearch.blogspot.com/ Moved to Field of Science network in July 2011; same URL (FoS is able to import Blogspot blogs)

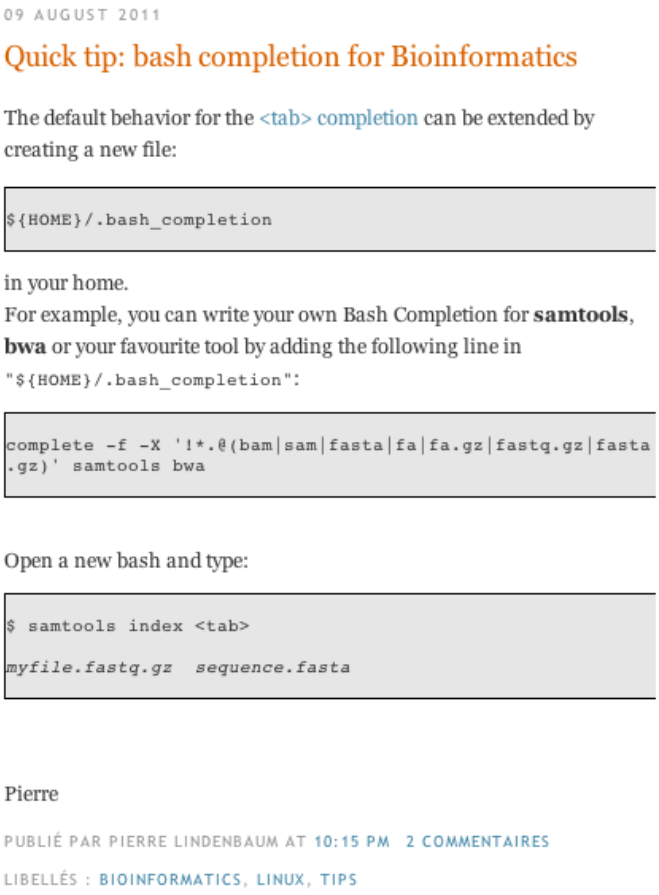
[82] Blogrolls, defined on p. 285, are the numbers of links to other people's blogs.

Name	Blogroll size[82]	Posts/Month (when active)	Blog Address
Peter Murray-Rust	0	4	Institutional domain; self-hosted http://blogs.ch.cam.ac.uk/pmr/
Anthony Salvagno	4	5	Blogger.com running on own domain http://gradness.iheartanthony.com/
Neil Saunders	0	4.8	Wordpress.com http://nsaunders.wordpress.com/
Deepak Singh	31	12	Self-hosted Business, bytes, genes, molecules http://mndoci.com
Pawel Szczęsny	0	4	Self-hosted http://www.pawelszczesny.org/
Mat Todd	0	3-4	Wordpress.com http://intermolecular.wordpress.com/
Bora Zivkovic	50	35.6	Currently on Scientific American blog network (also Blog Editor of network) Recently self-hosted http://blog.coturnix.org/, has been a member of ScienceBlogs.com

5.4 Participant's Science Blogging

This section will discuss the blogs of the participants: how they got started with blogging, what originally motivated them and whether that has changed; what the audience means to them; how blogging fits into their scientific work routines; and what they blog about and why. These sketches of blogging scientists add to the literature on blog authorship, scholarly communications and science blogging. To understand the size and scope of participant's science blogging and interactions, it is useful to review of how this blog interaction typically happens. Authors write a blog post on a topic, sometimes inspired by a post on another blog, someone else's tweet, a conference they have attended, or a paper they have read. Other topics include short essays about new ideas in work or open science, or explorations of current work they are doing. These idea-generating posts

can be matched to Efimova's three stages of idea development discussed in the literature review above` — *Awareness and Articulation, Sense-making* and *Turning into a product* (Efimova, 2009).



09 AUGUST 2011

Quick tip: bash completion for Bioinformatics

The default behavior for the <tab> completion can be extended by creating a new file:

```
${HOME}/.bash_completion
```

in your home.
For example, you can write your own Bash Completion for **samtools**, **bwa** or your favourite tool by adding the following line in "${HOME}/.bash_completion":

```
complete -f -X '!*.@(bam|sam|fasta|fa|fa.gz|fastq.gz|fasta
.gz)' samtools bwa
```

Open a new bash and type:

```
$ samtools index <tab>
myfile.fastq.gz  sequence.fasta
```

Pierre

PUBLIÉ PAR PIERRE LINDENBAUM AT 10:15 PM 2 COMMENTAIRES

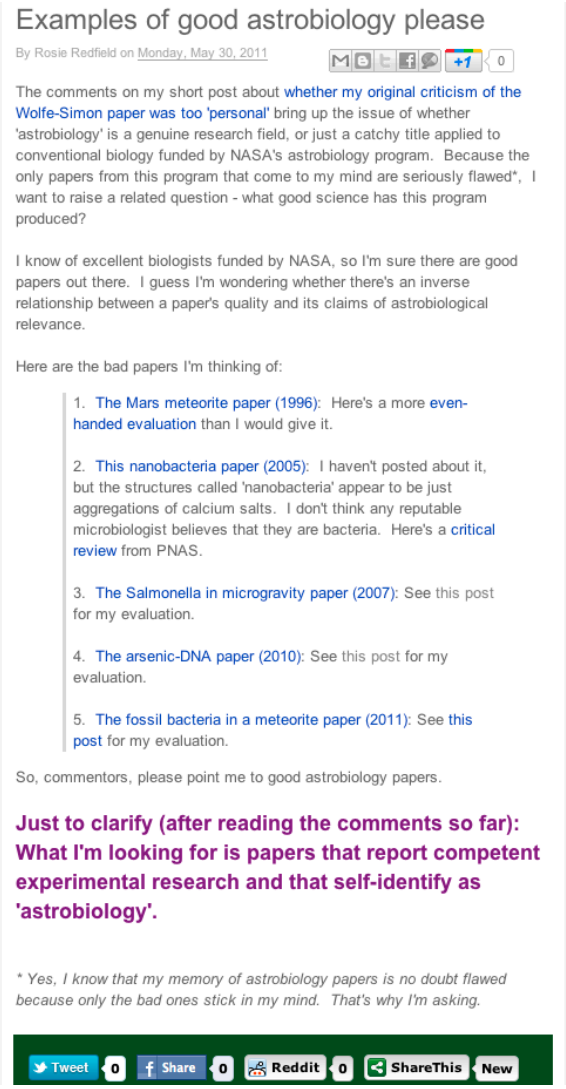
LIBELLÉS : BIOINFORMATICS, LINUX, TIPS

Figure 33. Idea-saving in a short blog post[83]

Lindenbaum mainly uses his blog as a research notebook: "every time I learn something new or I write something geek/cool, I write my notes on my blog" (Lindenbaum interview). The screenshot from Lindenbaum's blog exemplifies a simple *Awareness and Articulation* blog post, from Efimova's

[83] http://plindenbaum.blogspot.com/2011/08/quick-tip-bash-completion-for.html

categories. It creates a note which helps the writer remember a particular bioinformatics technique, including the actual code.



Examples of good astrobiology please

By Rosie Redfield on Monday, May 30, 2011

The comments on my short post about whether my original criticism of the Wolfe-Simon paper was too 'personal' bring up the issue of whether 'astrobiology' is a genuine research field, or just a catchy title applied to conventional biology funded by NASA's astrobiology program. Because the only papers from this program that come to my mind are seriously flawed*, I want to raise a related question - what good science has this program produced?

I know of excellent biologists funded by NASA, so I'm sure there are good papers out there. I guess I'm wondering whether there's an inverse relationship between a paper's quality and its claims of astrobiological relevance.

Here are the bad papers I'm thinking of:

1. The Mars meteorite paper (1996): Here's a more even-handed evaluation than I would give it.
2. This nanobacteria paper (2005): I haven't posted about it, but the structures called 'nanobacteria' appear to be just aggregations of calcium salts. I don't think any reputable microbiologist believes that they are bacteria. Here's a critical review from PNAS.
3. The Salmonella in microgravity paper (2007): See this post for my evaluation.
4. The arsenic-DNA paper (2010): See this post for my evaluation.
5. The fossil bacteria in a meteorite paper (2011): See this post for my evaluation.

So, commentors, please point me to good astrobiology papers.

Just to clarify (after reading the comments so far): What I'm looking for is papers that report competent experimental research and that self-identify as 'astrobiology'.

** Yes, I know that my memory of astrobiology papers is no doubt flawed because only the bad ones stick in my mind. That's why I'm asking.*

Tweet 0 Share 0 Reddit 0 ShareThis New

Figure 34. Sense-making blog post[84]

Redfield's post above asks for examples of astrobiology papers from readers, and provides an example of Efimova's *Sense-making* category, as it solicits clarification on the topic of astrobiology. It has 25 comments, which is a higher than average number on Redfield's blog, as most posts receive 0-

[84] http://rrresearch.fieldofscience.com/2011/05/examples-of-good-astrobiology-please.html

1 comments, although there are some notable exceptions, such as the Arsenic Life blog post discussed below in Section 5.7.

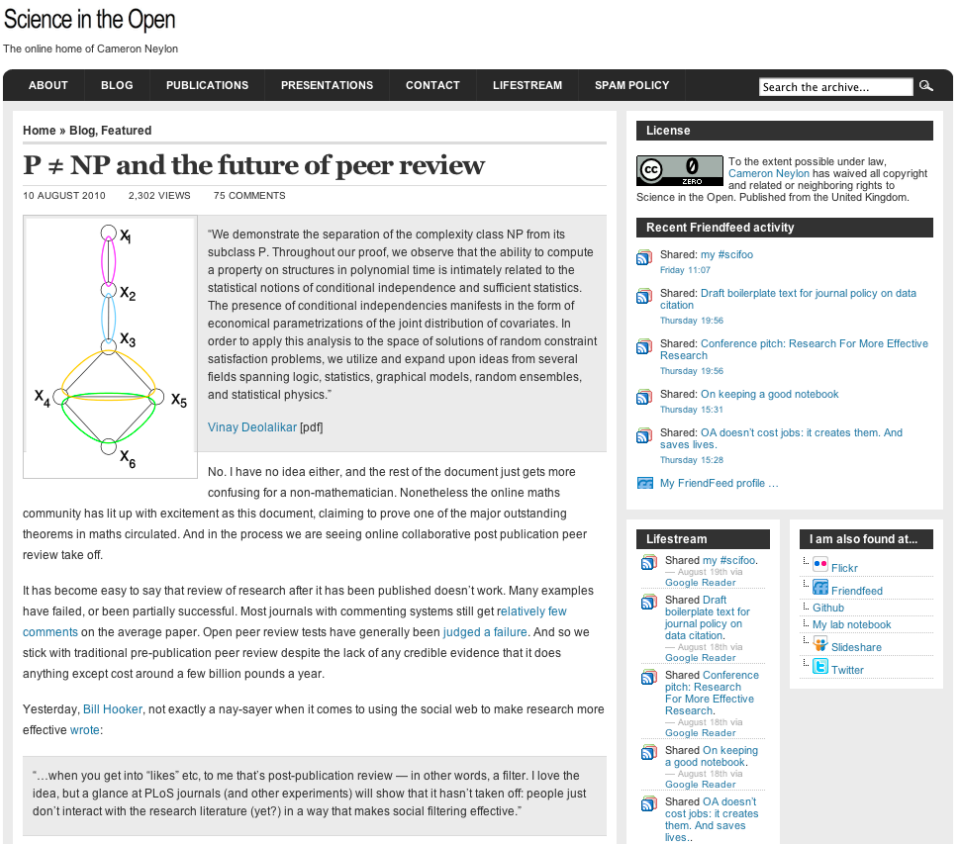


Figure 35. Turning into a product

Neylon's post above is discussing an instance of what he terms "web scale" peer review of a mathematics paper (Neylon, 2010b). It is 1,200 words long, has had 2,302 views (Aug 30 2011) and 75 comments (including retweets). On FriendFeed, it was liked by 35 people and commented on 30 times. It fits the category of long-form discussion, or writing longer posts which invite discussion from readers, called *Turning into a product* by Efimova.

After they press the "publish" link in their blogging client, their post will often be excerpted on FriendFeed and a link will be automatically Tweeted:



Figure 36. Blog post link in Twitter (Neylon, 2011a)

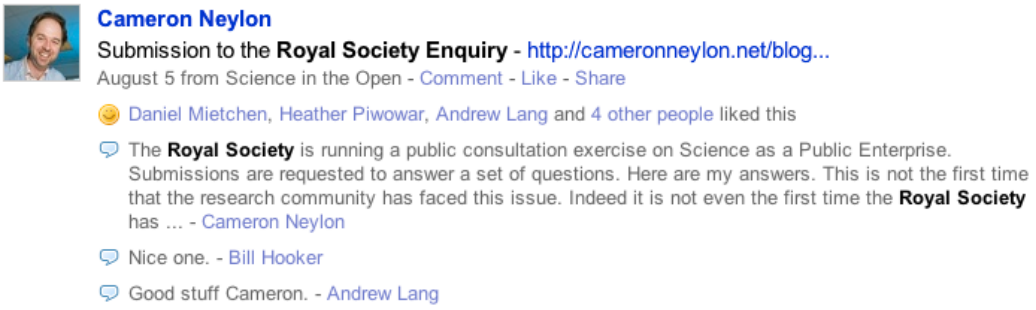


Figure 37. Blog post link in FriendFeed (Neylon, 2011b)

In the case of a blogger who does not use FriendFeed, the post will appear in the RSS readers of those who read the blog, and will be findable by search engines within a day or two (although the time taken to be added to the index will vary). Depending on the content of the post, discussion will take several forms: retweeted (the link is tweeted by others who read it and liked it but who may, or may not, have left comments, equivalent in some ways to "liking" in FriendFeed—see Figure 36); commented on at the blog post URL; or commented on at FriendFeed.

Blog authors choose whether they engage in the commentary discussion taking place—some will actively participate, others will note corrections or changes or add acknowledgements, and some will never engage with their

commenters (although they are using email for offline discussions, which Lang and Bradley both mentioned in interview). The participant's blogs and FriendFeed are often strongly interconnected, and for some, FriendFeed was responsible for growth in their online interactions:

> You know, you can look through the posts and there's not very much in the way of comments on any of them. And so until FriendFeed came along I didn't get a real lot of feedback. I didn't really think about that at all. ... I do get quite a lot more immediate commentary type feedback from FriendFeed than I ever got on the blog itself. And that's been very valuable ... I think of blogs as a conversation, I always have. I think of FF as just a way to augment that conversation. (Hooker interview)

This discussion can even happen without the blogger being aware that their work is being discussed:

> So I find quite a lot of comment on my blog on FF, it doesn't occur on my blog. Until I was alerted to that I didn't realise that. [on FF] they point to the blog and they say "PMR has overstepped the mark here" or "this is a good point" or just that they liked it. (Murray-Rust interview)

The nature of blog commentary and discussions depends on the relationship of the blogger with their audience, but for most of the participants in this research, discussion is similar to the interactions described in the FriendFeed chapter. A notable exception to this is discussed below, in the "Arsenic Life" section. One difference in the type of discussion on blogs compared to FriendFeed is that the authors will often receive email from those reading their blogs. Redfield and Zivkovic both commented that email from peers in their field has alerted them to the fact that their blogs are being read by people within their field.

Within the sample of this research the reasons for starting to blog varied greatly. Some began blogging on other topics, like politics, and then moved towards blogging about their science and science policy topics. Hooker "started out talking a lot more about politics" but as his interest in open science grew he gravitated to science topics (Hooker interview). Zivkovic was heavily involved in political blogging: "I wrote a political blog until November when Kerry/Edwards lost. And then I wrote three or four kind of postmortem blog posts ... but by this time I was already aware of several early science blogs" (Zivkovic interview). Singh began his first blog "because I was stuck in a hotel room with nothing to do … I realised that the posts that seemed to get the most attention, that people seemed to like, were the scientific ones. And those were the ones I also enjoyed writing the most" (Singh interview). Growing awareness of the online scientific community, and responses from that community sustained Singh, Zivkovic and Hooker's interest in science blogging.

Others began blogs as part of their Open Notebook Science approach, and although they adopted other platforms like wikis for their open notebooks, their blogs continued as discussion spaces for their work. Neylon began blogging to have a platform for discussions around his already-established open science work, whereas for Bradley a Blogger blog was his first OLN: "I started a blog, that's the UsefulChem blog, where I chronicle how I found the various projects and how we met all our collaborators and all of this, but the idea was basically simple—to just do

science transparently ... the only software I knew of at the time to communicate with people was blogging" (Bradley interview). The Blogger blog was replaced by the UsefulChem wiki, but Bradley keeps the original blog for announcements, sharing data such as conference presentations, and discussion of OLN data and research papers.

Expanding their teaching practice was another reason given: "initially I wanted a place where I could talk about chemistry things, … things kept coming up that I thought would be really interesting to expand on and explain to students" (Todd interview). Some blogs began as research notebooks and evolved into ONS platforms, while Grant, who had an institutional blog at the University of Sydney, was approached by Nature networks to blog for their network. Grant found his institutional blogging a good way of sharing his scientific practices: "anyone can do it. So anyone can say, hey, look, I've got this great experiment, or hey, I'm having problems with this Perl code, or whatever. You just suddenly start sharing things. Maybe not reagents, maybe not physical things, but you can share code, you can share methods, and I did a lot of that" (Grant interview). Murray-Rust was approached by his graduate students, who suggested he should start blogging.

The origins of research participants' motivations for blogging are also varied. Many started their science blogging as part of a research notebook, to keep track of their ideas, a theme also found in earlier studies of academic blogging (Ewins, 2005; Efimova, 2009; Price, 2010; Kjellberg,

2010). Hooker states: "I guess I mainly write the weblog to organise my own thoughts. First and foremost I do it for me. After that, I've found that some people are interested in the same things as I'm interested in, and some of the stuff that I write of is of use or interest to them, so that's another incentive to not just keep it on my own hard drive. I get some good feedback, and I've met some really great people through it" (Hooker interview). Koch states almost identical motivations: "to write a blog you collect your thoughts, and even if no-one read it, it would be good. And then there's a chance that someone's going to read it and benefit from it" (Koch interview).

Despite having kept a blog for more than six years, Singh still thinks the primary motivation should be your own, rather than audience-building: "I still think you write a blog for yourself as much as anybody else. I've tried writing for an audience and it doesn't work" (Singh interview). Similarly, Saunders also started writing for himself: "I thought of it as like a research diary. Just a sort of way of jotting down thoughts that I was having about research ... And then the surprising thing is that people start commenting" (Saunders interview). Bradley said of his UsefulChem blog: "I'm going to chronicle what I'm doing here, if I never get a visitor, I'm OK with that, because I was just building a record of what we were doing" (Bradley interview). In Bradley's view, the blog is a complement of his OLN work, providing a space for discussion of the methods and results detailed in the online notebook.

Avoiding waste

For Juhl Jensen, the blog is a place to put small ideas or projects that did not develop into papers but could be of use to someone else: "I do lots of stuff where I do a little analysis, I check some things and the results might be interesting to some people, ... you know, it's something that I spent one afternoon doing and I'm not really going to bother trying to make a peer reviewed paper out of it. ... I started thinking that it's a bit of a shame that I did those things and then it sort of died in my hard disk" (Juhl Jensen interview). This was echoed by Johnson, who says: "it's all useful kind of knowledge for ... the technical details of how you got the method to work or what have you. All that kind of lost knowledge, just as useful to have out there" (Johnson interview). In the sense that they use their blogs to preserve small amounts of useful information that are not big enough to put in a paper, these aspects of their blogs serve the same purpose as Bradley's OLN provides data on solubility measurements-small but useful pieces of scientific information which can help reduce unnecessary reproduction of results.

Having an audience

For Mesko, the audience is a key motivator, although he believes that blogging for an audience involves working everyday to keep that audience interested: "even if it takes plenty of time, because it does, I have to blog no matter [what]—as an example, when I was a medical student, or whether it's a national holiday, or Christmas—I have to blog". But for him the effort

has been beneficial: "I go to conferences internationally and talk about not just Web 2.0 and medicine but also Web 2.0 and science, because I think it's absolutely positive. Just because I have a blog, I've been invited to dozens of conferences, from California to the World Health Organisation" (Mesko interview). Redfield would also like to cultivate an audience, as her goal in blogging is to inform others about how science is really done: "I have this vague audience in mind ... I aim what I write [at] maybe an advanced undergraduate level" (Redfield interview).

However, the audience is often not the prime motivator, particularly for those who view their notebooks as research notes: "I'm not doing it for an audience really. I'm doing it because it's sensible and hopefully somebody – yeah, if one person gets something useful out of it and it makes their experience go better in the future, then it's kind of been worth it" (Johnson interview).

Bradley characterises writing for oneself and writing for an audience as quite different activities:

> I never really expected people to follow me. ... I think it's a completely different type of interaction if you say "I really pray I'm going to get, you know, 100 comments on this", whereas if you're saying "Let's not focus on how many or which kinds of comments you're going to get"—just focus on writing it as truly as you can to the point that you're trying to make. (Bradley interview)

Others find rewards from building an audience, whether deliberately or unconsciously, are of a longer-term nature than just reading and responding to blog post comments: "It's not the immediate feedback of

comments and conversation, but the longer-term feedback of contacts and people [that are valuable]" (Hooker interview).

As with Nardi et al.'s (2004) "studied minuet" between blog author and audience, the influence of the blog readers on the blog author varies widely: some bloggers are pleased to find they have an audience, while others write purely for themselves, keeping a research notebook, but most dance back and forth between their own and their audience's needs:

> The best thing was having someone like something I wrote and seemingly learning or having some kind of benefit from it. Another good thing is it's a permanent record - sometimes I link to my blog to answer some question. (Koch interview)

Interaction with readers and blogging for one's own research record are therefore not incompatible.

In relation to the perceived risks associated with blogging noted by Gregg (2006), Halavais (2006), Dennen (2008) and Ewins (2005), most of the research participants, while claiming not to feel like there is any risk, have considered how their employers could react to their blogging activities. Most inform their employers about their blog, some explicitly avoid blogging about their work, or more specifically people and incidents within their workplaces, while others are unconcerned about this.

Showing laboratory life

Redfield and Grant wanted to show the messy, human side of science to their readers. Grant states that he started blogging "for the University of Sydney ... because I wanted to show the human side of science" (Grant interview). Redfield's aim was similar: "I think it would be nice to have a

bigger audience ... I am motivated by the desire to make science more open, make people see that science can be more open and if nobody reads my blog, of course, they don't realise this, so getting a bigger audience is good in that way" (Redfield interview). Note that this approach may not be without risk; as is shown below in the Arsenic Life controversy around a post on Redfield's blog, revealing the mundane aspects of science may create fuel for those who wish to discredit scientific findings, who can point at dissent and argue that as a result, all scientific findings are inherently contentious and unstable. Most scientists would agree that contention and instability are signs of the normal practice of science.

Daily routines

The daily routine of blogging varies widely among participants. For some, blogging is a must-do daily activity that they fit in before or after work, or on their commute. "If you stumble even for a few days you will lose half of your readers. That's what the Internet is about. You have to update or they will leave you and I totally understand it" (Mesko interview). Although feeling less pressure than Mesko to blog, Salvagno also thinks of his audience: "I have at least a small reader base, so I feel like I let my imaginary fans down when I don't post anything!" (Salvagno interview). None of the bloggers in this research tend to compose their blog posts at work during their day jobs, they see it as something they do in their own non-work time, apart from those whose work is blogging, like Bora Zivkovic, who is employed full-time as Scientific American's blog editor.

For others, blogging is something they will do infrequently or at least not scheduled: "I don't try to blog at any frequency—it's just basically if there's something, if there's a milestone, something of interest to talk about, obviously if I go to a conference I'll tend to blog about it" (Bradley interview). Not trying to keep to a posting schedule seems to eliminate feelings of guilt about not blogging. Kjellberg (2009) found that the blogging physicists in her study made similar comments about time, and apologised to their audiences when they were unable to blog regularly.

Participants who had kept a blog for some time talked about their fluctuating patterns of blogging, and most mentioned feeling some kind of guilt at not "keeping up" with their own personal expectations of blogging. This guilt was particularly connected with those who were conscious of having an audience. For example, Koch mentions: "for the last couple of weeks I've been saying "I really should blog about our results" because we have new results, and they're interesting, and I really want to. I've just, I haven't been able to prioritise it I think" (Koch interview). For some, it also takes time to get from rough draft to something they consider publishable: "I have posts stacked up, which have been drafted, but not published yet. It's just a question of finding the time" (Todd interview). Similarly, for Redfield, waiting to finish posts makes them more of a burden: "I've got this blog post that's just sort of been hanging around for a month and a half, waiting for me to finish it ... I let it get too important in my mind and that makes it harder to finish" (Redfield interview). This aspect of academic

blogging may provide an explanation for the feelings of burden experienced, as time taken to write a blog post is time taken away from writing academic papers.

Several bloggers commented on how they can get "caught up" in the cycle of posting at a certain frequency, and feel guilt if they cannot sustain that frequency. Singh likens it to an obsession:

> There was a time, back in the day, where if you didn't write a blog post every day you felt - you put too much stress on yourself because you're trying to [write]` — I think it happens in the life of every blogger, the blog becomes big enough so you feel like you have to keep doing something. And then you realise after a while - when you have enough cachet, it doesn't matter. So it's sort of a peak, then you trough and then you flatten out in a comfortable place. (Singh interview)

The dimensions of academic blogging described above—not blogging at work unless it can legitimately be considered part of your job, feeling internal pressure to continue blogging, but finding it difficult to block out time to blog—have not previously been investigated in the research literature. The findings show that scientists regard their everyday work separately to their open science activism, and it is only when work and open science intersect—such as with Redfield, Todd, Bradley, Neylon and Koch, who all keep OLNs, described in Chapter Four—that they seem comfortable with blogging overlapping with their working hours. That said, many of them still do prefer to blog in their own time, when they can find a quiet hour or two on the weekend or in the evenings, because composing a blog post can be a writing activity which takes time, much as writing or reviewing a paper would.

5.5 Blog Networks and Aggregations

Considerable numbers of science blogs are now hosted within blog networks—individual or multi-author blogs sharing a common blogging platform hosted by a single entity (publisher or collective). Many of these have only existed for 3 years or less, and several (for reasons explained below) were created in late 2010. Many more science blogs (both those from networks and the independently hosted) are indexed by an aggregator, which collects the RSS feeds - either by finding blogs itself or by the blog authors registering with the site. Both networks and aggregations are attempts to help readers find science blogs, which will in turn drive traffic to the blogs. For some, traffic is an end in itself, as advertising-supported sites require audiences to generate revenue, and for others, traffic growth is a way to market their offline publications and build brand prestige (or not - see the discussion below about PepsiGate on ScienceBlogs.com).



Figure 38. Screenshot of ScienceSeeker.org home page, July 2011

Blog aggregations are groupings of science blogs based upon certain criteria. Scienceseeker.org is an aggregation of science blogs developed by six volunteers (Scienceseeker.org, 2011). People who consider themselves a science blogger may register their blog with the site. Aggregators assist both the blog networks and individual bloggers by helping readers find them, often creating aggregated RSS feeds of many blogs within a particular scientific discipline or blogging on a particular topic. For the individual bloggers, being aggregated may create a sense of belonging to the science blogosphere. There is evidence that journal clubs (regular meetings held by laboratory groups in which a few journal articles are presented and discussed) help researchers read and retain more knowledge (Linzer, Brown, Frazier, DeLong, & Siegel, 1988). It is possible that aggregations like

Researchblogging and Scienceseeker may contribute to knowledge retention in a similar way, by allowing engagement with the scientific arguments within a research paper. The table below shows examples of some current and defunct scientific blog aggregations.

Table 12

Examples of Scientific Blog Aggregations (August 2011)

Blog aggregation	Number of blogs	Rules and Regulations	Founded in
Researchblogging.org Created by Dave Munger, now owned by Seed Publishing	1200+	Must be blogging about a peer-reviewed scientific paper.	2007
ScienceSeeker.org (volunteers)	850+	Anyone who considers their blog a science blog may register	2010
Postgenomic (Nature Publishing Group)	810	Postgenomic succeeded by Nature.com blogs	2006 - 2009
Nature.com blogs	991	Highlights stories from Nature staff and Nature Networks, as well as science blogs from outside the Nature group. Blogs can ask to join the aggregation - it is simply a list.	2009

Researchblogging is an "opt-in" science blog aggregator. However, it only seeks to aggregate blog posts that discuss peer-reviewed research. Blog authors who join Researchblogging add particular code to posts discussing peer-reviewed research.



Figure 39. Example of Researchblogging.org icon in use on a blog post

The site's members (the blog authors) have authority in deciding what content fits the site guidelines and should be included. The fate of "borderline" content used to be discussed in the forum, however that discussion has subsequently moved to Twitter and comments on the news blog, and the procedure is that "registered users can flag individual posts that do not meet our guidelines and an editor will check them out and decide on the appropriate action" (Munger, 2011, personal communication).[85] Researchblogging.org's objective is that science bloggers use the icon when they write a "thoughtful blog post about peer-reviewed research, not just a link to a news article or press release" (Munger/Seed Publishing, 2008).

The full procedure for participation is:

1. Bloggers register with the site

[85] http://www.researchblogging.org/about.php

2. They continue blogging as normal, but when they write a post that meets our guidelines, they visit our site
3. They use our automated citation generator to create a properly-formatted research citation for the journal article they are citing
4. They paste the code we provide into their post
5. Our automated aggregator finds their post and publishes it on our front page, where thousands of readers can easily find it
6. A reference to their post remains permanently in our database, so users searching for posts on a specified topic can easily find it later (Munger/Seed Publishing, 2008)

Their guidelines for what constitute a research publication include that the publication should be: reviewed by experts in the field, edited, archived, published with clearly stated publication standards, and viewed as trustworthy by experts in the field.[86] This type of blogging about peer-reviewed research serves as another form of information filtering for those who pay attention to the scientific blogosphere, and has helped draw attention to the burgeoning informal interface between peer-reviewed research and post-publication peer review on blogs, which is discussed further below in the section on the *Arsenic Life* controversy.

Blog aggregations serve several purposes for the community, as they enable community members to find other like-minded bloggers, both within their disciplines and those who share common ideas across

[86] http://researchblogging.org/news/?p=53. Accessed 12 July 2011.

disciplines, such as an interest in open science and open data. As they are self-selecting—anyone considering themselves a science blogger can register for Scienceseeker, and anyone who blogs on peer-reviewed research may use Researchblogging icons (with the above caveats), they could be considered a neutral home for blogs outside networks, and another way of attracting blog traffic outside search engines. In the short to medium-term, they may also serve as quasi-archives of the scientific blog discussion space, providing links to content although not archiving it themselves.

Table 13

Scientific Blog Networks

(via Scienceblogging.com aggregator; list not exhaustive)

Blog network	Number of bloggers	Owner	Founded in
Scienceblogs.com	around 60	Seed Publishing	2006
Guardian science blogs	5	Guardian newspaper	2011
Nature networks	220	Nature publishing network	2007
Occam's Typewriter	14	Independent	Dec. 2010
Discover Magazine blogs	10	Kalmbach Publishing	2008
Scientopia	36	Independent	Aug. 2010

Blog network	Number of bloggers	Owner	Founded in
Field of Science	26	Independent	2008
LabSpaces	14	Independent	2005
CENtral Science (CEN = Chemical & Engineering news)	12	The American Chemical Society	2007
PLoS Blogs	14	Public Library of Science	Oct. 2010
Wired Science	7	Condé Nast	Sep. 2010

Becoming a member of a blog network is a goal for some blog authors, for several reasons: they do not have to deal with maintenance and upkeep of their blog; being a member of a network can increase traffic to their own blog, and in some instances, as discussed below, networks can become reference points for mainstream media, thus giving the bloggers involved some measure of media influence. Science blogging networks have proliferated since 2007, as shown in Table 13, above.

5.6 PepsiGate - Scienceblogs network controversy

Scienceblogs.com, established in early 2006, is an initiative of Seed Publishing, and has the largest audience of the science blogging networks at time of writing.[87] Invited scientific blog authors are supplied with a blogging platform and technical support if required. The Scienceblogs.com authors (colloquially named SciBlings internally) are not limited in what they may post, and are paid "a small amount" for their blogging based on the traffic their blog receives. Information on payment rates is not made openly available online.[88] Initially there were requirements for a certain frequency of blogging (at least once a week), but this rule was never enforced, as Zivkovic explains: "we took vacations, sick leaves, some bloggers went through traumatic personal life episodes, and community managers were all very sympathetic—community was more important than any individual blogger's monthly output, and network as a whole definitely produced a lot, with so many bloggers posting more than once a day" (Zivkovic, personal communication, 25 July 2011). The blog authors on Scienceblogs.com have been maintaining their own blogs for some time prior to receiving an invitation to join.

[87] http://seedmediagroup.com/

[88] I have not been able to find payment rates online, although there are a couple of references to this form of payment on Scienceblogs.com, notably in the PZ Myers comment reproduced in Footnote 90, below.

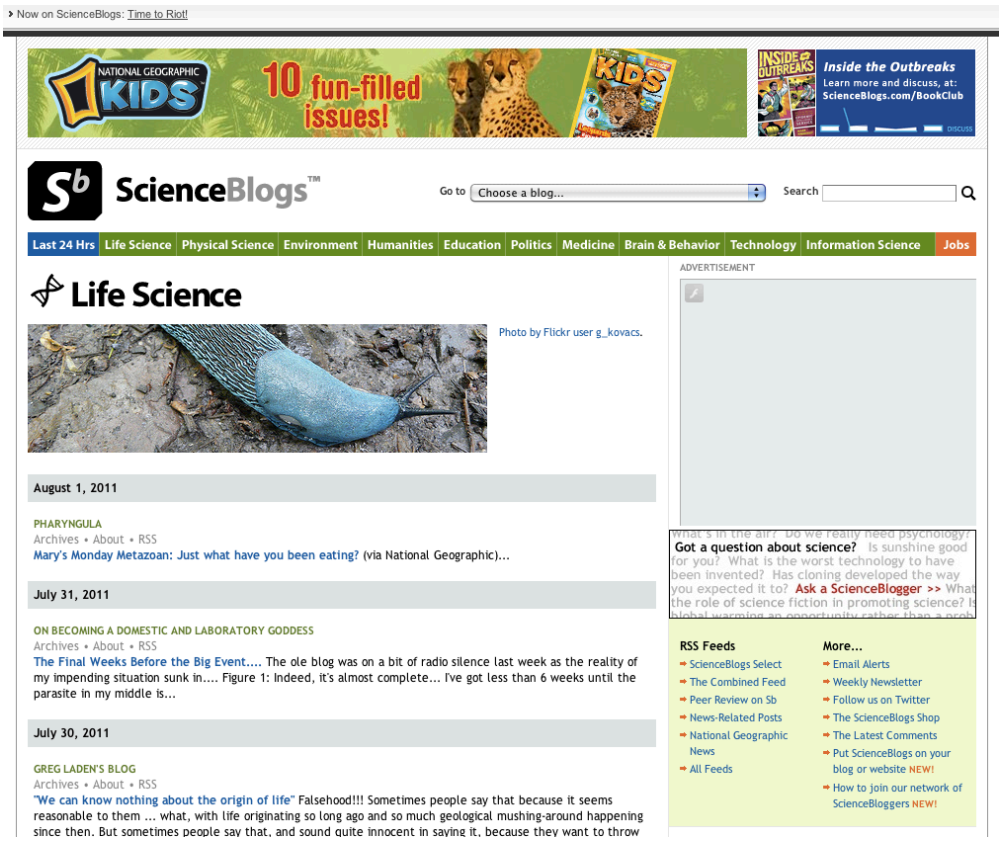


Figure 40. Screenshot of Scienceblogs.com Life Sciences home page, July 2011.

Zivkovic, a participant in this chapter's research, was invited to join Scienceblogs.com in 2006, and while on Scienceblogs used his blog (named *Coturnix*) for both his own writing and to promote PLoS publications, with Seed media's consent (he was PLoS community manager at the time). Zivkovic explained the attraction of blogging on a popular platform like Scienceblogs was partially due to the audience, who often came to Scienceblogs to read more popular blogs like Pharyngula[89], and stayed to

[89] Pharyngula http://scienceblogs.com/pharyngula (24 Aug 2011) is one of the most popular science blogs in the world, and rates highly among all blogs for traffic, not just science blogs. Wilkins (2008) characterises Paul Z. Myers Pharyngula blog as a blog about religion and politics, "and these topics attract the bulk of his massive readership of over 300 000 visitors per week. In between these populist posts, his science posts, usually on developmental biology, are works of art" (Wilkins, 2008).

read other blogs in the network. The effect of this increased traffic—traffic encourages more traffic due to more comments, links and trackbacks to the blogs involved—brought more attention to all of the bloggers on Scienceblogs.com:

> A significant rise in traffic and rank of scienceblogs.com resulted in all sorts of other deals. Choice posts of ours are linked from the New York Times science page. Likewise with the National Geographic site. Our blogs are sold on Amazon.com for Kindle. And the site is indexed not just in Google but also on Google News. (Zivkovic, 2010)

Zivkovic states that Scienceblogs.com became a reference point for mainstream media—when a big science story broke, journalists would look for responses on Scienceblogs.com, and their blogs would be quoted in national and international mainstream media. Zivkovic's explanation of the media power angle of Scienceblogs.com:

> This [high traffic and high Google search rankings] means not only that each one of us gets more direct traffic, and more potential indirect traffic from our SciBlings due to being on the network, but also an even larger and more powerful indirect traffic and visibility outside of the network. We are being closely watched, both by thousands of other bloggers and by the mainstream media. Whenever Scienceblogs.com explodes with a story, MSM [mainstream media] takes note. It is not by chance that some of the first reactions to the Pepsi scandal, even faster than on individual's blogs, appeared in places like The Guardian and the Columbia Journalism Review. As Jay Rosen and Dave Winer noted in their weekly podcast, the distance between us at Sb and the global media is very small. We are not just a loose collection of individuals blogging just for fun any more. (Zivkovic, 2010)

Into this environment Seed Publishing introduced a "nutrition blog" paid for and written by *Pepsico* staff (the soft drink and snack food conglomerate), without consulting any of the bloggers in the network.

Reaction by 30% of their existing bloggers was immediate—many leaving within 24 hours, almost all posting negative comments on their blogs. Their largest-traffic blogger (Pharyngula) has not left, although he posted negatively about the incident on his blog.[90] Seed Publishing responded quickly, removing the Pepsico blog (which had not yet posted any content) and apologising to the Scienceblogs.com bloggers within twenty-four hours.

This episode on Scienceblogs.com demonstrates the current volatility of science blogging networks, and the tensions between commercial ownership and science bloggers who value editorial independence and decry marketing posing as science, as Zivkovic put it: "a corporation paid to have a seat at the table with us. And that Seed made it happen ... this event severely undermined the reputation of all of us. Who can trust anything we say in the future? ... Once they [readers] look around the site and see that Pepsi has a blog here, why would they believe I am not exactly the same, some kind of shill for some kind of industry?" (Zivkovic, 2010). The Pepsi-gate controversy spawned several new blog networks and bolstered some already-existing networks as more than twenty high-traffic blogs left

---

[90] In "Say hello to…PepsiCo??!? WTF?" PZ Myers (Pharyngula's author) commented: "I get paid—on a fixed scale determined by number of page views. And I get paid reasonably well, not enough to convince me to quit my day job, but on a professor's salary, still a darned good bonus. But what makes it very difficult to pull up roots is 10,000+ posts and 1 million+ comments. That's a big investment in the scienceblogs.com/pharyngula brand. Seed has also been very, very good about avoiding any kind of meddling with blog content. There is zero pressure to conform, and no incentive to promote advertisers. We can trash PepsiCo all we want. It would be very hard to convince me to pull up stakes here. But one thing that could do it is if it turned into a corporate slum, where I'm walled in with fake blogs peddling high fructose corn syrup" (Myers, 2010).

Scienceblogs.com and sought new homes in the scientific blogosphere (Zimmer, 2010). Other media groups were beneficiaries of this exodus, as Discovery Magazine and Wired Science added several defectors from Scienceblogs.com to their already-existing blogs, and other media groups began their own networks, seeded with a few Sciencebloggers, for example, Scientific American Science Blogs (where Zivkovic is the new Blog Editor) and Guardian Science Blogs.

Despite the traffic and audience gains garnered from membership of Scienceblogs.com, almost 30% of their bloggers left after the Pepsico incident, and more are leaving after National Geographic's takeover of advertising and management in April 2011. Pharyngula, the highest-traffic blog on this network, has subsequently split his blog into two—one is at freethoughtblogs.com, and contains all of his content, and the same content resides at Scienceblogs.com, minus his anti-religious views.[91] Most recently, in August 2011 National Geographic are disallowing pseudonymous blogging at Scienceblogs.com, which is prompting a few Sciencebloggers who wish to keep their anonymity intact to leave the network (Drugmonkey, 2011).

We turn now to an examination of another controversial aspect of science blogging, post-publication peer review.

---

[91] Myers says: "I will still be maintaining my relationship with Scienceblogs and National Geographic, but only select content will appear there: that is, science, anti-creationism, that sort of thing...the openly anti-religious material will be on FtB only" Retrieved from https://plus.google.com/102600088268197126924/posts/WMHhpTAkr41?hl=en, 31 August 2011.

5.7 Arsenic Life - a blogging controversy

Another example of the tensions being played out in the scientific blogosphere is the discussion of research papers, known as post-publication peer review or "web scale" peer review (Neylon, 2011), discussed above in Section 5.4. The *Arsenic Life* story told below encapsulates the variety of reactions scientists have to this blogging phenomenon.

Rosie Redfield is Professor of Cell and Developmental Biology in the Zoology department at the University of British Columbia, based in Vancouver, Canada. Redfield has been blogging using Blogspot since August 2006. Redfield's blog was not part of a network but was known to many of the OLN users as Open Notebook Science (most of her research is discussed in her blog, but results are not live-blogged or fully open)—it is listed as an example on the ONS Wikipedia page[92], and the ONS aspects of her blog have been discussed in the previous chapter. However, Redfield states that she did not feel like she was part of an online community, as little traffic came to her blog, and her posts were rarely commented on.

> I just had a general sense of not being included in lists of people who were doing open science. Not being included in lists of women science bloggers ... the people of my lab read my blog, but even then, I'm not all that sure that they want to ... clearly some people who were reading my blog, because there were people who give me advice when I was trying to do bioinformatics. (Redfield interview)

On 2 December 2010, NASA held a press conference to herald research, published online the same day in *Science,* regarding the isolation of a

[92] http://en.wikipedia.org/wiki/Open_Notebook_Science

bacteria, named GFAJ-1[93] by NASA-sponsored researchers, which could allegedly assemble its DNA from arsenic instead of phosphorus, hitherto unknown/unfound on Earth. Media outlets responded with attention-grabbing headlines such as "*NASA Unveils Arsenic Life Form*" (Wired, 2 December 2010) and "*Arsenic-loving bacteria may help in hunt for alien life*" (BBC News Online, 2 December 2010). The response from the scientific blogging community was sceptical and fast.

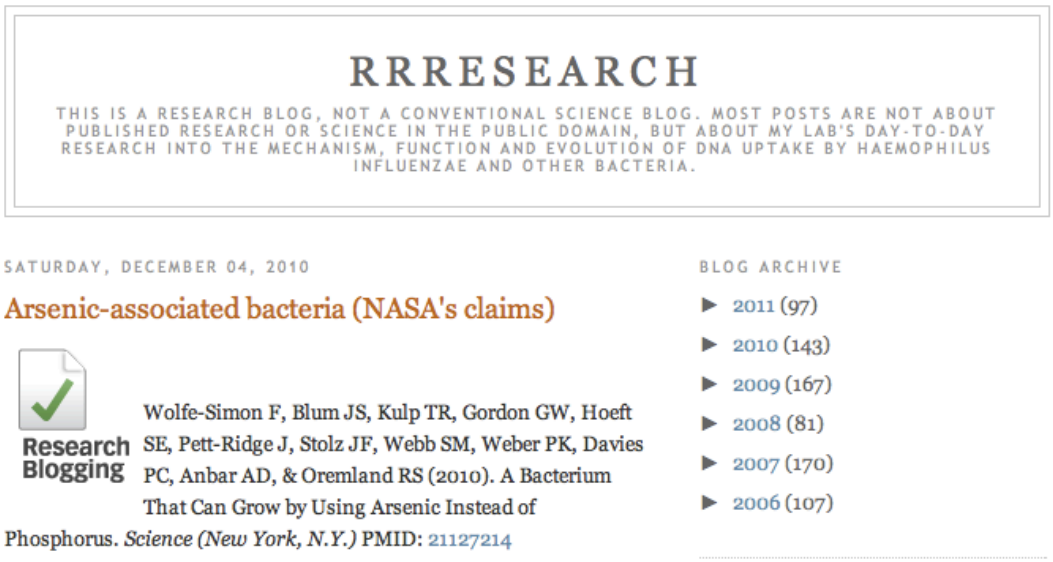


Figure 41. Rosie Redfield's blog, RRResearch, showing Dec 4 2010 post on *Arsenic-associated bacteria (NASA's claims)*. Note Research Blogging icon.

Redfield wrote a detailed scientific review and critique of the Wolfe-Simon et. al (2010) journal article on her blog on 4 December 2010. Redfield's review of the paper is no different in tone or approach to other reviews she has written on her blog: "when I blog about papers, it's partly

[93] "GFAJ" stands for "Give Felisa A Job" - Dr Felisa Wolfe-Simon is the lead author of the article, not employed by NASA but sponsored by it. The explanation of the acronym is from an article published in the Wall Street Journal by one of her co-authors (http://online.wsj.com/article/SB10001424052748703989004575652940497021092.html).

just ... to clear my own thinking about what it is and partly that, you know, if I get into a discussion with somebody about it, I can say well, I wrote all this up on my blog, just read my blog. And so that was basically what I was expecting to come of this" (Redfield interview). What distinguishes this critique from Redfield's other reviews are the responses of many scientists in the comments, drawn to the blog post by several online postings, including a comment by Redfield on the Pharyngula blog, several Twitter postings and re-tweets by others, and after a couple of days, links from many other science blogs.

As Redfield comments in an update to the 4 December post:

> Note to new readers: I wrote this post on Saturday Dec. 4, mainly to clarify my own thinking. I didn't expect anyone other than a few researchers to ever read it. Since then I've made a few minor corrections and clarifications (typos, decimal places, cells not cfu), but I haven't changed anything significant. Please read the comments - they contain a lot of good scientific thinking by other researchers. (Redfield, 2010)

There were many other posts and discussions about the "Arsenic Life" article in the scientific blogosphere. Redfield's blog post and the ensuing commentary and discussion are illustrative of the clash between blogging as an informal form of post-publication peer review and the "accepted" methods of peer criticism. The range of attitudes within comments on Redfield's post are summarised below.[94] Redfield speculates that her post, written on a Saturday afternoon after she read the paper, was mainly circulated via Twitter and a few links on other science blogs, resulting in

[94] I devised the coding categories and coded the comments.

over 100,000 hits on her blog, and 255 valid comments (normally Redfield receives fewer comments: 0-2 per post), of which 75 (28%) were anonymous.

Table 14

Summary of comments on Redfield arsenic blog post

Type of comment	Number	Percentage of total
Redfield response to comment	18	7.5%
Negative - expressing dismay about the way the blog post talks about the paper's authors, or criticising the concept of peer review on a blog	15	5.8%
Negative - saying that Redfield's science is wrong	4	1.5%
Neutral - scientific clarification or question, no opinion about format expressed	49	19.2%
Positive - the science in the blog post is correct, and/or this is a valid way to "do science"	64	25.0%
Positive - expressing a positive opinion and included scientific clarification or question	49	19.2%
Mixed - agreed with Redfield's scientific assessment but expressed dismay at the language used towards the authors of the paper	15	5.8%

Type of comment	Number	Percentage of total
Other - unrelated comments, grammar arguments	41	16.0%
Total	255	100%

Of interest are the breakdown of the comments—63.4% of comments contributed to the scientific discussion, or 81% if Redfield's responses are included—therefore the majority of scientists who responded regard the blog post as an advance of scientific discussion:

> If they decided to publish this paper too soon, then they must face the consequences of their decision and questionable science. The paper is out there for anybody to comment. I am willing to think that we are in a free world where I can comment a paper freely. Moreover, I think Rosie's post got actually refined by comments of readers here, and improved at the end. Indeed, this is a blog, and if someone wants to say the paper is a clear example of poor science, [they are] free to do it. (PetHW, 2010)

Similarly, comparing the post and discussion to peer review:

> I am a scientist in a completely different field, and have nothing to add to Dr. Redfield's critique of the original Science article. I do have to say, however, that I wish my field had a forum such as this for vetting and responding to published research. If we are looking for a viable alternative to the current system of peer reviewed publications, which often screens IN bad science and screens OUT good science with null findings, I think we've found it. In just a few days, we have achieved a thorough critique by an expert in the field, which has in turn been reviewed and further critiqued by other experts in the same and related fields. The result is a thorough and accurate assessment of the original work. Dr. Wolfe-Simon encouraged further debate under the auspices of peer review - namely, in

> Science. Well, I think she got it. This blog is the essence of peer review. Plus, free spelling and grammar checking. (Kennedy, 2010)

Some regarded the post as a normal scientific discussion to which they can contribute their scientific opinion:

> The discussion is excellent, even if, in my opinion, there are too many anonymous posters. Several people above have acknowledged their mistakes and have moved on. This is the only way that science can proceed. (Fornaro, 2010)

> I especially like your answer to the Anonymous who pointed out the error in your estimation of exponential growth: Check all my calculations! After all, that is the essence of science ... And any thinking scientist, when invited into such a discussion, takes nothing at face value—they repeat the calculation themselves, even if only in their head. They think through the same scenarios as you explain them and then see if they come to the same conclusions you come to. And then they join the discussion accordingly, either affirming your ideas or presenting their own. (3.14159rate, 2010)

Of the negative comments, which make up 7.3% of the total, few posters were able to find fault with Redfield's scientific assessment (4, or 1.5%), and the other 15 (5.8%) express similar viewpoints to the Arsenic paper's authors, regarding the impropriety of peer review taking place on a blog:

> I don't have a problem with you having a contrarian opinion on the recent findings. However, I think you are equally guilty of premature conclusions and using the media to create a circus. It appears that you know what you are talking about. Since you are apparently claiming a moral high ground why did you not submit your objections to the publishing journal. (Anonymous, 2010)

> The original post was extremely unethical and irresponsible. If you can refute the work, do so. And you refute work with your own work, or your published criticism, which gets reviewed, not with a blog. (Anonymous, 2010)

> Interesting commentary. I should note, right off the bat, that I used to work with one of the researchers associated with this paper. ... Your commentary crosses over from being a criticism of the science into calling into question the integrity of the researchers involved, which is rather poor form, and extremely unprofessional. It's one thing to be a skeptic, and another thing entirely to accuse people of lying (which also suggests that you are more clever than Science's reviewers, having seen through all this deception! Which is of course possible, but we have no way of knowing how your qualifications and experience stack up to theirs). I'm also a little skeptical of some of your critique, based on the number of mistakes that have emerged, and lack of clarity hear [*sic*] and there, but I recognize that this is just a blog post, and not, in fact, a scholarly article, and the standards are not the same. (Distant One, 2010)

> While I whole-heartedly agree with your critique of the primary findings of this paper, I find your editorial comments on their motivations rather pessimistic and premature. (inducedPluripotency, 2010)

A few commenters also expressed dismay that creationist websites were linking to this post as an example of the imperfections of science, summarised as:

> "Science by blog" is simply not a good idea, especially if scientists wish to casually question the character and motives of their peers ... Somehow, out here, truth is viewed as a majority-rule endeavor, and consequently, creationism, climate-denial, and 9-11 conspiracies thrive. ... And they're looking for any excuse whatsoever to perpetuate their belief that knowledge-via-evidence is just one more alternative worldview, one more test of their faith or politics or things they "just know".... Given that reality, expert public discourse of the type seen on this blog (with reckless speculations on scientific agendas and suppression of data) is not merely unhelpful; it may actually be dangerous and irresponsible". (Printz, 2010)

and:

> I followed links here from creationist's websites. They are cheering that Prof. Redfield has exposed science as the dishonest sham that all good Christians always knew it was. They don't understand the science, pro, or con. But they do

> love "flim flam," and "unscrupulously pushing NASA's ... agenda." Nor did they miss the accusations of data suppression and fraud in Redfield's remarks; "(or maybe they did but they didn't like the result)," and "(Might they have not presented assays using properly purified (washed) DNA because these turned out to not have any arsenic? Am I just paranoid?)". (Hurd, 2010)

Despite the polarised viewpoints of the commenters, many of both the positive and negative commenters urged Redfield to submit a technical comment to Science, which she began to shape the same day, and submitted a few days later (Redfield, 2011). It was one of the eight Technical Comments published by Science in May 2011 (of which further details below).

NASA had called the original press conference to announce the Arsenic Life paper, and then chose not to respond to the comments and questions in the scientific blogosphere, stating that these comments were not part of scientific due process: "Wolfe-Simon will not be responding to individual criticisms, as the agency doesn't feel it is appropriate to debate the science using the media and bloggers. Instead, it believes that should be done in scientific publications"(CBC News, 2010). Wolfe-Simon responded briefly on Twitter, echoing the same message, as shown in Figure 42. The authors did respond more fully in their Technical Comment published in *Science* (see below), but were not allowed to introduce new material to support their argument, as per long-standing editorial policy at *Science.*

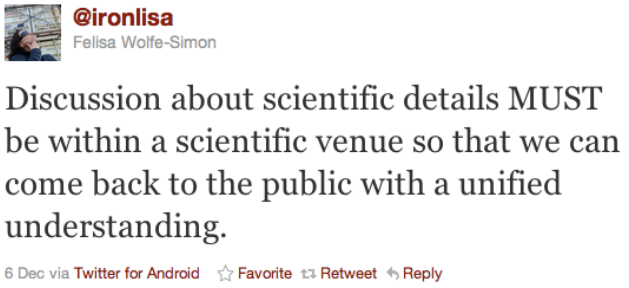


Figure 42. Felisa Wolfe-Simon's Tweet on the blog responses to the publication of the Arsenic paper.

On 27 May 2011, in *Science Express* (the in advance of print online edition of *Science)* and in their June 3 2011 print edition, *Science* published eight Technical Comments on the "A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus" paper, as well as the authors' response. An article in the print edition stated that this was the highest number of comments *Science* had ever published on a single paper, also noting that the debate was "finally being aired in the scientific literature rather than on blogs" although the substantive criticisms raised in the Technical Comments on the paper were identical to those raised on blogs within hours of its online publication (Pennisi, 2011). Pennisi also noted that twenty-five Technical Comments and Letters had been submitted, but there was much overlap between the concerns, so eight submissions which covered the majority of the comments on the article were finally selected by *Science's* editors (Pennisi, 2011).

This story is still in progress. Redfield has recently obtained samples of GFAJ-1 provided by the *Science* paper authors and is collaborating with several other scientists (in the Kruglyac laboratory at Princeton) to test the "Arsenic Life" hypothesis. Redfield is detailing all of her experiments in her

OLN blog in close to real-time. Redfield and her collaborators will also seek to publish their experiments with GFAJ-1.

This episode demonstrates two things—firstly, that the outcome of this particular post-publication peer review is "normal science" (except for the almost peripheral role of the publishing gatekeepers), speeded up many times, and secondly, that scientists who comment on blogs have a largely positive attitude to informal public discussion, and reasonably sophisticated arguments about why this discussion is a good thing for science.

As noted by Redfield and others, scientists routinely dissect and dismiss new papers in the tea-rooms and Journal Clubs of laboratory groups and departments or at field-specific conferences, but the dissemination of this discussion globally, within a day or two of publication, and having this scientific dissent and disagreement (which is part of normal science) picked up by mainstream media, is new (Redfield, 2011; Batts et al., 2008, Nerd Alert Podcast, 2011).

The controversies discussed above highlight that blogging currently occupies a particularly unstable niche in the online communication of science at the macro level, as the blog networks struggle for audience and stability—but at the individual level, blogging is continuing unabated, as scientists feel free to jump between blogging platforms and networks as they deem necessary. Despite this movement, science is happening on blogs in the form of knowledge-sharing and related discussions, as well

information-filtering, research-notebooking, and collaboration. The discussion below shows what these changes mean across the seven research aspects within the blogging ecosystem.

5.8 Results and Discussion

5.8.1 Scientific Roles

RQ: Are academic life scientists who blog extending their professional roles beyond their traditional ones?

Scientists who blog are willing to engage with each other in public, displaying the kinds of water-cooler discussion—analysis and deconstruction of other scientists' work that is historically enacted at conferences or in the laboratory hallways. Some fear that this publicity may devalue science in the public eye (for example the commenters on Redfield's "Arsenic Life" blog post who were concerned about creationist blogs linking to the post), although as shown from the comment analysis of that post, they are very much in the minority. Redfield commented that she may also not have bothered to write a Technical Comment to *Science* if her blog post had not received the attention it did, and she had not received the encouragement she did.

Scientists are using blogs to perform open or "web-scale" post-publication peer review and critiquing current scientific systems while being fully engaged with them—publishing their own research through both formal and informal channels. They also use blogs to share research thoughts and data. It does not appear that science blogging is incompatible

with scientific career development. Early-career bloggers see blogging as a way to connect with their peers, many more senior than themselves, as demonstrated by Salvagno, Lindenbaum and Beltrao's positive experiences.

Scientists who are members of blog networks also seem aware of their audiences, both their peers and the public, and will address both while writing blog posts, which for most is an extension of their roles, as previously their publication role was restricted to academic papers.

Scientists who blog are also having to find time to blog in their lives, and most seem to carefully make time for blogging outside their working hours, even though they are blogging about their work. So it could be that their profession is impinging on their private lives, although this infringement may not be sustained if blogging becomes an acceptable (or even mandated—many academic institutions have or are developing institutional blog platforms) part of professional academic life.

5.8.2 Online Collaboration

RQ: Does blogging facilitate inter-institutional and interdisciplinary collaboration for academic life scientists, and if so, how?

Scientists who blog are reconfiguring their disciplines (Dutton, 1999, 2005) and disciplinary culture as they discuss work outside their specialism, or comment on and discuss issues that it is not possible to discuss in journals, because lengthy treatment of the issue in journal paper format is not appropriate or required, as shown by Juhl Jensen, Saunders and Lindenbaum sharing code snippets and short pieces of analysis on small

datasets. As Singh remarks: "if you want to be a good blogger you have to pay attention to what other people are saying. You have to spread your mind and absorb a lot of information; and you learn a lot in the process" (Singh interview).

The present research indicates that blogging practices are part of an interconnected group of activities around doing science in public, or publicising open ways of working. Scientists who blog and engage in scientific social networking in general have reported new connections through these which result in new collaborations. An example within this chapter include Redfield's collaboration with a Princeton group to analyse the GFAJ-1 bacteria; other examples are heavily intertwined with either or both of the OLN and FriendFeed cases, and are described within those cases, such as Lindenbaum and Su's collaboration over FriendFeed and blogs, and collaborations that have arisen through OLNs and blogs, such as Bradley and Lang. Perhaps more important in an interdisciplinary sense are the cross-disciplinary connections they are making via their discussions of open science—the community they all report being part of, which connects them across a range of scientific disciplines, including chemistry, biophysics, mathematics and bioinformatics.

5.8.3 Science Communication

RQ: How do academic life scientists use blogs to publish different types of scientific data?

As shown above, many of the science bloggers featured in this chapter reserve their blogs for keeping ideas and notes about their research, or for discussing broader academic topics such as open science or peer review. However, the research ideas and notes they publish are a form of data, which prior to blogs were not kept publicly, but scribbled in the margins of laboratory notebooks or periodically typed into haphazardly-named text files on their personal computer. The blog allows scientists to enunciate and clarify their ideas, and whether or not they are conscious of an audience, the "publish" button provides an extra impetus to make their writing understandable to their peers and sometimes the public. In addition "you can actually go back and remember what you were thinking about, and all my academic progress so far [8 years] is actually online ... for that at least I'm convinced" (Beltrao interview). In this manner, blogs also provide a valuable chronology of the development of research thinking, which provides scaffolding for future research and a record of past achievements.

As with Herring et al.'s (2006) longitudinal study, blog writing matures, as authors tend to write more the longer they have been blogging, and filter less, or provide longer more thoughtful posts about things they are filtering, that is, links become catalysts for discussion rather than content. In addition, newer technologies like FriendFeed, Twitter and most recently Google+ (June 2011)[95] are becoming homes for filtered content—

[95] http://www.google.com/about/company/history.html#2011

demonstrating again the fluidity of Internet mediums and their users' flexible approaches to scientific communication.

Only a few of the scientists involved in this research tend to publish raw data on their blogs, or make raw data available for download. Juhl Jensen publishes ideas and some data, the snippets or afternoons of research that cannot be developed into a full paper but may be of use to others. Saunders has also done some analysis of scientists' FriendFeed activity, and will show code snippets and results, or provide code downloads on his blog. There is some evidence that scientists involved in this research are publishing a wider range of data: the code snippets mentioned above, Redfield's publication of grant applications on her group's website as soon as they are submitted, a practice also recently followed by Koch, who posted a failed grant application on his blog.

Discussions around open science could also be considered a form of data, as they provide data for social scientists, as they are a continuing record of the changes in scientific communication using the Internet. As this thesis argues, this data shows the development of a SIM and CM which seeks to change the way science is done, and blogs provide ample evidence of this.

5.8.4 Scientists and public engagement

RQ: How do academic life scientists see that their engagement with the public has changed since they started blogging?

The level of engagement with the public depends on the focus of a scientists' blog. Many bloggers are talking to an audience of their peers and colleagues and ancillary professionals such as librarians. However, many are also seeking to explain peer-reviewed research in a way that avoids journalistic simplification and accurately reflects the outcomes while maintaining a scientific approach to the data and hypotheses. The public focus of most of these blogs though has allowed more engagement with the wider scientific community rather than the general public (and the reason for this focus in this case is the open science movement) – there are other science bloggers focusing on public engagement, such as Pharyngula, mentioned in this chapter.

However, as also shown in the collaboration discussion, the peer-focused scientists who blog are interacting with a wider range of colleagues, from early-career scientists to professors, across different disciplines.

5.8.5 Gatekeeping

RQ: Does the use of blogs by academic life scientists circumvent existing gatekeepers in science and/or create new ones?

As the "Arsenic Life" and "PepsiGate" studies have shown, there is significant controversy and discussion around gatekeepers, old ones in the form of academic publishing's reaction to post-publication peer review, and new ones in the form of blog networks, their brand identities and the voices of the scientists who blog within them.

Most of the science bloggers interviewed for this chapter are able to cite positive outcomes of their blogging, ranging from new contacts to new collaborations. Two participants have found new jobs via their blogs: Zivkovic, previously PLoS community manager, now Scientific American Blog Editor, and Grant, now Information Architect at Faculty of 1000: "my [recent blog post] tagline was "I started a blog and all I got was this job" (Grant interview).

5.8.6 Status and Rewards

RQ: Do the various forms of online dissemination being used by academic life scientists yield rewards and status comparable to traditional outputs, or are there new mechanisms developing for reward and recognition of these activities?

There are as yet no official mechanisms for recognition of online discussion and dissemination of scientific work via blogging. Many of the existing research articles assert that there are risks involved in being public and non-anonymous, not least the threat that blogging may skew the opinions of hiring committees away from the candidate, or even that blogging could destroy a burgeoning academic career (Ewins, 2005; Gregg, 2009a). As NASA and *Science's* dismissive responses to the "Arsenic Life" blog comments on their paper demonstrate, formal scientific publishing is still unwilling to directly engage with the blogosphere, although indirectly it may have, as some scientists assert that *Science* would probably not have published as many as eight Technical Comments on the Arsenic paper had

it not attracted the levels of attention it did.[96] There are also, at time of writing, two groups testing the hypothesis with samples of GFAJ-1, and Redfield is doing so openly, documenting the process on her blog. Redfield also intends to formally publish the results of the experiment. Thus an argument can be made that blogging can be used to speed the process of normal science, as was also seen in the OLN chapter.

5.8.7 Openness of scientific data and discourse

RQ: Where in the chain of collecting and disseminating research materials are the incentives and barriers to open practices in the use of online tools by academic life scientists?

Several scientists participating in this study use their blogs to share ideas, methods, grant applications and code, even if they do not do Open Notebook Science (Juhl Jensen, Johnson, Saunders). Blogs are sometimes used for sharing other types of data. For example, Redfield recently started a blog to share papers that were not widely available but were being widely discussed—the Wolfe-Simon et al. *Arsenic* paper among them (*Science* later made this paper free to access). The comments about this action displayed attitudes towards this, as they ranged from pointing out that sharing this way was a potential violation of the publishing conditions of the paper, to pointing out already-established ways of sharing these kinds of papers, for example the FriendFeed *Papers Wanted* group. Incentives to continue

[96] Although this could of course be disputed, the average number of Technical Comments published in an issue of Science in the past six months is 0.3, discounting the Arsenic Life paper issue (June 8, 2011).

blogging are clearly enunciated by the scientists studied in this group. As previous researchers have found (Hsu & Lin, 2008; Karimi & Poo, 2009), blogging is an enjoyable activity, and they are motivated to continue and expand by the interactions with their audiences.

Barriers identified include the time costs of blogging—time must be set aside to think about posts—although this has also been cited as a positive by some, as thinking-by-writing helps them clarify their ideas, as has been noted about Redfield and Neylon. Several scientists talked about feeling a self-imposed pressure to blog once they started, particularly if they were conscious of having an audience—some research participants, such as Mesko, Singh and Koch feel guilt about not blogging. Lastly, although these concerns are dismissed by the bloggers themselves, evidence from outside the blogosphere in research and discussion in the *Chronicle of Higher Education* suggests that keeping a blog may be regarded negatively by others in their field (Kirkup, 2010; Tribble, 2005). Participants in this research are more likely to report peer indifference than opposition: "people know me online, I got those invitations from *Nature*, Japan etc, ... and then I meet the (French) community of bioinformaticians ... nobody knows about RSS feeds, about blogging ... that's depressing" (Lindenbaum interview). However, recent attention given to academic blogging by episodes such as Arsenic Life, and the subsequent reaction of the journal *Science*, which published an unprecedented number of Technical Comments

on the paper, indicate that science blogs are beginning to receive more attention from mainstream science.

5.9 Conclusion

Blogs are the third category of online tools described in this research. For some of the scientists participating in this study, blog use is interconnected with their activities in FriendFeed (and increasingly Twitter), which promotes and facilitates online discussion of their blog posts and/or their use of OLNs. For others, blogging is not consciously integrated into their online scientific networking, but it is a tool for disseminating their research ideas and assessments of other work in their own and related fields. Peer reaction to public and global reflections about science is largely positive—many scientists see it as a natural extension of the peer review process, while acknowledging that at least for the moment, it is supplementary to the formal publishing processes of science.

Blogs are a low-barrier-to-entry tool for scientists interested in communicating with their peers and the public. Starter bloggers feel no pressure to blog to a particular schedule, and no requirements to interact with their nascent audiences, rather blogging tends to start as a writing exercise for the authors themselves. An audience with its attendant pressures only develops later, if at all. The Arsenic Life case points out that before reviewing the Arsenic paper Redfield was relatively unsung and unread.

So we can see that blogging means different things to different people, and audiences vary, but the main findings are that it serves as a publication channel, and a way of informally discussing peer reviewed research (which is, as shown above, somewhat controversial). Mechanisms such as Researchblogging.org have arisen to link the blogs with the peer reviewed research they discuss, highlighting the connection between science blogging and the peer-reviewed research—those who discuss research on blogs are most often also scientific peers.

In terms of the trilogy of online tools described in this research, blogs can fit at the beginning and the end of the research process. FriendFeed (and lately, Twitter) are the information seeking and searching filter processes, and places to promote and sometimes discuss scientific content. OLNs document the raw data emerging from the laboratory, and blogs fit within the ecology of online tools as the longer-form discussion option—in the realms of informal communication they are more like peer review, the kind of peer review that happens at conferences, in laboratory corridors and lunch rooms—rumination, discussion, and argument. They can also serve, as we saw in this chapter, as places for idea generation and development.

Chapter 6
Synthesis

6.1 Introduction - Science and Technology advancing together

We have seen evidence of change in many aspects of science across the three cases of scientific networking using FriendFeed, laboratory work using OLNs, and the dissemination, discussion and dissent of science blogging. Resistance to these new approaches to scientific work is also evident. This chapter will lay out the changes being seen, and the resistance to that change, providing examples of the core activities and personnel of the combined SIM/CM, how they encounter resistance to their ideas and activities, and how, despite this resistance, the members of the open science movement are driving forward their open science goals.

This synthesis will remind the reader what characterises a SIM and a CM, and will show how the open science movement described in this thesis combines elements of both. The seven research aspects of scientific roles, collaboration, open data, scientific communications, gatekeepers, effects on career status and rewards, and interaction with the media and public provide a framework for revealing the specifics of the open science movement.

Although the activities chronicled in this thesis fit with prior research about how scientists work, this research also finds strong evidence of changes in science due to the affordances of computerisation including telecommunications and computer networking, and therefore extends the

body of research in the sociology of science by combining the complementary theories of Computerisation Movements (Kling & Iacono, 1994) and Scientific/Intellectual Movements (Frickel & Gross, 2005). These theories complement each other because computer technology is the prime facilitator of new forms of science communication taking place online which are in turn aiding the formation of a SIM aiming to permanently alter the landscape of science communications. Although some could argue that these communications are the same communication strategies wrapped in new technological frames, this thesis argues that these technological frames have several core affordances, and the way scientists have chosen to use them offers transformations in the scale, speed and the disciplinary makeup of communities that scientists are now able to form.

These transformations in use are characterised by scientific working practices and approaches to communication which are extremely open (sharing as much data and knowledge as possible with the community); flexible (using whatever tools seem like a current best fit for this dissemination, and discarding, extending or switching tools to suit the needs of the task); and fluid (operating with little regard for traditional boundaries).

The current system of science is built on a framework of status and rewards supplied by academic publishing—scientists must publish as many articles as possible in peer-reviewed journals, preferably high-impact ones (as measured by a publisher-derived index), and these publications are

used to judge worthiness for funding and jobs. Other foundations of the current system which are rewarded are the ability to write fundable research proposals, which also benefit from prior publications. Into this competitive environment the open science movement has introduced change, showing that the current system can still be used while conducting science "in the open"—sharing data and knowledge and assisting others. However, to many scientists within the current system this seems like a huge risk. They fear that others will use their data, or that their disciplines are so embedded in this competitive system that to be open would mean others might use their ideas to publish ahead of them. These fears are characterised below in terms of the resistance to particular aspects of open science. It should be noted that some of these examples are being drawn from outside the open science community, as this research was unable to find any instances of being scooped happening to those practicing open science.

Resistance to data-sharing is exemplified by *Climategate*, referenced earlier in Section 4.7.3 of the OLN chapter. Despite receiving valid Freedom of Information requests for their data, scientists at the University of East Anglia's Climate Research Unit did not share it. Later, their email server was hacked and several thousand of their emails were stolen and released (Secretary of State for Energy and Climate Change, 2010). Scientists involved were reluctant to share for two reasons: firstly, they maintain (as

many scientists in their position do)[97] that their data is too complex for anyone without their expertise and training to understand; and secondly, due to this complexity, it may be erroneously misinterpreted to serve some other purpose (particularly germane in contested areas of science such as climate change). This reasoning may well be valid to some extent, as to the untrained observer, translation of the findings from this data into plain language will always be necessary. However, this serves to underscore that this raw data is not a threat—rather, sharing it with other scientists publicly may throw up new solutions to old problems, new theories, or, as has already been shown in this research, provide faster ways of doing science by refining methods. Maloney's open laboratory notebook shows this, as a procedure that took him a year to perfect took a graduate student on the other side of the country only two weeks to reproduce using the information from his notebook (Maloney, 2011b).

The Arsenic Life controversy discussed in the blogging chapter sheds light on two more points of resistance: post-publication peer review and public discussion of science. Several fears were raised in regard to this, as the analysis of comments on Redfield's post shows. Firstly, again the concept of misappropriation of data was raised, in the form of fears that anti-science groups would seize on this demonstration of dissent and argument in science and use it to show that scientific findings are *all* subject to refutation or being displaced by more powerful knowledge. Of course,

[97] I have heard a similar claim made by a particle physicist at a conference.

all scientists know this to be true—this is why scientific theories are theories. Secondly, the concept of post-publication peer review was roundly criticised by a minority (5.8% percent) of commenters, who used words like "dangerous, irresponsible, and reckless" to describe post-publication peer review on blogs. More commenters (44.2%) praised the activity of post-publication peer review. However, the idea that this praise constitutes wholesale acceptance of the notion should be tempered by the fact that blog commenters are a minority, within a minority of scientists who read blogs. Having said that, the publication of eight Technical Comments by *Science* was unprecedented, and indicates that the "storm in the blogosphere" did receive wider attention from publishers and the scientific media.

The primacy of the current reward system in science is such that even those practicing completely open science are conservative with regards to the involvement of their early-career researchers and graduate students, in deference to the fear that openness could potentially harm their ability to publish. Most of the OLN-using scientists in this study made the use of an OLN optional for their graduate students and postdoctoral fellows. However, reports from the students who have embraced the use of OLNs have been positive. Maloney, in the Koch laboratory, was awarded a PhD based on his open dissertation, Mirza in the Bradley laboratory has also been awarded a PhD for work containing significant amounts of open science, and a postdoctoral fellow in Todd's laboratory has reported benefits in terms of faster production of research results.

The last form of resistance to open science is shaped by the fact that engaging with open science also requires close engagement with information and communication technologies. Some scientists are unwilling to engage with these technologies although learning how to use them is less complex than the information tasks they perform most days in the laboratory, such as processing and managing machine data and downloading PDF files of papers from publisher websites. Use of the open science technologies necessarily requires openness, and many scientists are cautious about putting their names on data and theories that have not received pre-publication peer review and the imprimatur of a journal. In addition, the most common issue for the scientists involved in open science did not have to do with institutional resistance, or active resistance by their peers, but time. It takes time to habituate oneself to using these technologies, however simple they are. At the moment, for most, this time is taken outside of one's paid job, at least for many science bloggers and FriendFeed users, but less so for the OLN users, as notebook upkeep is an essential task of their working days. However, those who do these activities acknowledge their positive impact on their work practices, and can foresee a time when they are an integral part of their scientific workflow.

To summarise, the active resistance to open science is based on unproven fears of negative repercussions of data-sharing, public peer review, and potential restraints to early-career researchers, and the real issue of time management within a scientific career. However, the

predominant reaction to this movement is passive—indifference (if that can be called a reaction), as on the surface open science poses no threat to current ways of working, as Saunders has remarked about his blogging: "in my first job they didn't really care. I think they knew about it but they had no strong feelings one way or the other" (Saunders interview). This indifference is encountered because the payoffs for open science are not seen to be great enough to overcome the perceived threats to scientific careers—one thing, peer reviewed articles, are what count for most scientists, and any distraction from that goal is seen as worthless. Therefore the forces of indifference must be overcome for this movement to succeed, in two ways, firstly: dispelling the fears with examples of successful open science, which are growing but not yet plentiful, as Neylon remarked in 2009: "the point at which this is going to become more useful for other people is when we've got some more complete stories, we don't have much in the way of complete stories" (Neylon interview). Secondly, in order to develop these stories, open scientists need to continue honing their ways of working, trying new technologies and developing their existing communication and dissemination solutions.

6.2 The Open Science Scientific/Intellectual Movement

Frickel and Gross's (2005) General Theory of Scientific/Intellectual Movements proposes a model for intellectual change in science. "SIMs are collective efforts to pursue research programs or projects for thought in the face of resistance from others in the scientific or intellectual community" (p.

206). A movement's goals are to "gain adherents, win intellectual prestige and ultimately acquire some level of institutional stability" (p. 205). Frickel and Gross's theory can be examined in two parts: firstly, they identify the features of a SIM, which are: having a "more or less coherent program for scientific or intellectual change or advance"; enacting practices which are contentious "relative to normative expectations within a given scientific or intellectual domain", displaying an orientation towards a political stance—not necessarily around creating more powerful positions for its members, but showing an interest in altering the uneven distribution of scarce resources across intellectual fields; participating in organised collective action; varying in intellectual aim and scope, and secondly, when examined from a historical viewpoint, they are episodic, in that they can be identified as historical entities existing for finite periods (Frickel & Gross, 2005, pp. 206-208).

The coherent program, contentious practices and political stance features assume that those becoming members of a SIM (and as is noted within this chapter, that membership need not be formalised) are doing so in resistance to, or in opposition to, some aspect of science that they wish to change. In turn, those belonging to a SIM would expect to encounter opposition to their aims and actions. The characteristics of being episodic and varying in intellectual aim and scope are less important descriptive factors for the purposes of this research, but mean that SIMs can take many

shapes and forms, and be of varying durations, and will either fade or stabilise into "normal science".

Frickel and Gross cite many examples of SIMs in support of their theory, some of which, like the SIM/CM of open science examined in this thesis, are supported or launched by technical developments, such as the growth of the PCR method which marked the start of the biological data deluge (Rabinow, 1997) and the move by physicists from university-based experimentation to expensive shared sites such as particle accelerators (Traweek, 1992), or the rise of radio astronomy as a legitimate field within astronomy, eventually overtaking optical astronomy (Mulkay & Edge, 1976). However, other SIMs are sparked by societal changes that do not involve technology, such as increasing entry by women and black people into academia in the 1960s which helped shape fields such as women's studies and African American studies (Messer-Davidow 2002; Small 1999), or the development of queer theory in cultural studies (Seidman, 1996). Other factors, such as intellectual opposition to prevailing ideas, are illustrated with Garfinkel's involvement in the development of ethnomethodology in the 1960s and his difference of opinion to his mentor Talcott Parson's interpretation of the role of rationality as a method of understanding social behaviours (Mullins, 1973).

Several examples from this research attest that the size of the open science community is disproportionately small compared to its impact on the overall science community. For example, Chapter Three, discussing

FriendFeed, provided several examples of editorials and comment pieces authored by open science advocates published in the popular science media (also see Appendix G). Chapter Two, Research Methods, described the engaged network community as a source of research participants in relation to the size of the population in the broad research field of the life sciences. The community of academic scientists is large, by some estimates up to seven million people, and the community that is actively championing open science is by comparison small, comprising a few hundred bloggers, a few hundred active FriendFeed users, and around a dozen OLN users. Prior research indicates that scientific and intellectual change in science is usually spearheaded by small, tightly-networked coherent groups like these, who interact often and provide a visible platform for communicating and coordinating the movement (Parker & Hackett, 2009; Collins, 1994).

Frickel and Gross also propose four key theoretical conditions which allow a SIM to emerge and grow: dissatisfaction with the *status quo*, having access to a variety of resources to aid growth, the ability to recruit people and resources, and resonance, or how the ideas of the SIM are represented to both insiders and outsiders.

Dissatisfaction - A SIM is more likely to emerge when high-status intellectual actors harbour complaints against what they understand to be the central intellectual tendencies of the day (Frickel & Gross, 2005).

Frickel and Gross characterise dissatisfaction as doubts around the wisdom of prevailing approaches, in this case to research practices. This,

they say, may be because of generational shifts—new scholars versus old; differences in outlook or values; becoming aware of theoretical approaches in other fields which could be incorporated into their own; or, in some instances, "new metaphors or technologies may call into question old ways of thinking about things"—or in the present case, old ways of *doing* things (Frickel & Gross, 2005, p. 210).

Participants in this research talk often about their motivations for doing science openly and promoting it. Most often they mention that "radical transparency" feels to them like a natural extension of the already relatively-open culture of science in the sense that all researchers, particularly academic scientists, already publish details of their experiments, methods, and theories in journals. However, they feel strongly that this is not enough, and their central dissatisfaction with the current system is that it slows research, disables relevant reproduction of research, requires unnecessary reproduction of redundant research due to lack of transparency and findability of previous results, and promotes secrecy and its attendant problems of delayed publication and competition. In addition, the current system of scientific publishing provides unequal access to scientific resources, as access to publications is curtailed by shrinking library and departmental budgets.

Access to Resources - SIMs are more likely to be successful when structural conditions provide access to key resources, including: financial support; opportunities for publication; and employment opportunities to

secure additional intellectual prestige. They will also abide by Collins' "law of small numbers"—that is, the theory that only a small number of groups can survive at the cutting edge at one time in high-consensus, rapid-discovery disciplines such as the life sciences (Collins, 1994, p. 157).

In the case of the open science SIM, access to resources is not constrained by financial conditions. The participants in this research have full access to the means of resourcing their movement with their existing funding regimes, without necessarily needing further financial or technical support. The changes to work practices the open science practitioners advocate fit within their normal programs of scientific funding (and some, such as free access to research publications, albeit after a short embargo, are strongly recommended by funders such as the National Science Foundation (NSF) in the USA and the Wellcome Trust).[98] As the OLN case study shows, small amounts of funding for open science projects are also being won. For example, the Koch laboratory won an Addgene award[99] for $5000 USD based on Maloney's write up of their open science procedures, Bradley's SubMeta funding for the Open Science Challenge, Todd's schistosomiasis

[98] The NSF's grant regulations refer to Dissemination and Sharing of Research Results, which strongly recommend "at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants" See: http://www.nsf.gov/pubs/manuals/gpm05_131/gpm7.jsp#734. The Wellcome Trust's "Position statement in support of open and unrestricted access to published research" states: "expects authors of research papers to maximise the opportunities to make their results available for free" and "affirms the principle that it is the intrinsic merit of the work, and not the title of the journal in which an author's work is published, that should be considered in making funding decision" See: http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD002766.htm

[99] http://www.addgene.org/awards/#past

project funding from the Australian Research Council and UNICEF, and Todd has also been successful in finding funding from other sources for related projects, specific funds to aid the development of the LabTrove OLN software at Southampton University.

With regard to employment, Frickel & Gross say that "the centralisation of employment opportunities in one institutional arena [academia] has tied the fates of intellectual fields directly to the fortunes of higher education" (Frickel & Gross, 2005, p. 214). Thus, a successful SIM needs to provide employment avenues or opportunities. Although the open science movement is not creating many new employment opportunities, two points can be made about this: firstly, being involved in open science does not appear to harm one's employment opportunities; indeed there are a few instances where the scientific social network has helped its members find new positions. Secondly, although it does not yet appear to be a broad avenue of opportunity, new roles have arisen in the management of online communications. For instance, Zivkovic, featured in the blogging chapter, has worked as PLoS community manager and recently moved to become the editor of *Scientific American's* blog network.

Gaining intellectual prestige is a pre-condition which I would argue is less relevant or neutral in this case. Frickel and Gross give examples of social conditions for historical SIMs which allowed a particular school of thought to flourish. However, two points should be made about the SIM

studied in this thesis—it is not an *intellectual* movement as such,[100] not based on a single intellectual paradigm, but one which is revolutionary because it moves scientists out of the relative isolation of their home laboratories to much wider worlds available online—and one which seeks to improve the *intellectual outputs* of science. Also, as shown in this research, there do not yet appear to be many negative ramifications of doing research in an open manner (exceptions within this research include the 13% of bloggers who commented negatively on Redfield's *Arsenic Life* blog post and the graduate students who chose not to use OLNs when given the option), although members of the open science movement have had to overcome various types of resistance to their ideas and actions. It is possible that in future the weight of peer indifference will take its toll on those involved in this movement, and although they will continue to work in an open manner, the current structures of science will prevail. However, even under those conditions, it seems unlikely that intellectual prestige is harmed by open science approaches. Others may regard open science as less prestigious, either due to perceived restrictions on formal publishing outlets, or a perception that open access publishing is less rigorous (for example, some scientists still mistakenly believe that open access journal

[100] Note that Frickel and Gross state that "SIMs can vary in intellectual aim and scope... Some SIMs distinguish themselves through new methods and rule-making procedures ... Other SIMs aim to alter the boundaries of existing scientific or intellectual fields" (Frickel & Gross, 2005, p. 208).

publications are not peer-reviewed).[101] At point of publication, however, it is the actual science produced that matters.

Recruitment - The greater a SIM's access to various micromobilisation contexts, the more likely it is to be successful. Micromobilisation contexts are in essence recruitment strategies. Frickel and Gross borrow the term from the social movements literature, where it means the "existence of local sites in which representatives of the movement and potential recruits can come into sustained contact with one another" (Frickel & Gross, 2005, p. 219). In the context of this research, the Internet augments the locality by providing online networks where scientists can indulge in water-cooler style conversation, like FriendFeed; share their everyday work, via OLNs; and hold both formal and informal longer-form discussions of work, in science blogs. "People think, ... 'well it's not real because you haven't met these people' but it's just as real as anything else, it's the human interaction, and you can still be friends, even if you've never actually physically met" (Grant interview).

> Like social movements, SIMs do not just happen, but once their key ideas are formulated, they must be orchestrated, coordinated, and collectively produced. For this to occur, opportunities for gaining access to resources are imperative. (Frickel & Gross, 2005, p. 213)

In order to grow, the SIM must continue to recruit new members and gain access to appropriate resources. As noted below in Section 6.6.2, Low

---

[101] See Peter Suber's summary of open access for a general guide—in general, most open access journals conduct the same level of peer review as non-open journals. See: http://www.earlham.edu/~peters/fos/overview.htm

Barriers to Entry, "membership" is an informal concept, and thus "recruitment" is also not strictly definable in terms of numbers of recruits. As is discussed above in the section about SIMs, success would be measured by if and when the various goals of open science became part of normal science—that is, if the much larger group of scientists who aren't currently part of the open science SIM/CM became open. Alternatively, a lack of success in recruitment would mean that the current levels of activity were maintained or dwindled.

Recruitment activities and products are influenced (aided or hindered) by the activities of the movement. Recruitment activities include, in this case: books, journal articles, newspaper articles, conferences, workshops at disciplinary conferences, training in particular tools, speaking tours and individual talks. These are detailed in Table 15, below.

Table 15

Illustrative Recruitment Activities and Products in the Open Science SIM

Publications	
Books	O'Reilly Beautiful Data
	The Open Laboratory (edited best of science blogging posts) 2006-2011
	Bradley group manuals
	Solubility Challenge book
	Reaction Attempts book
	Melting Points book
Articles	Many mentioned throughout this thesis; also see Appendix G.

Activities		
	Conferences	Science Online (USA) 2007-2011
		Science Online London 2007-2011
		The Open Science Summit
	Speaking tours and talks	Michael Nielsen Open Science, TED Waterloo
		Michael Nielsen speaking tour
		Mat Todd Google Tech Talk
	Workshops/Training	Future of Research Communications
		An August 2011 email thread on the Open Knowledge Foundation mailing list came up with 17 workshops
Organisations and Initiatives		
		The Open Knowledge Foundation
		Panton Principles
		Science Commons

Resonance - "The success of a SIM is contingent upon the work done by movement participants to frame movement ideas in ways that resonate with the concerns of those who inhabit an intellectual field or fields" (Frickel & Gross, 2005, p. 221). Academic publishing is experiencing a period of unprecedented upheaval with the advent of digital publishing and the Internet. Publishers have managed the transition to digital media (mainly in the form of PDFs of journal articles) quite well. It is now rare to find a scientist who regularly visits a library, and many have not set foot in a library for several years. However, other aspects of publishing, including peer review, who pays to publish, and impact factors—are all under review.

Open science has to some extent been able to take advantage of this instability, as academic budgets shrink worldwide, their introduction of low-cost ways of working and distributing and disseminating knowledge may be timely.

Thus, it can be seen that the SIM of open science fulfils Frickel and Gross's definition of a SIM, and is not greatly disturbed by various forms of opposition to its aims. The next section describes the theoretical conditions which allow a Computerisation Movement to form, and compares these to the data from the case studies, before proposing a combined theory which accounts for the findings of this research.

6.3 Computerisation and scientific work practices

Although the conditions for the existence of a SIM have been shown to exist, they could not have emerged without the use of the online tools described in this research. The open science movement consists of interwoven ideals and practices, and the technology used to enact these practices necessarily couples the theories of SIMs and CMs.

Members of the open science movement believe that technology and science advance together. Evidence for this resides not only in their everyday use of technology, but also in their interactions with those who build it. For example, Todd was invited by Google's Open Source Programs Manager, Chris de Bona, to give a Google Tech Talk at the main Google site in Mountain View, California. Todd explained his research and what open

science was, but also asked the listening developers to create a tool for open science:

> We need something that's very much more intuitive and allows data rich things to be posted, allows links to online pages where all your data are posted, allows links to online lab books for science. So basically the structure we need is something that is an intuitive lab book online, where all the data are linked with your experiment and can easily be analysed by somebody else as part of a collaboration, [and] collaborators can post to your web page. (Todd & Google, 42:08min, 2010)

As of August 31 2011 the YouTube video of the talk has been viewed 8,642 times (Todd & Google, 2010).

A Computerisation Movement (CM) is defined as a "movement whose advocates focus on computer-based systems as instruments to bring about a new social order" (Kling & Iacono, 1988, p. 228). Kling further defined these movements as based on core sets of technologies which were the "products of social movements rather than only as the products of research labs and industrial firms" (Kling, 1996, p. 1). Kling and Iacono's analysis is "grounded in a sociotechnical concept of computerisation and a classic sociological conception of a social movement" (Hara & Rosenbaum, 2008, p. 230). This provides the first link between SIMs and CMs. Computerisation Movements as defined by Kling and Iacono share several key features with the General Theory of SIMs: mobilising ideologies, activities (collective actions) and byproducts (which they characterise as advocates, public speeches and written works), and involve collaborations of participants with diverse interests, which in the present case would include participants from several scientific disciplines.

Kling and Iacono (1998) cite five examples of computerisation movements: Urban information systems, Artificial Intelligence, Computer-based education, office automation, and personal computing. Hara and Rosenbaum (2008) further expand this list with 26 examples using their refined typology of CMs, described below. These include personal digital assistants (PDAs), Geographic information systems (GIS), Email, eHealth, and eGovernment. Meyer (2008) provides a more recent example, detailing the switch by both professional and amateur photographers to digital media.

The core ideologies of the CM revolve around a set of propositions around computers, and activists within a CM will assert or imply the following about computer technologies: that they are a "historically unique opportunity for important social changes", and essential for work productivity; that it is necessary or strongly advisable to use the most advanced equipment available; that more computers are always better; that "real life is life on-line"; and that, within a CM, no one loses—that the CM fosters "cooperation and rationality" (Kling & Iacono, 1998, pp. 235-237).

The specific features of a CM which Kling and Iacono enunciate include that the action resides in technological action frames, that is, similar to Frickel and Gross's concept of resonance, the actions of the movement will be framed by collective concepts of why the movement is important. Kling and Iacono cite the examples of computers overcoming the tyranny of distance and shrinking geographical distances as one such framing.

Hara and Rosenbaum (2008) challenged and refined this definition to provide a typology of CMs based on five sorting criteria, along a continuum bounded by the endpoints of each pair (external/internal; market-driven/ non-market-driven; wide/narrow; stand-alone/ bundled and positive/negative, which they claim will allow for some assessment of the success or failure of a movement (Hara & Rosenbaum, 2008, p. 234). Under this typology, the open science movement displays good fit with some aspects of their definition—it would be classified as *external* (to organisational boundaries such as departments or universities)[102], *narrow* (affects a restricted domain), *non-market-driven* (used by academics, and the tools themselves are free to use, for now) and *positive* (aiming to have a positive impact on society, that is, not dystopian in outlook). The concept of using *bundled / stand-alone* software is difficult to classify within the open science SIM/CM. The scientists use various tools, and sometimes a suite of complementary tools for different parts of their scientific workflow, but with conscious choice, as there is no organisation dictating their use of particular tools, and they may pick and choose which tools they use. However, it would also be difficult to engage with the open science CM and not use these tools, as they are the core communication technologies of the SIM/CM, so they could be more strongly considered as a *bundle*. These classifications can be used to help assess characteristics of success or failure

[102] Hara and Rosenbaum do not define organisations, except to say that Kling and McKim did not talk about CMs outside them.

if applied over large groups of CMs that have progressed further in terms of adoption or rollout by organisations.

It is clear from the above characterisation that the scientists involved are pushing forward and promoting certain technologies simply by using them as communication and data dissemination tools. This is balanced by their flexible attitudes towards these technologies, discussed below in reference to the low barriers to entry of the community (Section 6.6.2) and informality of community membership (Section 6.6.3). As an example, it has been interesting to observe, across the years of this research, the rising popularity of Twitter as a communications tool within this SIM/CM. Twitter became more popular among the scientists studied in 2008, and has subsequently grown in use. For at least one scientist in this study (Saunders) Twitter's widespread use has led him to delete his FriendFeed account (in early 2011, two years after this research began), as he felt the declining performance of programmatic access to the archive and search functions degraded FriendFeed's usefulness, because as "online scientists", the community needs a way to archive, reference and analyse their conversations (Saunders, 2011). Saunders wanted the ability to keep his data: "publishing to services with an adequate degree of control—which essentially means the ability to retrieve or search the archive, [then] broadcasting published items to Twitter, for the benefit of anyone who might be interested" (Saunders, 2011). Thus, as the utility of particular technologies grows or

fades, each user decides how best to organise their working life, appropriating new online tools as appropriate, and dropping others.

This thesis takes the interpretation that appropriation of products intended for a multiplicity of uses, (for example, communication technologies adaptable for different uses, such as blogs, wikis and FriendFeed), can still be considered, according to Kling and Iacono's definition above, as the core technologies of a CM. This technological core has facilitated the growth of the SIM described above and the patterns of online scientific interaction found in the three case studies.

6.4 Synthesising Theories - SIMs and CMs

The diagram below shows the contributions of both theories to the current development of the open science movement. Both theories acknowledge that a movement (successful or not) will have recognisable collective actions towards its goals (such as publications and conferences); identifiable organisational structures and practices (though, as in this case, not necessarily formal or hierarchical); will mobilise resources (via conferences and promotion through scientific journals and similar activities); and display some kind of forward movement (in the sense that their development is based on the current open science system, but taking the concept of openness all the way to the laboratory bench, and to open collaboration and dissemination).

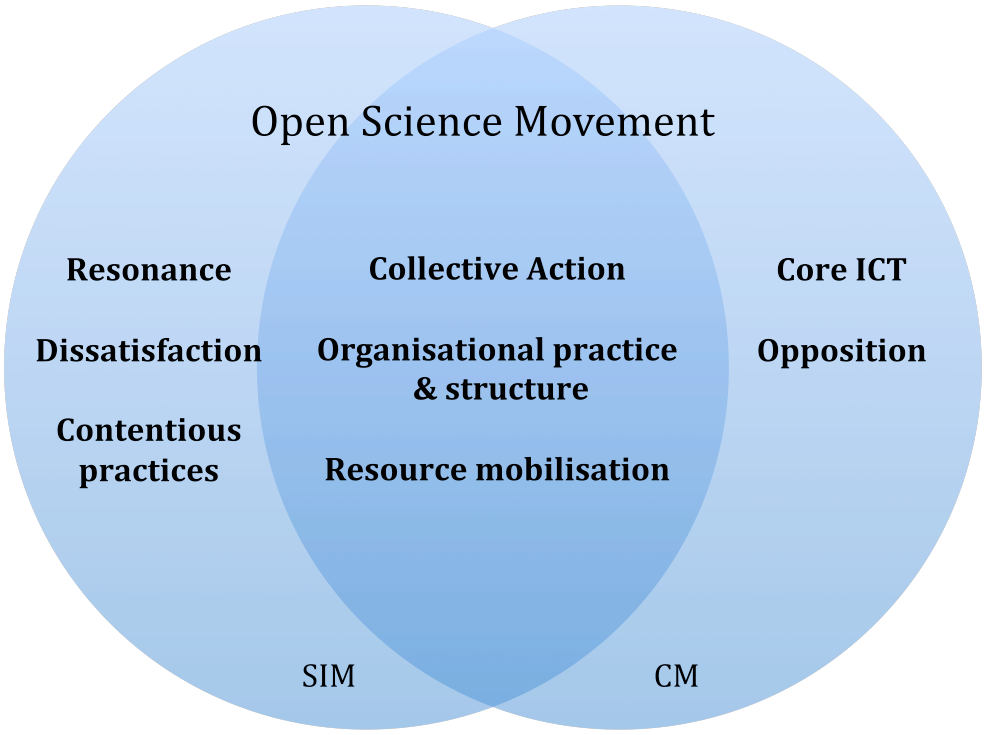


Figure 43. Combined SIM and CM theory concepts

Both SIMs and CMs, as shown in Figure 43, have common mobilizing factors which mutually reinforce - collective action, an organisational structure and organising practices (formal or informal) and the ability to mobilise resources in support of their goals. Together, they co-constitute and mutually reinforce the growth and support of an open science movement.

The theory of SIMs specifically attributes their formation to the concepts of dissatisfaction with the status quo, fellowship among its members ("this feels like the right way to do science"), and that its members are engaging in contentious practices which will encounter resistance (for example, OLNs). Computerisation Movements also contribute the core information

and communication technologies which enable these activities, and to a lesser extent an organised opposition.

Having established that a combined SIM/CM exists, the next sections elicit the key findings across the seven research aspects, and diagram the patterns of adoption and use of online tools in life science research. These patterns are central to the existence of this SIM/CM, or the open science movement, as they show the core ideologies of the movement, described above, being enacted.

6.5 Collaboration and Communication - Networking and Net-work.

Scientific collaboration and communication within the open science movement is marked by particular patterns of adoption and engagement which have been captured by using the thesis research aspects to analyse the scope of activities across the embedded cases.

Dutton's (2008) typology of collaborative network organisations (CNOs) is a useful base for classifying collaborative activities, typed as *sharing*, *contributing/deploying* and *cooperative co-creation*. These categories are explained below with examples from this research. Dutton's framework is referencing data from organisations, whereas the data from this research relates to the activities of individuals who are working towards a common goal of open science, using online tools to create and sustain a SIM/CM.

Nevertheless, there is a good fit with Dutton's categories, which suggests a universality of types of online interactions.[103]

Sharing - the ability to create linked documents and objects within a distributed network, thereby reconfiguring how and what information is shared with whom.

Example: using FriendFeed to share and comment on a variety of linked online resources, including blog posts, Tweets, other RSS feeds, and video.

Contributing/Deploying - the ability to employ social networking applications of the Web to facilitate group communication, thereby reshaping who contributes information to the collective group. In this case, contributing/deploying is highly informal, water-cooler-chat style interaction, with low barriers to entry, and aiming to solve small problems rather than collaborate on large distributed projects.

Example: keeping a blog with a blogroll of links to other science bloggers. Then feeding that blog into the sharing network of FriendFeed described above, thus enabling wider and more diverse communication options.

Cooperative co-creation - the ability to collaborate through networks that facilitate cooperative work toward shared goals thereby reconfiguring the sequencing, composition, and role of contributors.

---

[103] Also see Dutton, (2010) WWResearch Chapter 1, "Reconfiguring Access in Research: Information, Expertise and Experience"

Example: the OLNs of Bradley, Neylon, Koch, Redfield, Todd, and some of their graduate students. This could be seen as an extension of "normal" co-creation, as they are sharing, contributing and co-creating with a mindfulness/attitude towards "radical" openness—almost everything is public.

Patterns of adoption that can be identified fit in similar categories, but not necessarily in the same order, and potentially in different orders. Although sharing, deploying and co-creating are all present, the order in which they occur differs from individual to individual. For example, most of the Open Notebook scientists went open first, then started writing a blog, then discovered FriendFeed. However, others wrote blogs, then went open (see Section 6.6.3, Informal Community Membership, above). The order in which activities were adopted follows a pattern of individual sharing (blogging to keep track of one's own research thoughts, or starting an OLN), growing awareness of an audience and network (commenting on others' blogs, identifying as part of a group; for example, the self-styled Biogang), and reaching out via FriendFeed, completing small tasks together and collaborating informally.



Figure 44. Patterns of adoption of online tools observed in this research.

These patterns of adoption of online tools form part of the scientist's socialisation into the open science SIM/CM—although not all patterns are the same they show a general progression towards more openness, which may indicate a strengthening of the open science movement (for example, several of the OLN-using scientists began with their OLNs, then progressed into FriendFeed use and blogging, whereas most non-OLN users begin by blogging). As Fuchs points out, at "the uncertain and intensely competitive frontiers of a science, rapid discoveries and innovations are being made"; that is, scientists' jobs concern coping with uncertainty, so they should be good at innovating in this way (Fuchs, 2002, p. 34).

Table 16

Examples of patterns of online tool use by order of adoption

	ONS	Blog	FriendFeed
Neylon	1	2	3
Bradley	1	2	3
Koch	1	2	3
Maloney	1	n/a	2
Salvagno	1	2	3?
Todd	2	1	3
Hooker	n/a	1	2
Beltrao	n/a	1	2
Szczęsny	n/a	1	2
Jensen	n/a	1	2
Grant	n/a	1	2
Saunders	n/a	1	2
Singh	n/a	1	2
Lindenbaum	n/a	1	2

6.5.1 Informal small task collaboration

A feature observed in the informal networks described above is short-term collaboration for problem-solving, which the FriendFeed-using Biogang described as "bursty work"—individuals coming together to solve a specific small problem or issue; then dissolving, later re-forming with a different task focus and a somewhat different group of loosely-connected people. These small problems may be solved within minutes, hours, or days, and are usually so small that funding and attribution of individual contribution to the final scientific product are not required.

Bradley's *dosol* (do solubility) spreadsheet, described in Chapter Four, OLNs, provides a good example of this, as anyone may request an experiment; additionally FriendFeed, Chapter Three, highlights several examples of informal small task collaboration.

Occasionally the problem or task required is slightly larger, or expertise is required from a specialist such as an experienced programmer, in which case attribution or co-authorship may be offered if a publication results, but the problem is still approached in the same informal manner as a small task. Prior to use of these types of online tools, it would have been difficult to either find or make the approach to a programmer within one's institution, and the informality of the interaction could not be assumed. This approach to small task collaboration with its lowered barriers rewards scientists with new approaches to scientific questions, as the scientists involved can easily draw upon expertise in related and unrelated fields to their own. In addition, as shown in Section 6.6.2, Low Barriers to Entry, the tools and

practices themselves do not require much expertise to operate, so the scientists can concentrate on the collaboration activity.

Returning to Sonnenwald's description of the stages and factors of collaboration—Foundation, Formulation, Sustainment and Conclusion, shown in Table 1 of the Introduction to this thesis—it can be seen that informal small-task collaboration skips many of these stages and factors (Sonnenwald, 2007). The relationship it bears to formal collaboration can best be described as contingent and contextual, as it may or may not go through various stages or be influenced by particular factors depending on what the task is and what the expected outcomes are.

6.5.2 Formal collaboration

In contrast to informal small task collaboration, the OLN and FriendFeed cases have also shown examples of facilitation of "traditional", more formal collaborations, in the sense that although the initial connection happened online and informally, more extensive co-working is done than with the small task collaboration described above, and a scientific paper or papers is the expected outcome. Most participants remarked that prior to their involvement in scientific social media, they would have found new collaborators at conferences in their field, but participation in online activities has also enabled connections with scholars in other disciplines whom they would not otherwise have encountered.

Examples of collaborations which produced formal scientific publications highlighted in the case studies include the Saunders et al.

IMSB conference report (2009) in *PLoS Computational Biology*, Lindenbaum, Su et al. (2010) in *Nucleic Acids Research*, and Redfield and her collaborators, who hope to publish the results of their GFAJ-1 experiments in a peer-reviewed journal.

These publications demonstrate three points. Firstly, the members of the open science SIM are producing normal science, that is, work which is accepted by the whole scientific community as it goes through the standard processes of peer review prior to publication in a journal, which thus validates this mode of working. Secondly, the open scientists display a pragmatic attitude towards these formal publication styles, although they are happy to discuss their work and other peer-reviewed work informally online, they also actively participate in these formal scientific processes. Lastly, these formal outcomes are produced in an ad hoc, informal and public atmosphere—these papers are created quickly, often by people who met online and would not normally have the opportunity to work together, as demonstrated by the examples above.

6.5.3 Information filtering

With regard to science communication, participation in these networks can appear to outsiders to contribute a higher volume of information to the scientists' ever-present and growing "data deluge". In practice, participating scientists use their networks to control the deluge with strategies such as peer-filtered information-gathering, described in the

FriendFeed case, and they handle potential information overload by spreading the filtering tasks over their network.

Filtering is done by at first paying attention to one's network, reading and assessing the volume and relevancy of information to one's own work. Items that do not qualify as relevant will have their information streams either removed from the feed, or downgraded to a category like "read later". Several participants remarked that they paid most attention to a few people in their network whose "signal to noise" ratio was highest for them—these "topic filter" people tended to pick up relevant information from the information stream. An additional strategy is creating search filters on particular topics of interest, for example Google alert emails. However, this tends to be a less important type of filtering—most of the scientists relied on their personal networks to filter information for them.

With these strategies in place, the scientists seemed content that they were receiving manageable amounts of relevant information. However, it is of course still possible that this peer-filtering management is imperfect, and that the groups with similar topic interests were missing information useful to them, but this was not a concern expressed by any of the participants. It may even be that there are parallel groups of information seekers who do not overlap circulating in the Internet information sphere—however the scientists seem to trust that search engines (predominantly Google) are capable of finding everything of relevance on their topic.

6.5.4 Public Engagement

Despite nearly all of their communications being open and findable on the Internet, many of the scientists studied in this thesis are not seeking to engage with the general public, or a new public. Instead, they are mostly concerned with engaging with their peers, the scientific public, which given that scientists today number in the millions (see Chapter Two, Footnote 15), is still a large number of people. Participants in this research are having valuable discussions on a wide range of topics, not just about their own disciplines but also current issues around openness, publication, research funding and research assessment. They are also engaged in seeking connections with other scientists within and outside their disciplines, recruiting students and postdoctoral fellows, doing research and sharing research outcomes. The cases studied here focus on scientists working with their peers rather than the public, and as such their presentation of themselves and their data is mundane and technical, and not easily accessible to those without expertise in the topic. In short, their online activities tend to revolve around using online tools to promote a SIM of open science.

Some members of the general public are interested in and follow this discussion. For example, undergraduate science students (some have become participants in open science projects, for instance undergraduate David Bulger, as discussed in the OLN chapter), and academics from other disciplines such as the sociology of science.

A few of the scientists participating in this research are, in turn, engaging with the general public: blogging about patient empowerment (Mesko), blogging about the "messiness" of everyday science (Redfield, Grant), or starting competitions (Bradley, Todd). In addition, many scientists outside this study are heavily engaged with the general public. Outside the scope of this research, many current cases exist of successful public engagement; for example, citizen science, via the crowdsourced astronomy website Galaxy Zoo, or building past climate change records via transcription of ship's logs with the Old Weather project,[104] is enjoying a renaissance due to the capabilities afforded by the Internet, and some science bloggers actively engage with the public (for example, Scienceblogs.com's Pharyngula, discussed in the Blogging chapter). This blog engagement tends to be in the area of polemical and political issues (such as, for example, evolutionary biology or climate change). Others who engage with the public extensively are those who blog in a science journalism context—for example, blog networks such as *Scientific American Blogs* (including their blog editor Zivkovic, one participant in this study), many of the *Scienceblogs.com* blogs, and *Wired Science*.

As discussed in the Arsenic Life story in the science blogging chapter, some scientists are wary of drawing the non-scientific public's attention to the everyday struggles and disputes of science. They are concerned that displaying these scientific debates (which are not always polite or

[104] http://www.oldweather.org/

respectful) provides fuel for sceptics and fundamentalists who seek to undermine science. However, most scientists do not express this concern, and those involved in this research see far greater benefit than cost in open scientific debate.

6.6 Roles - Entry and Participation

6.6.1 Participation

The accounts of the research participants about how they discovered or were enrolled into the networks or online activities provide insight into how and why people join these networks, and why most of them are scientists or ex-scientists. Participants find that connecting with the open science movement extends their roles into interdisciplinary spaces which they may not have inhabited before as they become part of the SIM/CM of open science.

As described in each case study, the sites and activities conducted on them are open, and do not have strict membership rules or access controls. In some cases, users must sign up to the hosting website to view, add or edit content (for example, FriendFeed or Wikispaces), but this signup is free and unmoderated. In theory this freedom of access to scientific information and discussion could attract cranks or "trolls"[105], but in practice this appears not to be the case.

[105] Individuals who deliberately post inflammatory content online in order to create conflict.

The natural control on this type of behaviour in the case study scenarios may be the mundane nature of discussion. As mentioned in all of the case studies, the discussions are often centred around the everyday scientific work of the participants. Interaction is largely context-based, as participation is governed by the problem being investigated or the topic being discussed.

There are examples where discussion has become controversial, but even in these cases the argument did not stray far from the scientific topics at hand. The analysis of Redfield's blog post on the "Arsenic Life" story in the previous chapter demonstrates this, as although the commentary becomes heated and occasionally abusive, it stays on topic and is not derailed by trolls.

These examples point to the characteristics of the core participants in these networks. Participants are drawn towards contributing due to an interest in open science and related issues (either because they were already doing their science openly because working online felt natural to them, or they were interested in the concept of open science and wanted to learn more about it), or they interact with each other based on their wish to fulfil the work goals of their scientific specialisations, and were initially drawn to these activities by online searches, or various types of connections with others who were involved (for example, they knew each other from other forums such as Second Life).

6.6.2 Low Barriers to Entry

There are two aspects of ease of entry to networks—low barriers due to informality of interactions, and low barriers due to the ease-of-use of the online tools and resources.

In contrast to more formal top-down e-Science initiatives, such as Grid-based approaches, which require a high degree of technical skill or assistance to begin using, the activities reviewed in this thesis are more open because any interested scientist can join in with little or no training or guidance. Entry into the community of online networked science is informal and self-determined. There are no membership criteria. Scientists interested in these topics may choose to read and not participate in active discussion—they are known as "lurkers". In the context of open science lurking is not productive, but neither is it a drain on the movement, it is simply neutral, until the reader /lurker decides to become involved and start their own blog, or begin commenting on FriendFeed or Twitter. Those who make themselves visible in this manner are usually welcomed into the community, often with encouraging comments from highly active participants.

Flexibility of use is a key component of this access. Participants come and go, or vary the depth of their involvement according to the time and motivation they have for different initiatives or projects. However, the movement is kept intact by a core group who are heavily involved in several aspects of the SIM—exemplified in this case by scientists like Cameron Neylon and Jean-Claude Bradley, or self-identified groups like the

Biogang, whose activities across FriendFeed, open notebooks and blogs are described in this thesis.

Table 17

Technology use by networked scientists (mentioned in interviews)

Type	Usage	Most common mode of use
Email	One-to-one messaging	Private
Skype - online telephony	One-to-one voice or video calls, voice conference calling	Private
Web - blogs	One-to-many blog posts, discussion via comments	Public
Wikis	Keeping laboratory notebook, storing data, writing collaborative papers,	Public
FriendFeed	Repository for other Web items of interest, discussion space for blog posts, longer-form discussion of microblogging like Twitter, live conference reporting and discussion	Public or semi-public (open to anyone who asks)
Twitter	Sharing links; short comments on issues, live conference reporting	Public
Dropbox	Sharing files	Private
Google Documents	Writing collaborative papers, sharing files	Mix of private and public
Google Spreadsheets	Storing publicly-accessible data, backing up laboratory	Mix of private and public

Type	Usage	Most common mode of use
	data	
YouTube	Demonstrating laboratory procedures, publishing video of results	Public
SciVee	Demonstrating laboratory procedures, publishing video of results	Public
SecondLife (SL)	Creating molecular simulations	Public (users must log in to SL but access is free)

The online tools favoured by these scientists do not require great investments of time or skill to set up and use. The "Like" and "Comment" interactions on FriendFeed which are used by participants to signal interest and start and maintain discussions, are representative of this, with multiple research participants commenting on the ease of use of the site as an important factor for their continued use. (see Section 3.5, Research participants' Normal Use of FriendFeed, for more detail). No programming knowledge is required to participate, and the use of these tools does not require the assistance of a local or imported expert. In addition to ease of use, the most common features of popular online information-sharing and discussion tools are openness (anyone can join, and default features

encourage open groups or networks rather than exclusivity), and that they are free to use.

6.6.3 Informal community membership

Joining the community is simply a matter of participation, entering the conversation by commenting, liking, re-tweeting, adding people from the "coherent group" to your feed reader or social network, and continuing to participate in the conversation. Over time, as people become comfortable with interacting online, they may also start to contribute more, writing longer blog posts, participating virtually in conferences, testing and using new online tools as they become available, and eventually, going to offline meet-ups such as the science blogging conferences. Several of the participants in this research have met physically at these meetings. Prior to these meetings, they had often interacted online every day for some time, and as Lang comments about his first meeting with Bradley, meeting in person can be interesting: "I met him for the first time 10 minutes before our talk started, although we'd been collaborating for 18 months. ... he kept telling me it was hard for him to think of me, when he hears my voice he thinks of my Second Life avatar" (Lang interview).

It should be noted that "the community", although fluid in membership and flexible in approaches to using technology, is nevertheless something that all participants found easy to identify. After conducting the first few interviews the same names kept being mentioned as people who would be useful to interview. These people are the coherent group which drives

forward the open science agenda. There are many more members of this group than have been interviewed for this study, as it was not possible to arrange interviews with everyone recommended, and also during the course of the research the group has grown, and new members have found the movement via the popularisation of new tools such as Twitter. However, as discussed in Chapter Two, Research Methods, data saturation was reached with the number of participants interviewed.

From the evidence of this research, patterns of adoption of online tools tend to be from individual to community. Users often start a blog writing for their own reference, albeit publicly, then become conscious of an audience and interact with them. Through this they become aware of and join a wider community such as FriendFeed and contribute to it. They may then participate more frequently and help to grow the community by enrolling others, although as stated above their participation may wax and wane depending on their needs. Open notebook science is part of this in that it is individual work but also consciously done for the scientific community, to encourage others to practice open science.

Participation in these activities is time-consuming. Most research participants considered their online activities as something they should do in their own time, unless they are one of the few for whom online work is part of their job, like Zivkovic, who is the Scientific American blog editor. Other exceptions are made for ONS users, for whom creating an online laboratory notebook is definitely part of their daily work activities. Most

will catch up and comment on the information filtered to their computers or phones at least once a day, as soon as they wake up, or on their morning commute, lunch break or late in the evening. Many will keep their Net streams (FriendFeed, Google reader, Twitter) flowing all day, only shutting them down if they need to do "thinking work" like writing a grant proposal or paper. There is a great variety of approaches to production and consumption of the online news stream.

6.6.4 Bottom-up Innovation

The open science movement is also shaped by the provision of choice. Participation in these activities is fluid and flexible, as demonstrated by the informality and ease of use of the online tools adopted by the group. The online tools provide affordances for different types of interaction, and the participants use them for different things. For example, FriendFeed for some is a problem-solving network, for others it is a discussion space, for yet others it is a reference-finding tool. At different times it can be used for all three purposes by the same user. Distributed activities are supported by different online tools, completely separate from each other technically but linked by the participants' use of them; for example, aggregating conference blogs and tweets in a FriendFeed group, or embedding Google spreadsheets and YouTube videos within an ONS wiki hosted by Wikispaces or Open Wetware.

Participants do not, in general, use institutional infrastructures to facilitate these interactions (apart from institutional Internet access), nor do

many of them think that it is important or necessary to inform their institutions or departments of what they are doing. When they do consult in advance, the reactions of their employing institutions have been positive—partly because involvement in the movement, and use of these tools, does not in general require institutional support. This can be seen as an advantage. Those who want to connect with other disciplines, and the public, have no interest in using closed institutionally-based software or drawing on the already-scarce resources of the academy. However, it should also be acknowledged that despite not requiring institutional support, some movement members hope that future technology will bring a more unified solution (although not necessarily institutionally-run) in the form of a unified scientific network which fulfilled all of their requirements, but acknowledge that this is unlikely to happen in the near future. In the interim, they use the tools they have, which are inclusive, free tools that encourage a wide user base.

Serious attitudes to data management and preservation particularly among OLN users underlie their flexible approaches to participation. As described in the OLN chapter, in both public and private debates users have discussed strategies for archiving and backing up data and shared their disagreements about the platforms they use. These disagreements are not strident, and both sides embrace what is possible with current technology. The attitudes range between "we know it's not perfect but it's better than paper" and "this is one of the very early steps to a fully

integrated end-to-end scientific publishing system using the full power of the semantically linked web, which we must keep striving towards".

OLN users display different levels of frustration and dissatisfaction with current tools. The degrees of difference seem to be in how much time and organisational effort they are willing to devote to have a quasi-integrated solution which can incorporate and semantically understand all of their data types, and those that are reasonably happy that "at least" the text of their laboratory notebook is fully indexed by Google. As demonstrated in the OLN case study, the LaBLog used by Cameron Neylon is heavily structured, with particular rules for which data constitutes an OLN entry, whereas the wikis used by Bradley and Koch are more free-form text and relatively unstructured. All users seem to agree that better tools with a specific scientific focus will eventually evolve, even if they do not exist yet. Despite this variety of bottom-up approaches, as each individual makes their own technological choices, these OLN users have nevertheless united into a single open science movement.

6.7 Gatekeeping and Gatekeepers

6.7.1 Permeable Disciplinary Boundaries

Scientists studied in these case studies have flexible attitudes towards disciplinary boundaries. Traditionally, outreach to disciplines other than their own is relatively rare among scientists (Gabler and Frank, 2005), but those studied here are more willing to engage with scholars outside their own disciplines.

This openness to interdisciplinarity has been illustrated within the case studies, for example: Bradley (chemist) and Lang's (mathematician) collaboration on many aspects of Bradley's Open Notebook Science project, including co-authoring papers and the Solubility Challenge (OLNs, Chapter Four); Hooker (life scientist) participating as a judge in Bradley's (Chemist) Solubility Challenge (OLNs, Chapter Four); Neylon (biophysicist) co-authoring papers with librarians (FriendFeed, Chapter Three); Redfield (microbiologist) collaborating with Kruglyak, Rabinowitz and Reaves (also microbiologists, but with equipment Redfield does not have—a mass spectrometer) to test claims about the GFAJ-1 bacteria (Science Blogging, Chapter Five).

There are several reasons for this flexibility. Firstly, the scientists have met these extra-disciplinary colleagues online via FriendFeed, blogs, or Second Life, interacted with them and formed a professional opinion of them, much as they would with their disciplinary colleagues via conferences or meetings. This working relationship and daily or almost-daily discussion allows this assessment and breaks down disciplinary barriers. Shared connections via online tools and the same attitudes towards open science have prompted these connections, supporting the thesis that a combination of Computerisation and a Scientific/Intellectual Movement with shared aims is ushering in new forms of science communication and knowledge production.

Secondly, these scientists recognise that to fully exploit the data they are discovering on the Web or are producing themselves, a greater array of expertise is necessary than is traditionally found within their discipline (Arms & Larsen, 2007). Reaching out, for example, to programmers and librarians allows them to exploit each other's skills for mutual benefit. This outreach also serves as a form of enrolment for the open science SIM—contact with a wide range of outside influences makes more people aware of its existence and serves as promotion of the SIM.

6.7.2 Going Beyond Existing Boundaries

There is also evidence that the scientists studied in this research are using online tools to connect and "do science" with people beyond their disciplinary and interdisciplinary peers. Bradley's *Open Science Challenge* competition, detailed in Chapter Four, is an example of this outreach. His ONS Challenge seeks to create a public database of chemical solubilities and melting points, and contributing to this data is within the reach of anyone with the requisite equipment, which in reality means a chemistry laboratory. Several undergraduate students have participated in the Challenge. There are numerous other examples of citizen science, as scientists reach out to the public for raw data and data refinement. This is citizen science facilitated by Internet technology, scalable to thousands of participants and millions of data points (for example, the Galaxy Zoo, part of the Zooniverse group of projects).[106]

[106] http://www.zooniverse.org/

Another example, also discussed in the OLN chapter, comes from the Koch biophysics laboratory in New Mexico, where Andrew Maloney wrote his PhD dissertation on a public wiki, and welcomed feedback on it from anyone. Maloney's thesis examiners received a PDF version of the wiki text to assess, and were made aware that the thesis was public. They did not have any issues with its public nature, and Maloney's doctorate was awarded after minor corrections, and his thesis was subsequently seen by a much larger audience: as of August 2011 it has been viewed over 6,000 times, and the OLN chapter, Section 4.7.6, details Maloney's open science "success story" after a group from UCSC found his thesis and was able to reproduce his methods much faster using the data in his OLN). Bradley's group has also graduated a PhD student who did a significant portion of their work openly.

Another finding is in relation to "ignoring" institutional boundaries. Most of the scientists in this research did not consult anyone in their employing institutions before starting to blog or use FriendFeed, and in particular most of the Open Notebook Scientists, who were publishing their data on the public Internet, did not think it necessary to consult with their institutions, even though their actions may have contradicted institutional recommendations for recording laboratory data. This could reflect scientists' self-perception of independence in terms of their approaches to research.

The data from the OLN case supports the research proposition that scientists are going beyond existing institutional boundaries to get their work done openly, reaching out to scientific peers and sometimes the public, and the evidence suggests that their work been accepted within and across their disciplinary communities, that is, although many in their communities are unprepared to work in public themselves, they accept the products of open scientific work as normal artefacts of science (generally peer-reviewed publications, or in the case of PhD theses, committee-examined work).

6.8 Status and Rewards - Costs and Benefits of Participation

The questions of who gains and who loses from participation in these networks, or the distribution of rewards and costs, is difficult to assess—something that is perceived as a cost may turn out to be a benefit, in that the person's time spent on it gathers rewards such as new ideas that they turn into new research, or the time and effort spent setting up a platform for open science and starting to use it may be rewarded much later as publication turnaround time is reduced, or citations of work begin to rise. This thesis showed some obviously positive examples of this, for example, the citations of the Koch laboratory's open data in a quantitative biology paper, Todd's story of being able to generate results faster for his schistosomiasis research, and Neylon and Bradley's OLN work being featured in a *Nature* editorial. Over the three-year span of this research, there has also been some change identifiable with regard to individual and

institutional status and rewards, much of it positive; for example, the world's first completely open doctoral dissertation, Andrew Maloney's biophysics thesis in the Koch laboratory at the University of New Mexico, was written on a public wiki.

Systematic examination of the data can identify further costs and benefits. Lassi and Sonnenwald's (2010) synthesis of research on collaboratories allocates costs and benefits to categories, including careers (credit within, competition, rewards and concerns), personal factors (other than career), disciplinary and scientific advancement, costs of participation and costs to develop and sustain collaboratories. These categories can also usefully be applied to the present analysis with regard to the costs and benefits of participating in the open science SIM, and some of the relevant categories are discussed below.

With regard to *credit*, within the open science community credit is widely given and acknowledged—the friendly nature of their interactions ensures this, although it is not enforced by any visible means. Indeed, some data produced by research participants (for example, Bradley's solubility measures) does not require credit to be given for use, both within and outside the community. They produce their data with a CC0 (Public domain)[107] licence, which means that no attribution or credit is required when using the data. This is to avoid the problem known as "attribution stacking", or "cascading attribution", where data use from several sources

[107] http://creativecommons.org/publicdomain/zero/1.0/

would lead to a lengthy list of credits and attributions. Science Commons describes the potential problem as: "Would a scientist need to attribute 40,000 data depositors in the event of a query across 40,000 data sets?" (Science Commons, n.d.). This seeming altruism is therefore a design feature of open science, as to operate otherwise would render use of the free data impractical.

Competition is constantly cited as a risk or a barrier to open science by those outside the community, yet it is rarely observed. Scooping would be relatively uncomplicated to prove, yet may not be an issue within open science at the moment, due to the research niches occupied by the participants, and also the fact that the data is open—if others wish to use it in their research, they are welcome to, with no penalties, as discussed above.

Career rewards for those using online tools and participating in open science belong mainly to the categories of learning from colleagues and other disciplines, and increased efficiency because of shared resources, and also shared tools. There is already some anecdotal evidence of increased citations, or gaining citations directly from open data, and co-authorship has been offered in several of the cases. In terms of prestige, these scientists enjoy mutual respect within their community, but acknowledge that the lack of adoption by their own work groups and geographically local colleagues can be dispiriting, and they rely upon their virtual networks for encouragement.

The current open scientists may also be enjoying some early-adopter advantages. The novelty of the tools they use and their open science practices attract attention from the mainstream scientific media, demonstrated by the editorials and articles in prominent publications such as Nature and Science listed in Appendix G.

Career concerns revolve around two areas, firstly that there is currently low prestige and little or no credit given for producing data for others' use in the biosciences, and secondly that activities like blogging and FriendFeed are considered "risky" in terms of career advancement. Risk has been applied most often in the criticisms levelled at academic bloggers, reviewed in the blog literature and in the sample comments from Redfield's Arsenic Life blog post (Kirkup, 2010; Tribble, 2005). However, as assessed at the beginning of this chapter, the participants themselves do not subscribe to this notion of risk—to them their activities are a natural extension of normal science. Their involvement in the open science movement can also be seen as part of an effort to create reward structures around these kinds of knowledge-sharing activities.

Personal factors The scientists participating in this research often mentioned that they felt like they belonged to a community, most often talking about their participation in the "open science community" or "open science movement". Receiving comments and "likes" makes them feel supported and encouraged in their work, and some even feel like they have a measure of "Internet fame", as graduate student Salvagno remarked

about the attention he received from his use of an OLN in the Koch laboratory. The low overhead of "liking" something on FriendFeed had a particular effect of making participants feel like they were participating in the community, and supporting its efforts, by a click of their mouse.

Participants also enjoyed their position as early adopters of new tools, and as mentioned in both the FriendFeed and OLN chapters, they will begin experimenting with new tools early, and their trials and early use will spread rapidly among the community, accompanied by much online discussion about the affordances of the tools. In particular, this adoption happened with Twitter (which has remained popular and has been integrated with other tools) and Google Wave (which most of the community joined as soon as it was available). This pattern can again be seen in Google+, launched as a public beta in June 2011, as of time of writing (July 2011), everyone participating in this research has joined it and is posting information there, with ensuing discussions happening in the comments, as shown below in Figure 45. There is strong evidence from this research that participants enjoy being first or early movers in the online environment.

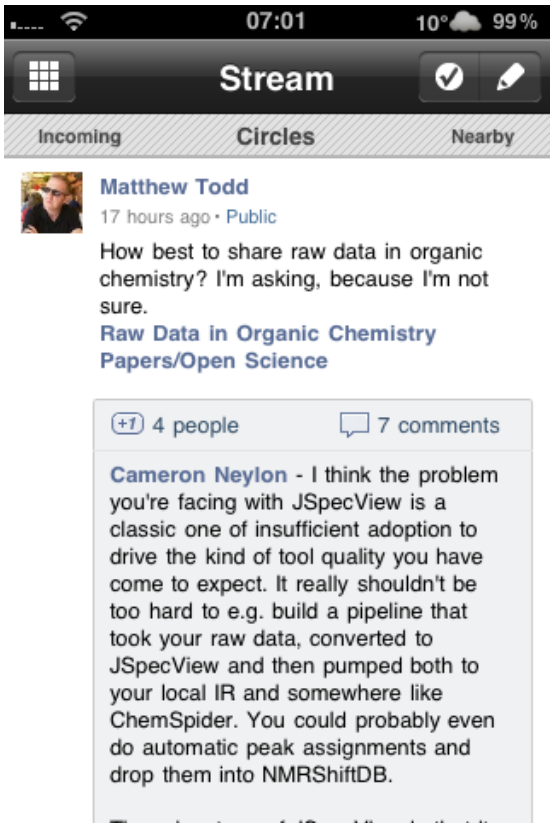


Figure 45. Todd and Neylon conversation in Google +

Disciplinary and scientific advancement - those involved in the open science movement firmly believe that their evolving work practices will speed up their research and improve research quality. Examples from this research include the effect of Maloney's open research in the Koch laboratory, which helped another laboratory in California reproduce a particular research technique in less than half the time it originally took to perform. Similarly, Todd's research on drug discovery in schistosomiasis drug development has been accelerated by the contributions of external scientists in the open research model he has adopted.

Community costs, in the form of lurkers or passivity. As seen in the FriendFeed chapter, lurkers, as opposed to active contributors, definitely

exist. However, at the cutting edge of a SIM/CM, the number of people driving things forward does not need to be large (Parker & Hackett, 2007). People active in the open science SIM/CM are not worried by lurkers, as although by definition their nature cannot be known, the open science movement members view witnesses to their work as positive. At this time, there appear to be enough active members of the SIM to sustain and continue to grow their activities, and the publicity of their efforts continues to grow, as evidenced by, among other things, the growth in online science conferences at which open science is a prominent topic of discussion (for example, Science Online UK and USA, SciFoo).

Costs Economic costs are low in terms of infrastructure, at this stage at least they are appropriating and using existing tools. Open scientists are using free online software, therefore their costs are negligible. They host most of their work and data in the cloud, and the only costs involved may be for backup servers to protect data, or to pay the small fees (less than $100 USD per year) for a self-hosted domain and blog. Specialist staff and developers are not usually hired—again, with some exceptions—Neylon's LaBLog originates from a Southampton University project which requires full-time staff to run it, although the software is free and can also be self-hosted.

Contribution barriers include time costs, cited several times by participants in this research as a difficulty. They make time in their personal lives to work on certain aspects of open science, such as blogging or

catching up with their social network, and only feel comfortable using working hours for these tasks where they can be obviously related to their work in a traditional sense, such as maintaining their OLN.

Participation barriers/Infrastructure barriers - at the moment, due to the use of off-the-shelf or free internet-based software, these are negligible, as all that the participating scientists really need is an Internet connection and a browser.

Thus, we have seen that the evidence presented in this thesis supports the proposition that a combined SIM/CM exists in the form of the open science movement. All that remains is to evaluate the main research question, "Is the use of online tools revolutionising, reinforcing or having unintended consequences for the research practices of academic life scientists?", which we will turn to next.

6.9 Conclusion

In this chapter, specific episodes from the case studies of scientific social networking, the use of online tools as Open Laboratory Notebooks, and science blogging have collectively been used to demonstrate the core tensions inherent within the origins and growth of a combined SIM and CM, the open science movement. We have seen the open science movement emerge from the simple action of moving research online, and witnessed the participants' struggle against peer indifference and the inbuilt forces of existing scientific practice, while handling the additional pressure of their

actions not conforming to current notions of academic credit for career advancement.

Hence, this thesis can assert that e-Research in the life sciences, as defined by the use of online tools within scientific research practices, is, within the three-year time span of this research, not revolutionising science but extending the boundaries of scientific work practices and reinforcing existing practices online; and also producing unexpected consequences for the scientists practicing it.

The participants' flexibility of approach extends beyond the current boundaries of many life scientists, advocating a "radical" approach to transparency of thoughts and data production in science. In this sense, the scientists studied in this research are embracing what most scientists perceive to be a risk, although the evidence offered in this thesis shows there is little risk.

For many in the movement, their activities have produced unexpected outcomes. Several have, through actions that felt completely natural to them, become de facto leaders of the movement (Neylon, Bradley). Others have found themselves at the centre of debates around the future forms of peer review (Redfield). For most, however, their participation in these open activities is for them a natural extension of scientific working practices—the way science "should be done". Participating scientists reinforce the academic research profiles of the institutions they work for, via flexible approaches to their communication and collaboration practices described in

this research. They have discovered new ways of working facilitated by their use of online tools, such as ad-hoc small task collaboration.

The participants in this research display a high degree of awareness about the unusual and still rare nature of their activities, which further indicates their membership of the open science movement, as an awareness of the aims of the movement is a core proposition for recruitment of new members and growth of the movement. This is reflected in discussion among themselves and self-aware analysis of their activities as well as engagement with scientific media, shown by the number of articles written about their activities in popular science journals such as Nature and Science, as well as the mainstream media (see Appendix G for a list).

Chapter Seven will conclude this thesis with a discussion of the contributions and limitations of this research, and ideas for further areas of study.

Chapter 7
Conclusions and Implications

Shared and open scientific communications emerged in the early modern era (1550-1700) with the advent of the printing press and the earliest data deluge (Blair, 2003; Eisenstein, 2005), and during the Republic of Letters when scholars formed an intellectual network linking Europe and the USA (1500-1800)[108] and further with the development of the modern journal format by Henry Oldenburg when Secretary of the Royal Society (Willinsky, 2006). All of these developments enabled wider sharing of knowledge within a sphere limited by personal networks and print distribution.

With the advent of networked computers, the Internet and digital publication, knowledge-sharing has become easy: personal publishing is possible with a single click of a mouse, and the distribution of information can be instant and world-wide. Yet the potential for dissemination of knowledge has not yet been fully realised, as the wider issues around openness related to open access publishing, open data and public understanding of science, show. Some disciplines have seized upon the opportunities afforded by the Internet—the use of the online arXiv to publish preprints by various branches of physics is but one example.

---

[108] See "Mapping the Republic of Letters", http://republicofletters.stanford.edu/, a Stanford University project, and associated publications in the journal *Republics of Letters: A Journal for the Study of Knowledge, Politics, and the Arts* (http://arcade.stanford.edu/journals/rofl/).

However, there are also wider issues which demonstrate that open access is still a battle being fought—physicists also still eventually publish their work in peer-reviewed journals, and the journal publication process can contribute to breaking the chain of freely-available scholarship, as bibliographic arXiv links are removed by many publishers (Nielsen, personal communication, 2011). Interpersonal scientific communications have gone through a similar slow development: from private letters and handwritten annotations in the margins of books, to faxes, and now the ubiquity of email in scholarly life. Computers have been central to changing the speed and reach of scholarly scientific communications (Borgman, 2007).

This struggle around openness in its widest sense, that scientific knowledge in its final and polished form should be made available to all, is, as this thesis shows, now also being played out across the full range of activities within the scientific research process, from the crafting of initial hypotheses, to the collection or re-use of data and chronicling its manipulation and management in laboratory notebooks, through to the more studied area of publication and peer review. All of these processes can now be recorded and published online, but few scientists are doing so. This thesis investigates the attempts of a group of scientists to extend the openness of science by opening up the research process to public scrutiny, and laying bare the inner workings of science—which they could not do without the online tools available today.

The open science SIM/CM is influenced by constraints and opportunities that are part of wider e-Research networks and infrastructures. These include existing technical research infrastructures, such as the Internet itself. Other influencing factors discussed within this thesis include copyright, data protection and privacy, commercialisation, academic reward systems and incentives, and formal publishing infrastructures. However, apart from where they interact with the open science SIMs members and actions, these external actors are not an area of focus for this research; for a fuller treatment of these issues see a variety of authors mentioned across this research, including Willinsky (2006), David & Hall, 2006, Wilbanks (2006), Hope (2008), Evans (2010), Dutton & Jeffreys (2010), and Nielsen (2011).

How do the participants in this research fit within this wider push for openness in scholarship? What does it mean to be an open scientist? We have seen, among the participants of this research, movement from the narrowness of traditional roles—experiment and publish—to expansion into several different worlds. A declaration that one is an open scientist means embracing computer technology, learning and discovering new tools, including, as documented in this thesis, wikis, blogs, and social networking tools such as FriendFeed and Twitter. It also means learning a new set of vocabulary around science policy and academic rewards, in order to be able to defend and promote a role as an open scientist. In addition it means becoming a publisher, and engaging with a wider range

of peers than one's immediate discipline or colleagues, and in some cases, engaging with the general public as they comment on science blogs.

With regard to career status and rewards, open scientists must declare themselves to be champions of alternative means of assessing research prowess, and pursue this approach in the face of current policies which distribute the academic rewards of promotion and tenure mainly through numbers of publications in peer-reviewed journals. Open scientists must argue that research outputs beyond publications, such as contributions to data sets, should also be rewarded.

As Willinsky (2006) comments in regard to the editorial gatekeepers of scholarship, digital systems could allow for a more distributed journal culture, particularly for what he terms peripheral scholarly communities in developing countries, for whom access to the scholarly literature is developing but still not equal. This thesis examined the growing culture of post-publication peer review, an activity contentious for a few commenters but seemingly normal for the majority, which also signals a shift in the boundaries of control over what scholars produce.

Scientific collaboration is the act of reaching out to one's peers to develop knowledge by sharing tools and expertise—but generally collaborations are open within themselves and closed to the public, such as the case of the development of the atomic bomb during the Manhattan project (Rhodes, 2005). The participants in this research take collaboration much further, not just reaching out to other scientists but also embracing

interdisciplinary colleagues and the public—anyone can, in theory, help (Bradley's solubility challenge, Todd's invitation to join his projects) and small problems are welcome (FriendFeed and small-task collaboration). The small-task lightweight collaboration which scientists would normally do just by asking the person in the next office or at the laboratory meeting, scientists can now do anytime and anywhere that they have Internet access—and just as importantly, access to a social scientific network. The same applies to communications when they are made lightweight and low-overhead—ease of communicating quickly over distance means that the Internet can augment the locality by providing a means to develop working relationships globally as well as locally. Much as the Health InterNetwork Access to Research Initiative (HINARI) project has made research literature widely available in developing countries free of publisher charges (Willinsky, 2006), this approach to collaboration could facilitate more distributed work, particularly with collaborators in developing countries. For example, Todd, featured in Chapter Four, OLNs, is already starting to do this in his open source antimalarial drug discovery project which will also contribute to India's open source drug discovery programme funded by the Council for Scientific and Industrial Research.

Not just communications, but also the outputs or products of science are changing due to the use of online tools. Fully open data, such as that produced by those using OLNs, has been shown to serve several purposes: correcting and updating the printed mistakes of the past (Bradley and the

production of the melting point database); enabling accurate replication of experiments when required (Maloney and the reproduction of new techniques by others); aiding teaching and learning of new scientific techniques; and producing high-quality scientific results in a shorter time span by facilitating open collaboration (Todd). Variations on existing collaboration and communication patterns have also been identified, such as informal small-task collaboration and networked information filtering. Several of the variations highlighted in Chapter Six are summarised below.

Table 18

Goals and Networked Activities in scientific online tool use.

Goal	Type of Network Activities
Small distinct problem to solve	Scientific social networking
Contact with like-minded	Blogging, scientific social networking
Recording thoughts and ideas about research	Blogging
Discussing research with others	Blogging, scientific social networking
Recording data and research plans	Open Laboratory Notebooks

Scientists instigating these changes through the use of online tools identify themselves—not formally, but through shared advocacy of a particular set of views and adherence to a shared set of practices—as part of an open science movement which this research has identified as a combination of a SIM and a CM. Frickel and Gross's (2005) SIM theory

allows for an accounting of the factors driving intellectual change, whereas Kling and Iacono's (1999) CM theory considers how technologies are introduced and utilised within society to change work practices. Both theories conceptualise collective action as mobilising resources to further the movement, and forms of shared organisational practices that are central to a SIM or CM. To these shared properties, SIMs bring coherence (how the ideas of the SIM are represented to insiders and outsiders), and practices that challenge existing norms and/or dissatisfaction with the scientific status quo, while CMs attempt to push a core technology or set of technologies enabling and driving the movement, and—again—try to overcome opposition (or indifference) to adopting and implementing them. Together, these factors have been shown to be at work in the open science SIM/CM.

The open science SIM/CM is a coherent movement, with a core of active participants, which may appear diffuse because of its informality and only loosely organised nature. However, this research has shown that despite this informality, the open science movement has attracted media and academic attention disproportionate to its small numbers and informal membership. What most scientists would characterise as radical transparency feels like a natural extension of the already-open nature of academic science to those practicing open science. The structural conditions around the use of online tools for open science also assist the movement's growth—the tools they use are free, and do not require the financial

support of their institutions or external funding regimes. However, these positives are accompanied by some opposition.

Opposition to the open science movement is comprised, among other things, of resistance to data-sharing, and an aversion to post-publication peer review and public discussion of the messy practices which lead to the formal publication of scientific work. Consideration of this resistance towards open science has meant that senior scientists practising open science are likely to allow their graduate students and early-career researchers to opt-in to open science rather than making it compulsory. However, as shown in Chapter Six above, this resistance is not coordinated, but most often appears as indifference to open science or non-adoption of open science practices. The issue of the time and effort it takes to change one's practices within a scientific career is also a valid concern, although during this study we have seen that a number of scientists successfully integrate use of online tools into their everyday scientific routine, whether by using them as OLNs or using them to help solve their everyday information management problems.

The open science SIM/CM fits within a larger drive by many other interconnected actors who wish to use Internet tools to make scientific data in a variety of disciplines more available and open (for example, the Open Knowledge Foundation and Science Commons cited in Chapter Six), to devise new methods of data analysis which can involve the public in citizen science projects (such as the Zooniverse group of projects also mentioned in

Chapter Six), and to make published peer-reviewed articles more widely available (for example, the movement promoting open access publishing). These approaches to data sharing and analysis share some of the characteristics identified in the open science SIM/CM—openness, flexibility (using the tool best fit for the task), and fluidity (having little regard for traditional boundaries). However, the scientists participating in this research take this openness and flexibility to its furthest extreme, exposing the entire research process.

7.1 Contributions

This research finds that the combination of SIM and CM theory extends previous research which typically regards new tools as a part of scientific communication or as confined to specialist areas like distributed collaboration, but not in terms of their impact on broader changes in science. As we have seen, it is not possible to divorce the push for openness in science from the tools and practices being used to integrate it into research practices.

In emphasising the relationship between open science practices and open online tools this study makes an overall contribution to our understanding of how science is changing—albeit in a way that is still emerging. In addition the thesis contributes to our understanding of online scientific communication by providing the first in-depth study of the use of open laboratory notebooks, of scientific blogging and of scientific social

networking in the life sciences, including how these online tools are integrated into everyday scientific work practices.

A further contribution is that although Frickel and Gross (2005) provide some brief examples of their General Theory of SIMs, they acknowledge that it has not been tested against a contemporary movement as their theory is founded on an examination of historical movements in science. Investigation of a contemporary well-documented movement like this study contributes to the understanding of SIMs and strengthens Frickel and Gross's theoretical model, though it also extends it (at least for the open science movement examined here) by incorporating the role of online tools for research and how these tools allow scientists to challenge existing practices and promote new ones. With regard to Kling and Iacono's (1994) CM theory, this thesis also strengthens their definition of core technologies, by showing that the products of industrial firms can be validly appropriated by social movements rather than built by social movements, as open scientists have used commercially-oriented products such as FriendFeed and Blogger.com to build their movement. Furthermore, those in the movement are not rigidly aligned to a particular set of technologies—this research has shown that a movement can use technology flexibly, appropriating the technology which shows best fit for their current needs, findings which also extend Kling and Iacono's CM theory.

There is a broader societal CM formed around informal online communications and social network connections, discussed in the literature

reviews for Chapter Three, Online Scientific Networking, and Chapter Five, Science Blogging. Hine (2006) highlighted the connections between broader CMs and the actions of those engaged in the SIM: "these champions do not necessarily develop their ideas about technologies in isolation: they are likely to be local advocates of more widely-held sets of beliefs". Hine's case study examined systematics, and this study adds to this evidence for the life sciences. The combined SIM/CM described in this thesis is specific to these cases, but also resonates with wider societal changes in ICT use, similar to Hine's (2008) description of cyberscience as "a localized manifestation of a dynamic that is not confined to science alone". The critical question here is - are the life scientists *not* involved with this SIM critical of the open science movement? As detailed in Chapter Six, Synthesis, the predominant response is indifference. In contrast, the examples of historical SIMs provided by Frickel and Gross (2005), such as queer theory, women's studies and the rise of ethnomethodology, encounter robust opposition from their peers and sometimes, society. Thus, the broader societal CM of online communications is assisting the development of the open science SIM.

7.2 Limitations of this research

The scope of this research was addressed in Section 1.5, outlining the reasons for choosing particular methods and the embedded cases. It should again be acknowledged that although the case study approach can be used to demonstrate developments within the context of each case, and provide

useful comparison points for others examining additional case studies over time, it does not allow the researcher to generalise further. Further limitations were discovered during the course of the research; and these can now be mentioned by way of suggestions for future research. Research time, discussed above in Research Methods, Chapter Two, means that this research has been a moving target. As Fry and Talja (2007) state: "digital genres analysed are not stable, but emergent and evolving" (p. 130). During the course of this research, Twitter began to be used by many more scientists, and had this project begun in 2010 there would probably have been a need to discuss the rise of Twitter and the way it is used by scientists for connection and discussion, a matter which is only lightly touched on in this research. However, I do not feel that this is a handicap, but rather an example of the changing landscape of scientific communications—many other online tools may rise and fall over the next few years, but the aims of those using them remain similar.

I did not interview or otherwise seek information from scientists sceptical about the impacts of the Internet on scientific work practices (though we have encountered a number of instances of resistance and/or indifference). It is fair to say that very few scientists have not been impacted by the Internet, even if the extent of their interactions is using email and downloading PDF files of journal articles from publisher websites. Most academic scientists do not engage with social media for work purposes, read or write blogs, or keep an online (or even electronic) laboratory

notebook. However, participants in this research observed that the same objections are always raised to their activities (fears of negative repercussions of data-sharing, public peer review, and potential restraints to early-career researchers). These objections and reasons for resistance have been extensively discussed in the Introduction to Chapter Six.

Initially it was hoped that some observation of work practices in the laboratory and other scientific work settings might take place. In practice this proved impractical, as most participants were remotely situated to me. At the same time, one of the advantages of the practices examined in this thesis is that they are all well-documented and in this way accessible, and that the researchers themselves provide thorough accounts of their work (which I have examined, and supplemented with interviews and other materials). The chapter for which observation might have proved most valuable was Chapter Four, Open Laboratory Notebooks. However, this potential limitation was negated by the fact that most of the OLN users record video of experiments within their laboratories, which enabled viewing of the work setting. In addition, within an examination of online work practices, there are virtues in using the same technology and approaches as the research participants, as this offers further insights into their practices, such as the observations made above about the difficulties of time management and fully engaging with these new tools. This thesis has provided a foundation of research in this new area of scientific online tool use, and it is hoped that it will provide a platform others can build on.

7.3 Implications for further research

As mentioned above, a qualitative study of scientists who reject the notion of open science as described in this thesis would provide a fuller understanding of the objections to it, and discover whether there are strong counterexamples to its practice, as the scientists participating in this study seem to override the known objections with ease. A research gap noticed during the course of this research is the lack of prior study of the concept and practices of "scooping" in biomedical research. Although most scientists are convinced it is a real risk, they can only provide second-hand anecdotal evidence of it happening, and there is no comprehensive quantitative or qualitative research data available on the frequency of the practice, or even whether it actually happens. There is a small body of literature on scientific deviance, fraud, misconduct and disputes, and some research on simultaneous discovery in science (Merton, 1961; Kuhn, 1977; Chubin, 1985; Ben-Yehuda, 1986; LaFollette, 1992), but no studies on the impact and incidence of scooping could be found. Also noted during this research is a gap in the current research literature on blogging, namely a study on the time it takes to blog consistently, and how academic bloggers allocate time for blogging—do they consider it part of their research or a personal pursuit?

Further studies could also investigate the role of other Web tools in the life sciences; for instance, in some scientific communities Facebook is used to communicate with the public, and some of the other purpose-built applications for scientists such as *ResearchGate* and *Colwiz*, and others

mentioned in Section 3.8.4, could be investigated to see if the patterns found in this research exist elsewhere and could potentially join up to add to a more powerful open science movement. Other work that could inform the findings of this thesis would include analysis of other views of impact, such as within the academic publishing sphere. Webmetric and bibliometric study of citation patterns within the open science community could contribute to this analysis.

Overall, then, the participants in this research are enacting a version of open science facilitated by technology, using freely available online tools adapted for performing e-Research. They are promoting a set of tools which turn the end-to-end research process—from hypothesis to experiment to publication—into an open and public process. They welcome other scholars to join them, albeit with the caveats noted in this thesis, as new recruits must be willing to engage with this technology, and be flexible about it, as the next best tool may launch or be found tomorrow. This activity fits within the larger open science and open access movement as the first to embrace an open end-to-end approach to scientific research.

The success of the open science movement described in this thesis is not guaranteed, but as documented here, this movement consists both of ideals and practices, interwoven in a belief that better science and better technology go hand in hand. The present research adds to our understanding of the impact of computerisation on scientific work, particularly with regard to appropriation of existing technologies to shift

the boundaries of long-established scientific practices by extending open science to all stages of the research process.

When ideas about openness and open online tools are combined, they support the two theories operationalised in this thesis: The General Theory of Scientific/Intellectual Movements, which focuses on new ideas in science, and the theory of Computerisation Movements, which focuses on the use of new technologies. When tied together with the seven research aspects examined in this thesis and in previous research of change in scientific roles, the gatekeepers of science, scientific communication and collaboration, career status and rewards, openness of scientific data, and scientific interaction with the media and public, it is possible to account for contemporary changes in the life sciences.

This thesis provides a balanced analysis that does not uncritically accept the promise of a technological utopia—the present research has found that the use of online tools is mediated by several forces such as opposition to traditional ways of performing scientific work, the challenges of time management and learning new technologies. However, it it is also undeniable that the use of online tools is changing scientific ways of working—as this research shows, the tools described here have proved essential to a group of open scientists who may well be the harbingers of a disciplinary or even wider change in the way academic work is done. The movement towards opening up the end-to-end research cycle in the life sciences could not happen without the online tools that facilitate

publishing, sharing and discussion—and the features of those online tools that allow public sharing.

Appendix A

List of Research Questions

Is the use of online tools revolutionising, reinforcing or having unintended consequences for the research practices of academic life scientists?

Research Aspect	Research Question
Roles of scientific researchers	Do online tools extend the professional roles of academic life scientists beyond their traditional ones?
Scientific collaboration practices	Do online tools facilitate inter-institutional and interdisciplinary collaboration for academic life scientists, and if so, how?
Openness of scientific data and discourse	Where in the chain of collecting and disseminating research materials are the incentives and barriers to open practices in the use of online tools by academic life scientists?
Informal and formal scientific communication practices	How do academic life scientists use online tools (social networks, open laboratory notebooks, blogs) to disseminate different types of scientific data?
The roles of gatekeeper institutions	Does the use of online tools by academic life scientists circumvent gatekeepers in science and/or create new ones?
Scientific career status management and rewards	Do the various forms of online dissemination being used by academic life scientists yield rewards and status comparable to traditional outputs, or are there new mechanisms developing for reward and recognition of these activities?
Scientific engagement with the media and public	How do academic life scientists see that their engagement with the public has changed since they adopted online tools?

Appendix B

Invitation Emails

Samples of a invitations to interview. Each email was tailored to suit the potential research participant according to their activities and interactions with others in the SIM.

Dear Professor [name],

I am a fourth year DPhil student at the Oxford Internet Institute, University of Oxford, where my dissertation project is looking at e-Research in the Life Sciences. I've spoken with a number of open scientists, including Cameron Neylon and Jean-Claude Bradley, as well as other science bloggers, several of whom recommended I talk to you.

I am interested in speaking with you specifically about science blogging and open science. If you have some time next week, I would appreciate an hour (or less) of it to conduct an interview with you, by phone or Skype. Any day, any time before 2pm [city] time would be great.
Kind regards,
Lucy

Lucy Power
University of Oxford - Oxford Internet Institute - DPhil Candidate
lucy.power@oii.ox.ac.uk
http://www.oii.ox.ac.uk
http://people.oii.ox.ac.uk/power/

See next page for another email.

Dear [name],

Last week I spoke with Cameron Neylon who mentioned your name in connection with the use of FriendFeed by scientists.

I am a second year DPhil student at the Oxford Internet Institute, University of Oxford, where my dissertation project is looking at e-Research in the Life Sciences, with a specific focus on online tools and their use in science communications and research dissemination.

I would be very interested in speaking with you about FriendFeed, open science and Web-based research and communication tools. If you have some time before June 21st I would appreciate an hour (or less) of it to conduct an interview with you, by phone or in person (if geography permits).

Kind regards,
Lucy

Lucy Power
University of Oxford - Oxford Internet Institute - DPhil Student
lucy.power@oii.ox.ac.uk
http://www.oii.ox.ac.uk
http://people.oii.ox.ac.uk/power/

Appendix C

Schedule of Interviews

n = 22.

Name	Date Interviewed DD/MM/YY	Face-to-Face (F2F) or Skype	M/F	Location
Nick Anthis	26/10/09	F2F	M	U.K.
Pedro Beltrao	18/02/10	Skype	M	USA
Jean-Claude Bradley	10/06/10	Skype	M	USA
David Bulger	06/02/11	Email	M	USA
Richard Grant	17/06/09	F2F	M	U.K.
Bill Hooker	12/06/09	Skype	M	USA
Lars Juhl Jensen	12/03/10	Skype	M	Denmark
Martin Johnson	21/02/11	Skype	M	U.K.
Steve Koch	12/08/10	Skype	M	USA
Andrew Lang	03/02/11	Skype	M	USA
Pierre Lindenbaum	04/08/10	Email	M	France
Andy Maloney	07/09/10	Skype	M	USA
Berci Mesko	15/03/10	Skype	M	Hungary
Peter Murray-Rust	07/12/09	F2F	M	U.K.
Cameron Neylon	02/06/09	F2F	M	U.K.
Rosie Redfield	25/07/11	Skype	F	Canada
Anthony Salvagno	02/09/10	Skype	M	USA
Neil Saunders	29/10/09	Skype	M	Australia
Deepak Singh	12/11/09	Skype	M	USA
Pawel Szczęsny	03/03/10	Skype	M	USA
Mat Todd	28/04/11	Skype	M	Australia
Bora Zivkovic	02/09/10	Skype	M	USA

Appendix D

Consent Form



e-Research in the Life Sciences
Interviewee information sheet

Researcher
Lucy Power
Oxford Internet Institute, University of Oxford
1 St Giles, Oxford OX1 3JS
lucy.power@oii.ox.ac.uk
+44 (0)1865 287 210

Background to the research
The Internet has brought new ways of working to most professions. This research seeks to quantify and analyse the extent of these new ways of working and their effects both positive and negative on research processes and outputs, publication, networking and knowledge dissemination.

Why is this research being conducted?
The aim of the research is to better understand the role of the Internet in professional experience and the effect of networking and knowledge dissemination as a result of the Internet. The research is timely in that Internet use among scientists is known to be extensive but little is known of the detail of how scientists use the Internet in their work processes. This study will provide useful information and analysis to enable future developments to be planned around scientists' needs and desired working patterns.

What you will be asked to do should you agree to participate.
You will take part in an interview in which you will be asked questions regarding Internet use and networking activities among the scientific community. It will take about 60 minutes to complete.

The basis of your selection for the research.
The project involves interviewing about 40 scientists with a range of experience at different stages of their career with experience in the life sciences.

Your confidentiality.
With your agreement the interview will be recorded and later transcribed. The recording will then be destroyed and the transcriptions kept on a private secure server at the Oxford Internet Institute. You may choose whether your interview is anonymised or not at the conclusion of the interview. Your participation is voluntary: you do not have to answer every question. You have the right to withdraw from the study at any time.

Questions and feedback.
Further information concerning the project may be obtained by contacting the principal researcher, Lucy Power. Feedback of the results of this study will be made available to you on request. The study is being supervised by Dr. Ralph Schroeder and Dr. Eric Meyer of the OII, and has received ethical approval from the Social Sciences and Humanities Inter-divisional Ethics Committee of the University of Oxford. There are no known risks to participating in this study.

By signing below I confirm that I have read and understood the information package and in particular, have noted that:

- I understand that my involvement in this research entails taking part in an interview, related to researchers' use of the Internet
- I have had any questions answered to my satisfaction
- I understand the risks involved
- I understand there will be no direct and immediate benefit to me from participation in this research.
- I understand that my participation is voluntary, and that responses are completely anonymous and that no identifying information is required
- I understand that if I have any additional questions or concerns, I can contact the researcher at the address provided
- I understand that I am free to withdraw at any time without comment or penalty
- I agree to participate in the project

Name_____ Age: _____

Signature_____ Date _____

I would like this interview to be anonymised: ☐ Yes ☐ No

Further information

Employer: _____ Job Title: _____

Education:

☐ PhD ☐ Masters ☐ Bachelor's degree ☐ Current student (degree): _____

Field of research: _____

PLEASE RETAIN THIS PAGE FOR LATER REFERENCE

Appendix E
Interview Questions

I'd like to talk about a few things with you today – the use of technology in your daily routine, and where ONS, blogging, FF and other online activities work within that. Also talk in a bit more detail about how you started blogging, and what it means to you.
Lastly I'll ask more about FriendFeed, as I'm interested in how that is being used by scientists.

Background
Can you describe the path that led to your current job?

Daily routine
Can you describe your average daily work routine?
How is your routine now different from when you didn't use online tools?

Computer use
Could you tell me how things have changed with your work and computers between your student years and now?

Open notebooks
Can you talk me through how you got started with your open notebook?
What processes and people were involved in setting it up?
What reactions do you get when you talk to others about it:
for example, your colleagues, students, the public?
Are they mostly positive or negative reactions? Can you describe them?
How do you use the data you produce and store on your notebook?
Is anyone else using your data that you're aware of?
Why would they? Why wouldn't they?
Does your online notebook fulfil the same role that the paper notebook did?

What have been the best and worst things about the open notebook?
What makes [worst thing] so difficult? What makes [best thing] so enjoyable?

Blogging

Can you talk me through how you got started with blogging – [cite date if known]
How has your blog changed over time?
Can you step me through your process of writing a blog post, from initial inspiration to writing?
What have been the best and worst things about blogging?
What makes [worst thing] so difficult? What makes [best thing] enjoyable?
What usually happens after you publish a post?
What are the range of reactions you get to posts?
Have reactions to any posts changed the way you do anything?
What do you know about your audience?
How are your activities affecting your field?
What impact do you think it is having?

FriendFeed users

How did you get started using FriendFeed?
Can you give me some examples of how you use it?
What have you seen others using it for? What do you think of that?
Do you think the type of connections you make through FriendFeed, would have been made via email? Would you have run across these people any other way?

Research aspect-specific

What does the term "open science" mean to you?
In publishing? Public understanding of science? Collaboration?
How would you describe "Web 2.0" to someone else?

How do the tools you use fall into this category?
What specific contributions do you think Internet tools make to your research?
What are you doing with these tools?
Do you have ideas for improvements?
If not now, then what are you hoping for in the future?
Tell me a bit about how and where you seek to publish your work.

Conferences

Tell me which conferences you attend in a year, which are essential to you, and what you take away from them.
Can you give me some examples of the types of data you present at conferences?
Is the data you present usually already published or not?
Papers or posters mostly?

Collaboration

How do you find collaborators for projects? Has anything changed about that?

Closing interview

We haven't really talked much about [research aspect name]. Do you have anything you'd like to say about [research aspect name]?
Roles, Gatekeepers, Rewards & Status, Open Science, Collaboration, Communications, Public Understanding of Science
You're an early adopter. Where would you like to see this all end up? Do you have a vision for this in the future?
Can you think of anyone else I should talk to?

Appendix F

Content Codes

Listed below are the coding categories used to code the interview transcripts and other codeable data such as blog and FriendFeed posts. Use of the computer-assisted qualitative analysis software is discussed in the Methods chapter.

background: participants' background and history

case>blog: Science Blogging

case>FF: FriendFeed

case>OLN: Open Laboratory Notebook

computer_use: history with computers, expertise

conference: conferences attended

library: mentions of going/not going to libraries

online_tool: tools mentioned that were not any of the three cases, SecondLife or Twitter

online_tool>secondlife: SecondLife (SL) use

online_tool>Twitter: Twitter use

people: [details below]

people>CameronNeylon: mentions of Neylon

people>JCBradley: mentions of Bradley

people>SteveKoch: mentions of Koch

question: Question asked by interviewer

quote: Something that might make a good quote

role: context code for role

routine: daily routine

secrecy: talk about scooping, secrecy or lack thereof

synthesis: highlighting similarities and overlap - useful for synthesis chapter

theme>collaboration: Do academic life scientists using online tools engage in more inter-institutional and more interdisciplinary collaboration than their colleagues?

theme>communication: Do academic life scientists use online tools (blogs, scientific social networking, open laboratory notebooks) to publish more data and more types of data than they did before using these tools?

theme>gatekeepers: Does the use of online tools by academic life scientists circumvent gatekeepers in science and/or create new ones?

theme>openscience: Where in the chain of collecting and disseminating research materials are the incentives and barriers to open practices in the use of online tools by academic life scientists?

theme>public: Do academic life scientists using online tools engage with the public more than other life scientists?

theme>roles: Are academic life scientists using online tools to extend their professional roles beyond their traditional ones?

theme>status_rewards: Do the various forms of online dissemination being used by academic life scientists yield rewards and status comparable to traditional outputs, or are there new mechanisms developing for reward and recognition of these activities?

theory>Computerisation: evidence of Computerisation Movement (CM)

theory>SIM: SIMs, in Frickel and Gross's account, emerge through marshalling resources (for example, controlling publication channels, securing positions within professional bodies, consolidating groups in disciplinary specialisms) behind an agenda that challenges established ideas or scientific orthodoxies.

Appendix G

List of Articles in Popular Science Publications

Butler, D. (2005). Electronic notebooks: A new leaf. *Nature*, 436(7047), 20-21. doi:10.1038/436020a

Economist (2004, Jun 10). An open-source shot in the arm? *The Economist*. Retrieved from http://www.economist.com/node/2724420

Goetz, T. (2007). Freeing the Dark Data of Failed Scientific Experiments. *Wired*,. Retrieved from http://www.wired.com/science/discoveries/magazine/15-10/st_essay

Griffiths, M. (2007). Talking physics in the social web. *Physics World*, 20(1), 24-28.

Neylon, C. (2009). Stitching science together. *Nature*, 461, 881.

Nielsen, M. (2009). Doing Science in the Open. *Physics World*,. Retrieved from http://physicsworld.com/cws/article/print/38904

Sanderson, K. (2008). Data on display. *Nature*, September. doi:10.1038/455273a

Shneiderman, B. (2008). Science 2.0. *Science*, 319(5868), 1349-1350. doi:10.1126/science.1153539

Woelfle, M., Olliaro, P., & Todd, M. H. (2011). Open science is a research accelerator. *Nat Chem*, *3*(10), 745-748. doi:10.1038/nchem.1149

Velden, T., & Lagoze, C. (2009). Communicating chemistry. *Nature Chemistry*, 1(9), 673-678. doi:10.1038/nchem.448

Waldrop, M. M. (2008). Science 2.0: Great New Tool, or Great Risk? *Scientific American*, 9 Jan. Retrieved from http://www.sciam.com/article.cfm?id=science-2-point-0-great-new-tool-or-great-risk

References

3.14159rate. (2010). *Arsenic-associated bacteria (NASA's claims).* [Web log comment] Retrieved January 12, 2011 from http://rrresearch.blogspot.com/2010/12/arsenic-associated-bacteria-nasas.html

Adams, G., & White, J. (1994). Dissertation Research in Public Administration and Cognate Fields: An Assessment of Methods and Quality Author(s). *Public Administration Review*, *54*(6), 565-576.

Arms, W. Y., & Larsen, R. L. (2007, 17-19 April). *The future of scholarly communication: building the infrastructure for cyberscholarship*. Proceedings from a workshop sponsored by the National Science Foundation (NSF) and the Joint Information Systems Committee (JISC)., Phoenix, Arizona.

Bailey, M. T. (1992). Do Physicists Use Case Studies? Thoughts on Public Administration. *Public Administration Review*, *52*(1), 47-54.

Bar-Ilan, J. (2005). Information hub blogs. *Journal of Information Science*, *31*(4), 297-307.

Bar-Ilan, J. (2008). Informetrics at the beginning of the 21st century—A review. *Journal of Informetrics*, *2*(1), 1-52.

Barabasi, A. L. (2003). *Linked: How Everything Is Connected to Everything Else and What It Means.* New York: Plume.

Barjak, F. (2006). The role of the internet in informal scholarly communication. *Journal of the American Society for Information Science and Technology*, *57*(10), 1350-1367.

Barjak, F., & Thelwall, M. (2008). A statistical analysis of the web presences of European life sciences research teams. *Journal of the American Society for Information Science and Technology*, *59*(4), 628-643.

Batts, S. A., Anthis, N. J., & Smith, T. C. (2008). Advancing science through conversations: bridging the gap between blogs and the academy. *PLoS Biology*, *6*(9), e240. doi:10.1371/journal.pbio.0060240

Bazerman, C. (1988). *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science.* Madison, Wisconsin: University of Wisconsin Press.

Becher, T. (1989). *Academic tribes and territories : intellectual enquiry and the cultures of disciplines.* Milton Keynes, England ; Bristol, PA., USA: Society for Research into Higher Education : Open University Press.

Benford, P., & Standen, P. J. (2010). The use of email-facilitated interviewing with higher functioning autistic people participating in a grounded theory study. *International Journal of Social Research Methodology*. doi:10.1080/13645579.2010.534654

Ben-Yehuda, N. (1986). Deviance in Science. *British Journal of Criminology*, *26*(1), 1-27.

Berners-Lee, T., Hall, W., Hendler, J., A, O'Hara, K., Shadbolt, N., & Weitzner, D., J. (2006). A Framework for Web Science., Retrieved from: http://eprints.ecs.soton.ac.uk/13347/.

Bettencourt, L. M. A., Kaiser, D. I., Kaur, J., Castillo-Chávez, C., & Wojick, D. E. (2008). Population modeling of the emergence and development of scientific fields. *Scientometrics*, *75*(3), 495-518.

Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology*, *57*(13), 1758-1770. doi:10.1002/asi.20380

Bishop, J. (2007). Increasing participation in online communities: A framework for human-computer interaction. *Computers in Human Behavior*, *23*(4), 1881-1893. doi:10.1016/j.chb.2005.11.004

Blair, A. (2003). Reading Strategies for Coping with Information Overload ca. 1550-1700. *Journal of the History of Ideas*, *64*(1), 11-28.

Blake, J. A., & Bult, C. J. (2006). Beyond the data deluge: data integration and bio-ontologies. *Journal of Biomedical Informatics*, *39*(3), 314-320.

Blood, R. (2002). Introduction. In J. Rodzvilla (Ed.), *We've Got Blog: How Weblogs Are Changing Our Culture* (pp. ix-xiii). Cambridge, MA.: Perseus Publishing.

Blumer, H. (1954). What is wrong with social theory? *American sociological review*, *19*(1), 3-10.

Bonetta, L. (2007). Scientists Enter the Blogosphere. *Cell*, *129*(3), 443-445.

Borgman, C. L. (2007). *Scholarship in the digital age: Information, infrastructure, and the internet.* Cambridge, MA.: The MIT Press.

boyd, d., & Ellison, N. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer Mediated Communication*, *13*(1), 210-230.

Boyle, R. (2001). *The Correspondence of Robert Boyle, 1636-91* (1), ed. Hunter, M. C. W., Clericuzio, A., & Principe, L. London: Pickering & Chatto (Publishers) Ltd.

Bradley, J-C. (2006). *Open Notebook Science*. Retrieved 15 August, 2011 from http://drexel-coas-elearning.blogspot.com/2006/09/open-notebook-science.html

Bradley, J-C. (2007). *Open Notebook Science Using Blogs and Wikis*. Retrieved from http://dx.doi.org/10.1038/npre.2007.39.1

Bradley, J-C. (2011). *More on 4-benzyltoluene and the impact of melting point data curation and transparency*. Retrieved 25 August, 2011 from http://usefulchem.blogspot.com/2011/06/more-on-4-benzyltoluene-and-impact-of.html

Bradley, J-C., & Samuel, B. (2000). SMIRP-A Systems Approach to Laboratory Automation. *Journal of the Association for Laboratory Automation*, *5*(3), 48-53. doi:10.1016/S1535-5535(04)00074-7

Brandtzaeg, P. B., & Heim, J. (2009). Why people use social networking sites. *Lecture Notes in Computer Science*, *5621 LNCS*, 143-152.

Brembs, B., Mietchen, D., Neylon, C., Pikas, C. K., Steel, G., & Todd, M. H. (2009). *Title: A Friend in Feed is a Co-author Indeed: The social aggregator as a tool for user-led collaborative science*. Retrieved 5 May, 2010 from http://ff4s-paper.wikidot.com/

Burns, E. (2010). Developing Email Interview Practices in Qualitative Research. *Sociological Research Online*, *15*(4). doi:10.5153/sro.2232

Butler, D. (2005). Electronic notebooks: A new leaf. *Nature*, *436*(7047), 20-21. doi:10.1038/436020a

Caldas, A. (2003). Are newsgroups extending "invisible colleges" into the digital infrastructure of science? *Economics of Innovation and New Technology*, *12*(1), 43-60.

Campbell, E. G., Clarridge, B. R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N. A., & Blumenthal, D. (2002). Data withholding in academic genetics: evidence from a national survey. *JAMA*, *287*(4), 473.

Carter, S. (2005, April 2-7). *The role of the author in topical blogs*. Proceedings from CHI-05, Portland, OR, USA.

CBC News (2010, December 6). NASA's arsenic microbe science slammed. *CBC News*. Retrieved from http://www.cbc.ca/news/technology/story/2010/12/06/arsenic-microbe-dna-nasa-wolfe-simon.html

Celli, F., Di Lascio, F., Magnani, M., Pacelli, B., & Rossi, L. (2010). Social Network Data and Practices: The Case of Friendfeed. *Advances in Social Computing,* 346-353. doi:10.1007/978-3-642-12079-4_43

Charmaz, K. (2003). Grounded theory: Objectivist and constructivist methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies for qualitative inquiry* (2nd ed., pp. 249-291). Thousand Oaks, CA: Sage.

Chavis, D. M., & Pretty, G. M. H. (1999). Sense of community: Advances in measurement and application. *Journal of Community Psychology*, *27*(6), 635-642.

Chin, A., & Chignell, M. (2006, August 22-25). *A social hypertext model for finding community in blogs*. Proceedings from the seventeenth conference on hypertext and hypermedia, Odense, Denmark.

Chubin, D. E. (1985). Misconduct in research: An issue of science policy and practice. *Minerva*, 23(2), 175-202.

Collins, H. M. (1974). The TEA Set: Tacit Knowledge and Scientific Networks. *Social Studies of Science*, *4*, 165-185.

Collins, R. (1994). Why the social sciences won't become high-consensus, rapid-discovery science. *Sociological Forum*, *9(2)*, 155-177.

Comai, L., & Cartwright, R. A. (2005). A Toxic Mutator and Selection Alternative to the Non-Mendelian RNA Cache Hypothesis for *hothead* Reversion. *Plant Cell*, *18*, 2856-2858.

Constant, D., Sproull, L., & Kiesler, S. (1996). The kindness of strangers: The usefulness of electronic weak ties for technical advice. *Organizational Science*, *7*(2), 119-135.

Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology*, *52*(7), 558-569.

Cummings, J. N., & Kiesler, S. (2005). Collaborative research across disciplinary and organizational boundaries. *Social studies of science*, *35*(5), 703-722.

David, P. A. (2004). Towards a cyberinfrastructure for enhanced scientific collaboration. *OII research report*, *4*.

David, P. A., den Besten, M., & Schroeder, R. (2005, December 7-10). *How open is e-Scienc*e? Proceedings from the Second IEEE International Conference on e-Science and Grid Computing, Brisbane, Australia.

David, P. A., & Hall, B. H. (2006). Property and the pursuit of knowledge: IPR issues affecting scientific research. *Research Policy*, *35*(6), 767-771.

David, P., A. (2004). Understanding the emergence of 'open science' institutions: functionalist economics in historical context. *Ind Corp Change*, *13*(4), 571-589. doi:10.1093/icc/dth023

de Moor, A., & Efimova, L. (2004, 2-3 June). *An argumentation analysis of weblog conversations*. Proceedings from 9th International Working Conference on the Language-Action Perspective on Communication Modelling (LAP 2004), New Brunswick, NJ,.

de Solla Price, D. J. (1961). *Science Since Babylon.* New Haven, USA: Yale University Press.

den Besten, M. (2003). *The Rise of Bioinformatics An historical approach to the emergence of a new scientific discipline.* M. Phil. University of Oxford, Oxford.

Dennen, V. P. (2009). Constructing academic alter-egos: identity issues in a blog-based community. *Identity in the Information Society, 2(1).* 1-16. doi:10.1007/s12394-009-0020-8

Dougherty, M., Meyer, E., Madsen, C., Van den Heuvel, C., Thomas, A. and Wyatt, S. (2010) 'Researcher Engagement with Web Archives: State of the Art', Joint Information Systems Committee Report. http://ssrn.com/abstract=1714997

Drugmonkey. (2011). *Pseudonymous blogging at Science Blogs is over*. Retrieved from http://scienceblogs.com/drugmonkey/2011/08/pseudonymous_blogging_at_scien.php

Duque, R. B., Ynalvez, M., Sooryamoorthy, R., Mbatia, P., Dzorgbo, D. B., & Shrum, W. (2005). Collaboration Paradox: Scientific Productivity, the Internet, and Problems of Research in Developing Areas. *Social Studies of Science*, *35*, 755-785.

Dutton, W.H. (1999). *Society on the Line*. Oxford: Oxford University Press.

Dutton, W.H. (2005). The Internet and Social Transformation: Reconfiguring Access. In Dutton, W., Kahin, B., O'Callaghan, R. and Wykoff, A. (Eds.), *Transforming Enterprise: Economic and Social Implications of Information Technology* (pp. 375-394). Cambridge, MA: MIT Press.

Dutton, W. H. (2008). The Wisdom of Collaborative Network Organizations: Capturing the Value of Networked Individuals. *Prometheus*, *26*(3), 211-230.

Dutton, W. H., Carusi, A., & Peltu, M. (2006). Fostering Multidisciplinary Engagement: Communication Challenges for Social Research on Emerging Digital Technologies. *Prometheus*, *24*(2), 129-149.

Dutton, W.H., Helsper, E.J., and Gerber, M.M. (2009) The Internet in Britain: 2009. *Oxford Internet Institute*, University of Oxford.

Dutton, W. H., & Jeffreys, P. (2010). *World Wide Research: Reshaping the Sciences and Humanities.* Cambridge, MA: The MIT Press.

Efimova, L. (2009, June 29 - July 1). *Weblog as a personal thinking space*. Proceedings from 20th ACM conference on Hypertext and hypermedia, Torino, Italy. doi:10.1145/1557914.1557963

Eisenstein, E. (2005). *The Printing Revolution in Early Modern Europe.* Cambridge: Cambridge University Press.

Elliott, M. S., & Kraemer, K. L. (2008). *Computerization Movements and Technology Diffusion.* Medford, NJ: Information Today, Inc.

Evans, J. A. (2008). Electronic Publication and the Narrowing of Science and Scholarship. *Science*, *321*(5887), 395-399.

Evans, J. A. (2010). Industry Induces Academic Science to Know Less about More. *American Journal of Sociology*, *116*(2), 389-452.

Evans, J. A. (2004). *Sharing the harvest? The uncertain fruits of public/private collaboration in plant biotechnology.* Ph.D. Stanford University.

Ewins, R. (2005). Who Are You? Weblogs and Academic Identity. *E-Learning*, *2*(4), 368-377.

Fisher, A. T., Sonn, C. C., & Bishop, B. J. (2002). *Psychological sense of community: Research, applications, and implications*. Springer US.

Fornaro, J. (2010). *Arsenic-associated bacteria (NASA's claims)*. [Web log comment] Retrieved January 12, 2011 from http://rrresearch.blogspot.com/2010/12/arsenic-associated-bacteria-nasas.html

Fortson, L., Masters, K., Nichol, R. C., Borne, K., Edmondson, E., Lintott, C. J., … Wallin, J. (2011). *Galaxy Zoo: Morphological classification and citizen science*. arXiv:1104.5513 [astro-ph]

Fox, M. F. (2001). Women, science and academia: Graduate education and careers. *Gender and Society*, *15*(5), 654-666.

Frickel, S., & Gross, N. (2005). A general theory of scientific/intellectual movements. *American Sociological Review*, *70*(2), 204-232.

Fry, J., & Talja, S. (2007). The intellectual and social organization of academic fields and the shaping of digital resources. *Journal of Information Science*, *33*(2), 115-133.

Fuchs, S. (1992). *The professional quest for truth: A social theory of science and knowledge.* Albany: State University of New York Press.

Fuchs, S. (1993). A sociological theory of scientific change. *Social Forces*, *71*, 933-954.

Fuchs, S. (2002). What makes sciences scientific? In J. Turner (Ed.), *Handbook of Sociological Theory* (pp. 21-35). New York: Kluwer Academic/Plenum Publishers.

Gabler, J., & Frank, D., John. (2005). The Natural Sciences in the University: Change and Variation over the 20th Century. *Sociology of Education*, *78*(3), 183-206.

Galperin, M., Y. (2008). The molecular biology database collection: 2008 Update. *Nucleic Acids Research*, *36*(suppl_1).

Garg, S., Gupta, T., Carlsson, N., & Mahanti, A. (2009). *Evolution of an online social aggregation network: an empirical study*. Proceedings from 9th ACM SIGCOMM conference on Internet measurement, New York, NY. doi:10.1145/1644893.1644931

Glaser, B., & Strauss, A. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Chicago: Aldine Publishing Company.

Global Voices Online. (2010). *About*. Retrieved 15 August, 2011 from http://globalvoicesonline.org/about/

Goetz, T. (2007). Freeing the Dark Data of Failed Scientific Experiments. *Wired*, 15 Oct. Retrieved from http://www.wired.com/science/discoveries/magazine/15-10/st_essay

Goldstein, A. M. (2009). Blogging Evolution. *Evolution Education Outreach*, *2*, 548-559.

Gooding, D. (1990). Mapping Experiment as a Learning Process: How the First Electromagnetic Motor Was Invented. *Science, Technology & Human Values*, *15*(2), 165-201. doi:10.1177/016224399001500202

Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology*, *78*(6), 1360-1380. doi:10.1086/225469

Granovetter, M. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, *1*, 201-233.

Gregg, M. (2009). Banal bohemia: Blogging from the ivory tower hot-desk. *Convergence*, *15*(4), 470-483.

Gregg, M. (2006). Feeling Ordinary: Blogging as Conversational Scholarship. *Continuum*, *20*(2), 147-160.

Griffith, B. C., & Mullins, N. C. (1972). Coherent Social Groups in Scientific Change. *Science*, *177*(4053), 959-964.

Griffiths, M. (2007). Talking physics in the social web. *Physics World*, *20*(1), 24-28.

Groth, P., & Gurney, T. (2010, April 26-27th). *Studying Scientific Discourse on the Web using Bibliometrics: A Chemistry Blogging Case Study*. Proceedings from WebSci10: Extending the Frontiers of Society On-Line, Raleigh, NC.

Gruhl, D., Liben-Nowell, D., Guha, R., & Tomkins, A. (2004). Information diffusion through blogspace. *SIGKDD Explor. Newsl.*, *6*(2), 43-52. doi:10.1145/1046456.1046462

Grushcow, J., M. (2004). Measuring Secrecy: A Cost of the Patent System Revealed. *The Journal of Legal Studies The Journal of Legal Studies*, *33*(1), 59-84. doi:10.1086/381288

Gupta, T., Garg, S., Mahanti, A., Carlsson, N., & Arlitt, M. (2009, May 17-20). *Characterization of FriendFeed – A Web-based Social Aggregation Service*. Proceedings from Third International AAAI Conference on Weblogs and Social Media, San Jose, CA.

Hackett, E. J., Amsterdamska, O., Lynch, M., & Wajcman, J. (2007). *The handbook of science and technology studies, third edition.* Cambridge, MA.: The MIT Press.

Hagstrom, W. O. (1974). Competition in Science. *American Sociological Review*, *39*(1), 1-18.

Halavais, A. (2006). Scholarly Blogging: Moving towards the Visible College. In A. Bruns & J. Jacobs (Eds.), *Uses of Blogs* (pp. 117-126). New York: Peter Lang.

Hara, N., & Rosenbaum, H. (2008). Revising the conceptualization of computerization movements. *Information Society*, *24*(4), 229-245.

Harley, D., Acord, S. K., Earl-Novell, S., Lawrence, S., & King, C. J. (2010). *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines*. UC Berkeley: Center for Studies in Higher Education. Retrieved from: http://escholarship.org/uc/item/15x7385g.

Herring, S., Scheidt, L. A., Kouper, I., & Wright, E. (2006). Longitudinal content analysis of blogs: 2003 - 2004. In M. Tremayne (Ed.), *Blogging, Citizenship and the Future of Media.* New York: Routledge.

Herring, S. C., & Paolillo, J. C. (2006). Gender and genre variation in weblogs. *Journal of Sociolinguistics*, *10*(4), 439–459. doi:10.1111/j.1467-9841.2006.00287.x

Hesse, B., Sproull, L., Kiesler, S., & Walsh, J. P. (1993). Returns to Science: Computer Networks in Oceanography. *Communications of the ACM*, *36*(8).

Hey, T., & Trefethen, A. (2003a). e-Science and its implications. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *361*, 1809-1825.

Hey, T., & Trefethen, A. (2003b). The data deluge: An e-science perspective. In *Grid computing: making the global infrastructure a reality* (pp. 809-824). John Wiley & Sons, Ltd. doi:10.1002/0470867167.ch36

Hindman, M. (2009). *The Myth of Digital Democracy.* Princeton, N.J.: Princeton University Press.

Hine, C. (2007). Connective Ethnography for the Exploration of e-Science. *Journal of Computer-Mediated Communication, (12),* 618-634. doi:10.1111/j.1083-6101.2007.00341.

Hine, C. (2008). *Systematics as Cyberscience: Computers, Change, and Continuity in Science (Inside Technology).* Cambridge, MA.: The MIT Press.

Holmes, F., L., Renn, J., & Rheinberger, H.-J. (2003). *Reworking the Bench: Research notebooks in the history of science* (7). Dordrecht: Kluwer Academic Publishers.

Hong, W., & Walsh, J., P. (2009). For money or glory? Commercialization, competition and secrecy in the entrepreneurial university. *Sociological Quarterly*, *50*(1), 145-171. doi:10.1111/j.1533-8525.2008.01136.x

Hookway, N. (2008). `entering the blogosphere': Some strategies for using blogs in social research. *Qualitative Research*, *8*(1), 91-113.

Hope, J. (2008). *Biobazaar: The Open Source Revolution and Biotechnology.* Cambridge, MA.: Harvard University Press.

Hsu, C.-L., & Lin, J. C.-C. (2008). Acceptance of blog usage: The roles of technology acceptance, social influence and knowledge sharing motivation. *Information & Management*, *45*(1), 65-74. doi:10.1016/j.im.2007.11.001

Hurd, G. (2010). *Arsenic-associated bacteria (NASA's claims).* [Web log comment] Retrieved January 12, 2011 from http://rrresearch.blogspot.com/2010/12/arsenic-associated-bacteria-nasas.html

inducedPluripotency. (2010). *Arsenic-associated bacteria (NASA's claims).* [Web log comment] Retrieved January 12, 2011 from http://rrresearch.blogspot.com/2010/12/arsenic-associated-bacteria-nasas.html

Jankowski, N. (Ed.). (2009). *E-Research: Transformation in Scholarly Practice*. New York: Routledge.

Johnson, M. (2011). *I declare my science offically open! Mostly*. Retrieved 15 August, 2011 from http://martwine.wordpress.com/2011/02/03/i-declare-my-science-offically-open-mostly/

Kaplan, K. (2010). Academia: The changing face of tenure. *Nature*, *468*, 123–125. doi:10.1038/nj7320-123a

Karimi, F., & Poo, D. C. C. (2009, November 6-11). *Personal and External Determinants of Medical Bloggers' Knowledge Sharing Behavior*. Proceedings from ASIS&T Annual Meeting, Vancouver.

Kazmer, M., & Xie, B. (2008). Qualitative interviewing in internet studies: Playing with the media, playing with the method. *Information, Communication & Society*, *11*(2), 257-278. doi:10.1080/13691180801946333

Keller, E. F. (1978). Gender and Science. *Psychoanalysis and Contemporary Thought*, *1*(3), 409-433.

Kennedy, J. (2010). *Arsenic-associated bacteria (NASA's claims).* [Web log comment] Retrieved January 12, 2011 from http://rrresearch.blogspot.com/2010/12/arsenic-associated-bacteria-nasas.html

King, D. W., Tenopir, C., Choemprayong, S., & Wu, L. (2009). Scholarly journal information-seeking and reading patterns of faculty at five US universities. *Learned Publishing*, *22*, 126-144. doi:10.1087/2009208

Kirkup, G. (2010). Academic blogging: Academic practice and academic identity. *London Review of Education*, *8*(1), 75-84.

Kjellberg, S. (2009). Scholarly blogging practice as situated genre: An analytical framework based on genre theory. *Information Research*, *14*(3).

Kjellberg, S. (2010). I am a blogging researcher: Motivations for blogging in a scholarly context. *First Monday*, *15*(8).

Kleinman, D. L. (1998). Untangling context: Understanding a university laboratory in the commercial world. *Science, Technology & Human Values*, *23*(3), 285-314.

Kling, R., & Iacono, S. (1988). The Mobilization of Support for Computerization: The Role of Computerization Movements. *Social Problems*, *35*(3), 226-243.

Kling, R., & Iacono, S. (1994). Computerization Movements and the Mobilization of Support for Computerization. In L. Starr (Ed.), *Ecologies of Knowledge.* NYC: SUNY Press.

Kling, R., McKim, G., & King, A. (2003). A bit more to it: Scholarly Communication Forums as Socio-Technical Interaction Networks. *Journal of the American Society for Information Science and Technology*, *54*(1), 47-67.

Klokmose, C. N., & Zander, P. O. (2010, May 19-21). *Rethinking Laboratory Notebooks*. Proceedings from 9th International Conference on Designing Cooperative Systems, Aix-en-Provence, France.

Knapp, L.G., Kelly-Reid, J.E., and Ginder, S.A. (2010). *Employees in Postsecondary Institutions, Fall 2009, and Salaries of Full Time Instructional Staff, 2009-10* (NCES 2011-150). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved 25 September 2011 from http://nces.ed.gov/pubsearch.

Knorr-Cetina, K. (1999). *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA.: Harvard University Press.

Knorr-Cetina, K. (2007). Culture in global knowledge societies: knowledge cultures and epistemic cultures. *Interdisciplinary Science Review*, *32*(4), 361-375.

Knorr-Cetina, K. (1981). *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science.* London: Pergamon Pr.

Koku, E., Nazar, N., & Wellman, B. (2001). Netting Scholars: Online and Offline. *American behavioral scientist*, *44*(10), 1752.

Kouper, I. (2010). Science blogs and public engagement with science: Practices, challenges, and opportunities. *Journal of Science Communication*, *9*(1), 1-10.

Kousha, K., & Thelwall, M. (2007). How is science cited on the web? A classification of google unique web citations. *Journal of the American Society for Information Science and Technology*, *58*(11), 1631-1644.

Kuhn, T. (1970). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Kuhn, T. (1977). *The essential tension : selected studies in scientific tradition and change*. Chicago: University of Chicago Press.

Kvale, S. (1996). *InterViews: An Introduction to Qualitative Research Interviewing.* Thousand Oaks, CA.: Sage Publications, Inc.

LaFollette, M. C. (1992). *Stealing into print : fraud, plagiarism, and misconduct in scientific publishing*. Berkeley: University of California Press.

Lamont, M. (2009). *How Professors Think: Inside the Curious World of Academic Judgment.* Cambridge, MA.: Harvard University Press.

Lassi, M., & Sonnenwald, D. (2010). Identifying factors that may impact the adoption and use of a social science collaboratory: a synthesis of previous research. *Information Research*, *15*(3). Retrieved from http://InformationR.net/ir/15-3/colis7/colis710.html

Lathe, W., Williams, J., Mangan, M., & Karolchik, D. (2008). Genomic Data Resources: Challenges and Promises. *Nature Education*, *1*(3).

Latour, B. (1988). *Science in Action: How to Follow Scientists and Engineers through Society.* Cambridge, MA.: Harvard University Press.

Latour, B., & Woolgar, S. (1986). *Laboratory Life.* Princeton, N.J.: Princeton University Press.

Lau, C. Y., Duan, H., Wang, F., He, C. B., Low, H. Y., & Yang, J. K. W. (2011). Enhanced Ordering in Gold Nanoparticles Self-Assembly through Excess Free Ligands. *Langmuir*, *27*(7), 3355-3360.

Leahey, E. (2006). Gender Differences in Productivity: Research Specialization as a Missing Link. *Gender & Society*, *20*(6), 754-780.

Lenhart, A., & Fox, S. (2006). Bloggers: A Portrait of the Internet's new storytellers. *PEW Internet,* 1-25.

Lenhart, A., Purcell, K., Smith, A., & Zickuhr, K. (2010). Social Media & Mobile Internet Use Among Teens and Young Adults. *Pew Internet & American Life Project,* 1-37.

Leonelli, S. (2007). Arabidopsis, the botanical Drosophila: from mouse cress to model organism. *Endeavour*, *31*(1), 34-38.

Lievrouw, L., & Carley, K. (1990). Changing patterns of communication among scientists in an era of "telescience.". *Technology in Society*, *12*, 457-477.

Lievrouw, L. (2004). Biotechnology, intellectual property, and the prospects for scholarly communication. In S. Braman (Ed.), *Biotechnology and Communication: The Meta-Technologies of Information* (pp. 145-172). Mahwah, NJ: Lawrence Erlbaum Associates.

Lievrouw, L., A. (2010). Social Media and the Production of Knowledge: A Return to Little Science? *Social Epistemology: A Journal of Knowledge, Culture and Policy*, *24*(3), 219-219. doi:10.1080/02691728.2010.499177

Linzer, M., Brown, T., Frazier, L., DeLong, E., & Siegel, W. (1988). Impact of a Medical Journal Club on House-Staff Reading Habits, Knowledge, and Critical Appraisal Skills. *Journal of the American Medical Association*, *260*(17), 2537-2541. doi:10.1001/jama.1988.03410170085039

Long, J. S. (1990). The Origins of Sex Differences in Science. *Social Forces*, *68*(4), 1297-1316.

Luzón, M. J. (2008). Scholarly hyperwriting: The function of links in academic weblogs. *Journal of the American Society for Information Science and Technology*, *5 Sept*.

Lynch, M. (2002). Protocols, practices, and the reproduction of technique in molecular biology. *British Journal of Sociology*, *53*(2), 203-220. doi:10.1080/00071310220133304

Lynch, M. (1985). *Art and Artifact in Laboratory Science: A Study of Shop Work and Shop Talk in a Research Laboratory (Studies in Ethnomethodology).* London: Routledge & Kegan Paul.

Mabe, M., & Amin, M. (2001). Growth dynamics of scholarly and scientific journals. *Scientometrics*, *51*(1), 147-162. doi:10.1023/A:1010520913124

Mackay, W., E., Pothier, G., Letondal, C., Boegh, K., & Sorensen, H. E. (2002, October 27-30). *The missing link*. Proceedings from the 15th annual ACM symposium on User interface software and technology - UIST '02, Paris, France. doi:10.1145/571985.571992

Maloney, A. (2011a). *User talk:Andy Maloney*. Retrieved 15 August, 2011 from http://www.openwetware.org/wiki/User_talk:Andy_Maloney#Committee.27s_response_to_1st_draft_of_dissertation_and_Andy.27s_presentation

Maloney, A. (2011b). *Yep, open science works!* Retrieved 15 August, 2011 from http://ff.im/yH6Yb

McCain, K. W. (1991). Communication, competition, and secrecy: the production and dissemination of research-related information in genetics. *Science, Technology & Human Values*, *16*(4), 491. doi:10.1177/016224399101600404

Mead, J., A., Shadforth, I., P., & Bessant, C. (2007). Public proteomic MS repositories and pipelines: available tools and biological applications. *Proteomics*, *7*(16), 2769-2786.

Meadows, A. J. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities.* Chicago: University of Chicago Press.

Meadows, A. J. (1998). *Communicating Research.* San Diego, CA.: Academic Press.

Merton, R. K. (1936). The Unanticipated Consequences of Purposive Social Action. *American Sociological Review*, *1*(6), 894-904.

Merton, R. K. (1961). Singletons and Multiples in Scientific Discovery: A Chapter in the Sociology of Science. *Proceedings of the American Philosophical Society*, 105(5), 470-486.

Meyer, E. T., Rosenbaum, H., & Hara, N. (2005, Oct 6-9). *How Photobloggers are Framing a New Computerization Movement*. Proceedings from Internet Generations 6.0: Annual conference of the Association of Internet Researchers, Chicago, Illinois, USA.

Meyer, E. T. (2006). Socio-Technical Interaction Networks: A Discussion of the Strengths, Weaknesses and Future of Kling's STIN Model. In J. Berleur, M. I. Numinen, & J. Impagliazzo (Eds.), *IFIP International Federation for Information Processing, (223). Social Informatics: An Information Society for all? In Remembrance of Rob Kling* (pp. 37-48). Boston: Springer.

Meyer, E. T. (2011). *Splashes and Ripples: Synthesizing the Evidence on the Impact of Digital Resources*. London: JISC.

Meyer, E. T., & Schroeder, R. (2009). Untangling the web of e-Research: Towards a sociology of online knowledge. *Journal of Informetrics*, *3*(3), 246-260.

Messer-Davidow, E. (2002). *Disciplining Feminism: From Social Activism to Academic Discourse*. Durham, NC: Duke University Press.

Miles, M., & Huberman, A. (1984). *Qualitative data analysis: A sourcebook of new methods.* Beverly Hills, CA: Sage.

Moed, H., F. (2007). The effect of 'open access' on citation impact: An analysis of arXiv's condensed matter section. *Journal of the American Society for Information Science and Technology*, *58*(13), 2047-2054.

Mortensen, T., & Walker, J. (2002). Blogging Thoughts: Personal Publication as an Online Research Tool. In A. Morrison (Ed.), *Researching ICTs in Context* (pp. 249-279). Oslo: InterMedia.

Mulkay, M., & Edge, D. (1976). Cognitive, Technical, and Social Factors in the Growth of Radio Astronomy. In G. Lemaine, R. MacLeod, M. Mulkay, and P. Weingart (Eds.), *Perspectives in the Emergence of Scientific Disciplines* (pp. 153-186).Chicago, IL: Aldine.

Mullins, N. (1973) The Development of Specialties in Social Science: The Case of Ethnomethodology. *Science Studies*, 3(3), pp. 245-273.

Munger, D. (2010). *Blogging out of balance*. Retrieved June 10, 2011 from http://seedmagazine.com/content/article/blogging_out_of_balance

Munger, D. (2011). *Researchblogging*. Retrieved 10 July, 2011 from http://www.researchblogging.org/

Myers, P. Z. (2010). *Say hello to…PepsiCo???!? WTF?* Retrieved 25 Aug, 2011 from http://scienceblogs.com/pharyngula/2010/07/say_hello_topepsico_wtf.php

Nardi, B. A., Schiano, D. J., & Gumbrecht, M. (2004, November 6-10). *Blogging as social activity, or, would you let 900 million people read your diary?* Proceedings from the 2004 ACM conference on Computer supported cooperative work, Chicago, Illinois, USA.

Nentwich, M. (2004). *Cyberscience: Research in the age of the internet.* Austria: Austrian Academy of Sciences.

NESTA & RIN. (2009). *Open to All? Case studies of openness in research.* Retrieved from http://www.rin.ac.uk/our-work/data-management-and-curation/open-science-case-studies

Newton, Isaac. "Portsmouth Add. MS. 3973". *The Chymistry of Isaac Newton*. Ed. William R. Newman 2006. Retrieved June 16, 2011 from: http://purl.dlib.indiana.edu/iudl/newton/ALCH00109.

Neylon, C. (2009). Stitching science together. *Nature*, *461*, 881.

Neylon, C. (2009). *Call for JCOM papers on participative media*. Retrieved May 5, 2010 from http://friendfeed.com/cameronneylon/3301eb89/call-for-jcom-papers-on-participative-media

Neylon, C. (2010a). *The truth…well most of the truth anyway*. Retrieved 15 August, 2011 from http://cameronneylon.net/blog/the-truth-well-most-of-the-truth-anyway/

Neylon, C. (2010b). *P ≠ NP and the future of peer review*. Retrieved 15 August, 2011 from http://cameronneylon.net/blog/p-%E2%89%A0-np-and-the-future-of-peer-review/

Neylon, C. (2011a, August 5). *My submission to Royal Society enquiry on #openscience "Science as a Public Enterprise" http://bit.ly/oBloil Enquiry: http://bit.ly/quC9dl* [Twitter post]. Retrieved from http://twitter.com/#!/CameronNeylon/status/99431733577003009

Neylon, C. (2011b, August 5). *Submission to the Royal Society Enquiry*. [Friendfeed post] Retrieved from http://friendfeed.com/cameronneylon/ab0fd7ae/submission-to-royal-society-enquiry

Nielsen, M. (2009). Doing Science in the Open. *Physics World*, *May 1*, 1-9.

Nielsen, M. (2011). *Reinventing Discovery: The New Era of Networked Science*. Princeton University Press.

Nonnecke, B., & Preece, J. (2001, August 3-5). *Why Lurkers Lurk*. Proceedings from AMCIS Conference. Boston, USA.

Nonnecke, B., & Preece, J. (2000, April 1-6). *Lurker demographics: counting the silent*. Proceedings from Conference on Human Factors in Computing Systems, The Hague.

Olson, G. M., Zimmerman, A., & Bos, N. (2008). *Scientific Collaboration on the Internet (Acting with Technology).* Cambridge, MA.: The MIT Press.

ONS Challenge Wiki. (2010). *RSC Awards 2010 - Rules and Application Form*. Retrieved July 15, 2011 from http://onschallenge.wikispaces.com/RSCAwards2010

ONS Challenge Wiki. (2011). *Open Notebook Science Challenge*. Retrieved July 15, 2011 from http://onschallenge.wikispaces.com/

Papacharissi, Z. (2006). Audiences as media producers: Content analysis of 260 blogs. In M. Tremayne (Ed.), *Blogging, Citizenship and the Future of Media.* New York: Routledge.

Parker, J., & Hackett, E. J. (2012). Hot Spots and Hot Moments in Scientific Collaborations. *American Sociological Review, 77*(1),21-44 . doi: 10.1177/000312241143376

Parkhe, A. (1993). "Messy" Research, Methodological Predispositions, and Theory Development in International Joint Ventures. *The Academy of Management Review*, *18*(2), 227-268. doi:10.2307/258759

Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods.* Thousand Oaks, CA.: Sage Publications, Inc.

Pennisi, E. (2011). Concerns About Arsenic-Laden Bacterium Aired. *Science*, *332*(6034), 1136-1137. doi:10.1126/science.332.6034.1136

PetHW. (2010). *Arsenic-associated bacteria (NASA's claims)*. [Web log comment] Retrieved January 12, 2011 from http://rrresearch.blogspot.com/2010/12/arsenic-associated-bacteria-nasas.html

Polonsky, A., Six, A., Kotelnikov, M., Polonsky, V., Polly, R., & Brey, P. (2006, June 12). *Semantic Laboratory Notebook: Managing Biomedical Research Notes and Experimental Data*. Proceedings from ESWC 2006 Industry Forum, Budva, Montenegro.

Powell, W. W., White, D. R., Koput, K. W., & Owen-Smith, J. (2005). Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. *American Journal of Sociology*, *110*(4), 1132-1205. doi:10.1086/421508

Power, L. (2008). *Scientific Blogging: Content and Motivation*. Unpublished manuscript, Oxford Internet Institute, University of Oxford, Oxford.

Preece, J., Nonnecke, B., & Andrews, D. (2004). The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior*, *20*(2), 201-223. doi:0.1016/j.chb.2003.10.015

Price, D. H. (2010). Blogging Anthropology: Savage minds, Zero Anthropology, and AAA blogs. *American Anthropologist*, *112*(1), 140-142.

Price, D. J. de S. (1963). *Little Science, Big Science.* Columbia University Press.

Printz, S. (2010). *Arsenic-associated bacteria (NASA's claims).* [Web log comment] Retrieved January 12, 2011 from http://rrresearch.blogspot.com/2010/12/arsenic-associated-bacteria-nasas.html

Rabinow, P. (1997). *Making PCR: A Story of Biotechnology.* Chicago: University Of Chicago Press.

Redfield, R. (2010). *Arsenic-associated bacteria (NASA's claims)*. Retrieved January 12, 2011 from http://rrresearch.blogspot.com/2010/12/arsenic-associated-bacteria-nasas.html

Redfield, R. J. (2011). Comment on "A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus". *Science*, *332*(6034), 1149. doi:10.1126/science.1201482

Rhodes, R. (1995). *The Making of the Atomic Bomb*. New York: Simon & Schuster.

Rodriguez, V. (2009). Access to data and material for research: putting empirical evidence into perspective. *New Genetics and Society*, *28*(1), 67-86. doi:10.1080/14636770802670274

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.

Rogers, M. E., & Onken, J. (2008). The National Institute of General Medical Sciences Glue Grant Program. In G. M. Olson, A. Zimmerman, & N. Bos (Eds.), *Scientific Collaboration on the Internet (Acting with Technology)* (pp. 197-219). Cambridge, MA.: The MIT Press.

Rubin, D. H. J., & Rubin, D. I. S. (2004). *Qualitative Interviewing: The Art of Hearing Data.* Thousand Oaks, CA.: Sage Publications, Inc.

Sanderson, K. (2008). Data on display. *Nature*, September 2008. doi:10.1038/455273a

Saunders, N. (2011). *Farewell FriendFeed. It's been fun*. Retrieved Aug 25, 2011 from http://nsaunders.wordpress.com/2011/01/31/farewell-friendfeed-its-been-fun/

Saunders, N., Beltrao, P., Jensen, L., Jurczak, D., Krause, R., Kuhn, M., & Wu, S. (2009). Microblogging the ISMB: A New Approach to Conference Reporting. *PLoS Computational Biology*, *5*(1). doi:10.1371/journal.pcbi.1000263

schraefel, m. c., Hughes, G., Mills, H., Smith, G., Payne, T. and Frey, J. (2009, April 24-29). *Breaking the Book: Translating the Chemistry Lab Book into a Pervasive Computing Lab Environment*. Proceedings from CHI 2004, Vienna, Austria.

Schroeder, R. (2007). e-Research Infrastructures and Open Science: Towards a New System of Knowledge Production? *Prometheus*, *25*(1), 1-17.

Science Commons. (n.d.). *Protocol for Implementing Open Access Data*. Retrieved Aug 25, 2011 from http://sciencecommons.org/projects/publishing/open-access-data-protocol/

Scienceseeker.org. (2011). *About*. Retrieved Aug 22, 2011 from http://scienceseeker.org/about/

Secretary of State for Energy and Climate Change. (2010). *Government Response to the House of Commons Science and Technology 8th Report of Session 2009–10: The disclosure of climate data from the Climatic Research Unit at the University of East Anglia.* Norwich: The Stationery Office.

Seidman, S, ed. (1996). *Queer Theory/Sociology*. Cambridge, England: Blackwell.

Shaevitz, J. (2010). *My Lab's Wiki-based Electronic Lab Notebook System*. Retrieved 15 August, 2011 from http://www.forcesoflife.com/2010/06/my-labs-wiki-based-electronic-lab.html

Shema, H., & Bar-Ilan, J. (2011, June 14). *Characteristics of Researchblogging.org science Blogs and Bloggers*. Proceedings from altmetrics11: Tracking scholarly impact on the social Web, Koblenz, Germany.

Shneiderman, B. (2008). Science 2.0. *Science*, *319*(5868), 1349-1350. doi:10.1126/science.1153539

Shrum, W., Genuth, J., & Chompalov, I. (2007). *Structures of Scientific Collaboration.* MIT Press.

Singh, D. (2008). *The BioGang*. Retrieved March 15, 2010 from http://biogang.wikispaces.com/

Small, M.L. (1999). Departmental Conditions and the Emergence of New Disciplines: Two Cases in the Legitimation of African American Studies. *Theory and Society 28*, 659–707.

Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology*, *41*, 643-681.

Stack, S. (2004). Gender, Children and Research Productivity. *Research in Higher Education*, *45*(8), 891-920.

Stake, R. E. (2005). Qualitative Case Studies. In Denzin & Lincoln (Eds.), *Handbook of qualitative research* (3rd ed.). Thousand Oaks, CA: Sage.

Stephens, N. (2007). Collecting data from elites and ultra elites: Telephone and face-to-face interviews with macroeconomists. *Qualitative Research*, *7*(2), 203-215.

Tabard, A., Mackay, W., E., & Eastmond, E. (2008, November 8-12). *From individual to collaborative: the evolution of prism, a hybrid laboratory notebook*. Proceedings from the 2008 ACM conference on Computer supported cooperative work, San Diego, CA, USA. doi:10.1145/1460563.1460653

Tang, J.-T., Wang, T., & Wang, J. (2009, July 12-15). *Measuring the influence of social networks on information diffusion on blogspheres*. Proceedings from 2009 International Conference on Machine Learning and Cybernetics. Baoding, China. doi: 10.1109/ICMLC.2009.5212771

Taylor, G. (2006). *About The Synaptic Leap*. Retrieved Aug 20, 2011 from http://www.thesynapticleap.org/about

Taylor, K. T. (2006). The status of electronic laboratory notebooks for chemistry and biology. *Current Opinion in Drug Discovery & Development*, *9*(3), 348-353.

Tenopir, C., & King, D. W. (2008). Electronic Journals and Changes in Scholarly Article Seeking and Reading Patterns. *DLib Magazine*, *14*(11/12).

The Distant One. (2010). *Arsenic-associated bacteria (NASA's claims).* [Web log comment] Retrieved January 12, 2011 from http://rrresearch.blogspot.com/2010/12/arsenic-associated-bacteria-nasas.html

The Economist (2004, Jun 10). An open-source shot in the arm? *The Economist*,. Retrieved from http://www.economist.com/node/2724420

The Nielsen Company. (2011). *Blogpulse.com*. Retrieved 18 August, 2011 from http://blogpulse.com/

The Royal Society. (2011). *Knowledge, Networks and Nations: Global scientific collaboration in the 21st Century* (RS Policy Document 03/11). London: Elsevier.

Thelwall, M., & Kousha, K. (2008). Online presentations as a source of scientific impact? An analysis of PowerPoint files citing academic journals. *Journal of the American Society for Information Science and Technology*, *59*(5), 805-815.

Thelwall, M., & Stuart, D. (2006). Web crawling ethics revisited: Cost, privacy, and denial of service. *Journal of the American Society for Information Science and Technology*, *57*(13), 1771-1779.

Todd, M. H. (2010). *Comment on AJ Cann's Friendfeed: Where's the Open Science? 7 Mar 2010*. Retrieved Aug 22, 2011 from http://friendfeed.com/ajc/84486d9d/where-open-science

Todd, M., & Google. (2010, April 6). *Chemistry on the Web: How Can we Crowdsource Chemistry to Solve Important Problems?* [YouTube video] Retrieved August 31, 2011 from http://www.youtube.com/watch?v=yWnTJw042OM

Trammell, K. D., & Keshelashvili, A. (2005). Examining the new influencers: A self-presentation study of a-list blogs. *Journalism and Mass Communication Quarterly*, *82*(4), 968-968.

Trammell, K. D., Tarkowski, A., Hofmokl, J., & Sapp, A., M. (2006). Rzeczpospolita blogow [Republic of Blog]: Examining Polish Bloggers Through Content Analysis. *Journal of Computer-Mediated Communication*, *11*(3), 702-722.

Traweek, S. (1992). *Beamtimes and lifetimes: the world of high energy physicists.* Cambridge, MA.: Harvard University Press.

Tremayne, M. (2006). *Blogging, Citizenship and the Future of Media.* New York: Routledge.

Trevino, E. M. (2005, October 6-9). *Blogger motivations: Power, pull, and positive feedback*. Proceedings from Internet Research 6.0, Chicago.

UNESCO. (2010). *UNESCO Science Report 2010 The Current Status of Science around the World*. Retrieved July 25, 2011 from http://www.unesco.org/new/en/natural-sciences/science-technology/prospective-studies/unesco-science-report/

UNESCO. (2011). *Researchers by field of science (FTE and HC)*. Retrieved July 25, 2011 from http://stats.uis.unesco.org/unesco/TableViewer/tableView.aspx?ReportId=2642

University of Oxford Research Services Webmaster. (2010). *Keeping Laboratory Notebooks*. Retrieved 18 August, 2011 from http://www.admin.ox.ac.uk/rdm/dmp/documentation/labnotebooks/

Vakkari, P. (2008). Perceived influence of the use of electronic information resources on scholarly work and publication productivity. *Journal of the American Society for Information Science and Technology*, *59*(4), 602-612.

Vaughan, L., & Shaw, D. (2005). Web citation data for impact assessment: A comparison of four science disciplines. *Journal of the American Society for Information Science and Technology*, *56*(10), 1075-1087.

Velden, T., & Lagoze, C. (2009). Communicating chemistry. *Nature Chemistry*, *1*(9), 673-678. doi:10.1038/nchem.448

Waldrop, M. M. (2008). Science 2.0: Great New Tool, or Great Risk? *Scientific American*, 9 Jan.

Walker, J. (2006). Blogging from Inside the Ivory Tower. In A. Bruns & J. Jacobs (Eds.), *Uses of Blogs* (pp. 1-11). New York: Peter Lang.

Walsh, J. P., Kucker, S., Maloney, N. G., & Gabbay, S. (2000). Connecting minds: Computer-mediated communication and scientific work. *Journal of the American Society for Information Science*, *51*(14), 1295-1305.

Walsh, J. P., & Maloney, N. G. (2007). Collaboration Structure, Communication Media, and Problems in Scientific Work Teams. *Journal of Computer-Mediated Communication*,. doi:10.1111/j.1083-6101.2007.00346.

Weitzman, E., & Miles, M. (1995). *Computer programs for qualitative data analysis.* Thousand Oaks, CA: Sage.

Welsh, E., Jirotka, M., & Gavaghan, D. (2006). Post-genomic science: cross-disciplinary and large-scale collaborative research and its organizational and technological challenges for the scientific research process. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *364*(1843), 1533-1549.

Whitley, R. (1985). *The intellectual and social organization of the sciences.* Oxford: Oxford University Press.

Wickman, C. (2010). Writing Material in Chemical Physics Research: The Laboratory Notebook as Locus of Technical and Textual Integration. *Written Communication*, *27*(3), 259-292. doi:10.1177/0741088310371777

Wilbanks, J. (2006). Another reason for opening access to research. *British Medical Journal*, 333(7582), 1306-1308. doi:10.1136/ sbmj.39063.730660.F7

Wilkins, J. S. (2008). The roles, reasons and restrictions of science blogs. *Trends in Ecology & Evolution*, *23*(8), 411-413.

Williams, J., & Clark, J. D. (1992). The Information Explosion: Fact or Myth? *IEEE Transactions on Engineering Management*, *39*(1), 79-84.

Willinsky, J. (2006). *The Access Principle: The Case for Open Access to Research and Scholarship*. Cambridge, MA.; London, England: MIT Press.

Wolfe-Simon, F., Blum, J. S., Kulp, T. R., Gordon, G. W., Hoeft, S. E., Pett-Ridge, J., … Oremland, R. S. (2011). A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus. *Science*, *332*(6034), 1163-1166. doi:10.1126/science.1197258

Woolgar, S. (1982). Laboratory studies: A comment on the state of the art. *Social studies of science*, *12*(4), 481.

Wuchty, S., Jones, B., F., & Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*, *316*(5827), 1036-1039. doi:10.1126/science.1136099

Yin, R. K. (2003). *Case study research: Design and methods: 005*. Thousand Oaks, CA.: Sage Publications, Inc.

Yong, E. (2010). *Arsenic bacteria – a post-mortem, a review, and some navel-gazing*. Retrieved Aug 25, 2011 from http://blogs.discovermagazine.com/notrocketscience/2010/12/10/arsenic-bacteria-a-post-mortem-a-review-and-some-navel-gazing/

Zimmer, C. (2010). *Oh, Pepsi, What Hast Thou Wrought?* Retrieved July 25, 2011 from http://blogs.discovermagazine.com/loom/2010/07/07/oh-pepsi-what-hath-thou-wrought/

Zivkovic, B. (2011). *Science blogs and public engagement with science*. Retrieved 10 August, 2011 from http://blog.coturnix.org/2010/03/08/science_blogs_and_public_engag/

Zivkovic, B. (2010). *A Farewell to Scienceblogs: the Changing Science Blogging Ecosystem*. Retrieved 15 August, 2011 from http://blog.coturnix.org/2010/07/19/a-farewell-to-scienceblogs-the-changing-science-blogging-ecosystem/

Zivkovic, B., Bradley, J-C., Stemwedel, J., Edwards, P., & Vaughan, K. T. L. (2007). Opening science to all: Implications of blogs and wikis for social

and scholarly scientific communication. *Proceedings of the American Society for Information Science and Technology*, *44*(1), 1-3.

Zuccala, A. (2006). Modeling the invisible college. *Journal of the American Society for Information Science and Technology*, *57*(2), 152-168. doi:10.1002/asi.20256

Zuckerman, H. (1977). *Scientific Elite: Nobel Laureates in the United States.* New York: Free Press.