Examining the Structures and Practices for Knowledge Production within Galaxy Zoo - an Online Citizen Science Initiative

Dissertation submitted to the University of Oxford for the degree of

DPhil in Education

Bipana Bantawa   Lincoln College, Trinity 2014

Table of Contents

Abstract

This study examines the ways in which public participation in the production of scientific knowledge, influences the practices and expertise of the scientists in Galaxy Zoo, an online Big Data citizen science initiative. The need for citizen science in the field of Astronomy arose in response to the challenges of rapid advances in data gathering technologies, which demanded pattern recognition capabilities that were too advanced for existing computer algorithms. To address these challenges, Galaxy Zoo scientists recruited volunteers through their online website, a strategy which proved to be remarkably reliable and efficient. In doing so, they opened up the boundaries of scientific processes to the public. This shift has led to important outcomes in terms of the scientific discovery of new Astronomical objects; the creation and refining of scientific practices; and the development of new forms of expertise among key actors while they continue to pursue their scientific goals.

This thesis attempts to answer the over-arching research question: How is citizen science shaping the practices and expertise of Galaxy Zoo scientists? The emergence of new practices and development of the expertise in the domain of managing citizen science projects were observed through following the work of the Galaxy Zoo scientists and in particular the Principal Investigator and the project's Technical Lead, from February 2010 to April 2013. A broadly ethnographic approach was taken, which allowed the study to be sensitive to the uncertainty and unprecedented events that characterised the development of Galaxy Zoo as a pioneering project in the field of data-intensive citizen science. Unstructured interviewing was the major source of data on the work of the PI and TL; while the communication between these participants, the broader Science Team and their inter-institutional collaborators was captured through analyses of the team emailing list, their official blog and their social media posts.

The process of data analysis was informed by an initial conceptualisation of Galaxy Zoo as a knowledge production system and the concept of knowledge object (Knorr-Cetina,1999), as an unfolding epistemic entity, became a primary analytical tool. Since the direction and future of Galaxy Zoo involved addressing new challenges, the study demanded periodic recursive analysis of the conceptual framework and the knowledge objects of both Galaxy Zoo and the present examination of its development.

The key findings were as follows. The involvement of public volunteers shaped the practices of the Science Team, while they pursued robust scientific outcomes. Changes included: negotiating collaborations; designing the classification tasks for the volunteers; re-examining data reduction methods and data release policies; disseminating results; creating new epistemic communities; and science communication. In addition, new kinds of expertise involved in running Galaxy Zoo were identified.

The relational and adaptive aspects of expertise were seen as important. It was therefore proposed that the development of the expertise in running citizen science projects should be recognised as a domain-expertise in its own right. In Galaxy Zoo, the development of the expertise could be attributed to a combined understanding of: the design principles of doing good science; innovation in methods; and creating a dialogic space for scientists and volunteers. The empirical and theoretical implications of this study therefore lie in (i) identifying emergent practices in citizen science while prioritising scientific knowledge production and (ii) a re-examination of expertise for science in the emerging context of data-intensive science.

Keywords: Citizen Science, Big Data Science, Scientific Expertise, Relational Expertise, Scientific Knowledge Production, Epistemic Architecture

Abbreviations

AE	Adaptive Expertise
CAS	Complex Adaptive Systems
CI	Collective Intelligence
CSCW	Computer Supported Cooperative Work
D(A)	Domain Expertise in Astronomy
D(CS)	Domain Expertise in Citizen Science
D(ML)	Domain Expertise in Machine Learning
D(T)	Domain Expertise in Technical Development
DI	Distributed Intelligence
EL	Educational Lead
ET	Education Team
FAQs	Frequently Asked Questions
GZ	Galaxy Zoo
HEFCE	Higher Education Funding Council for England
IE	Interactional Expertise
LSST	Large Survey Synoptic Telescope
MNRAS	Monthly Notices of the Royal Astronomical Society
MZ	Moon Zoo
NAM	National Astronomy Meeting
NASA	National Aeronautics and Space Agency
NSF	National Science Foundation
PI	Principal Investigator
RE	Relational Expertise
SDSS	Sloan Digital Sky Survey
SETI	Search for Extra Terrestrial Intelligence
SL	Science
ST	Science Team
STS	Science and Technology Studies
TL	Technical Lead
UBRET	User-Based Research Enabling Toolkit

Acknowledgements

Throughout my DPhil, I have watched the mist of doubt dance around the mountain of questions I had built, revealing a mesmerising landscape of inquiries. Not only has this epistemic journey has been deeply humbling, but it certainly has been a tremendously rewarding experience, for which I must thank many.

Professor Anne Edwards has been the most compassionate and engaging supervisor any student could ever ask for. Her enthusiasm, passion and dedication in her research has been a constant source of inspiration. Her thorough and detailed critique has been invaluable in the coming together of my doctoral thesis. My DPhil would not have been possible without her inexhaustible encouragement, patience and whole-hearted support.

Special thanks to Chris and Arfon, who were absolutely critical to my research as they welcomed me into their world and embraced my research without knowing what was going on. I had the opportunity to study incredibly dynamic individuals at work and I am very grateful for their openness, kindness and commitment to citizen science. I would also like to extend my gratitude to the entire Galaxy Zoo Team who allowed me to participate during the first Zooniverse meeting in Chicago.

It would be a grave error to assume that academic ability alone leads to a DPhil. As the bubble of self-doubt and obsessiveness takes over, it becomes increasingly important that one is reminded that ‘it is just a thesis’ and there is more to life - friendship, laughter, compassion, and empathy. There are many individuals who ensured that my wellbeing and sanity was not not sacrificed at the alter of knowledge. Carmella, thank you for your optimism, warmth and your tireless support. To the Bean Crew, for being the friendly faces that cheered me up every day and to the Missing Bean for being my refuge and my second home in Oxford. To the Alpha bar, crew for making time to chat about my progress in spite of the long queues. To staff of Lincoln College, the Department of Education and the Bodleian libraries, who have supported me in more ways than I perhaps realise. Special thanks to the very sympathetic college porters who let me work in the library beyond official hours during the last few weeks before my thesis submission. To the momoship for being the excuse I got to see many of my friends every Wednesday and for nurturing disillusioned students with some steamed goodness. To Jo at the International office for all her help with the tedious immigration procedures.

Thanks to my dear dudes, Tania and Nahal, for being my family in Oxford and the collective craziness and laughter without which I would have not survived the DPhil. To Mahmoud, the grappler, for being so wonderfully curious and unexpectedly ridiculous. To Pegah, for your warmth and our morning conversations over coffee. To Muneera, for not letting me be the only monkey in the room. To my fellow PRS colleagues, Nalini, Wan Ying, Jessica, Andrew and Anna for being such a lively bunch. To Ashish, for looking out for me. To Janeth, Joe and Annie for being such lovely souls with a generous dash of silliness. To Henry and Gunnar, for patiently answering my science-related questions and the general hilarity. To Tim, for your randomness. To James, for your encouragement and sharing your curiosity; and for Port O’ Clock. To Henning and Max for your love of whisky and thoughtful conversations. To Rich, Passmore, Xavier, Hong Sheng and Jorge, for putting the merry in college Christmas dinners. To Alex, for your inquisitiveness and PhTea. To Usha and Amogh, for checking up on me and the dosa-fuelled debates. To Vikaran and Challenger, for Sltrings 2 Cosmos. To all the people I met along the way who kindly entertained my ideas, shared their stories and offered advice.

To Oxford, for being my second home.

To Apo, Pia and Googs, for always being there for me. To Rohan and Mikey, for your laughter and letting me crash your kitchen. To Shreya, Deepti, Adu and Kiz for your friendship.

To Shravan, thank you for your wonderful sense of openness to life, for always being my cheerleader and for being the more patient member of Team BS.

To my school teachers, Lalitha aunty, Sudha aunty, Urshia di and Gopalan uncle for allowing me the freedom to enquire beyond what was prescribed.

To mama and maiju,for supporting me in so many ways since my undergraduate days. To my grandparents, for their courage in life and their emphasis on the value of education.

Finally, my deepest gratitude goes to my parents. Their unconditional love, support and understanding has been the greatest and the most infallible force in my life. To Amaji, for your resilience and your ability to enjoy simple things. To Manis, for your spirited sense ofadventure and inquiry, for all the stories you read to me, for the freedom I had to question and disagree and for teaching me that every person is a human being first. Manis and Amaji, I dedicate this thesis to you.

List of Figures and Tables

Figures

Tables

Chapter 1

Introduction to the Study

1.1 The Age of Digital Astronomy and Its Challenges

Recent advances in computational and informational technologies in the field of Astronomy, has led to a change in traditional methodologies and approaches to data analysis (Bredekamp, 2012). Way et al. (2012) suggest that the current epoch of Astronomy research could be aptly named ‘the Age of Digital Astronomy’ owing to the reliance on digital data. They suggest that, historically, data catalogues of Babylonian stargazers date back to 1500 BC with records of Astronomical data ranging from stone carvings to written records, hand-drawn illustrations, to photographs and currently to digital media (Goodman and Wong, 2009). While the invention of the telescope (c.1600) and access to the electromagnetic spectrum beyond the capacity of human vision (c.1930) resulted in qualitative changes within Astronomy research, the sheer volume of data captured through digital technologies remains unrivaled (Goodman and Wong, 2009). For instance, the Large Synoptic Survey Telescope (LSST) that is currently being designed, and is set for operations in the year 2023, will be equipped with the world’s largest camera with 3200 Megapixels, generating 30 Terabytes of data every night. This has significant implications for the ability to undertake data analysis, especially since

“[a]rchive massive amounts of intermediate computations, complex calibration tables, and theoretical simulations has altered scientific computation in major ways.” (Way *et al.,* 2012: xvii).

While there have been improvements in the data handling and processing capabilities, understanding and addressing the complexity of the prevalent condition of unprecedented amount of data (Way *et al*., 2012) still lacks attention particularly in Digital Astronomy, resulting in challenges not only in Astronomy, but a number of other fields (Baranuik, 2011, Hey *et al*., 2009).

One of the earliest pioneering solutions to the data deluge problem was SETI @ Home (Search for Extra Terrestrial Intelligence @ Home) which was launched in May 1999; a project led by several universities and the US government. It was the first of its kind based on a distributed computing model, whereby vast amounts of data recorded from radio telescopes were sent to personal computers of volunteers around the world connected by the Internet. The computers download ed the data and analysed it during the time permitted by the volunteers, and the results were sent back to SETI's headquarters at the University of Berkeley. As of 23 October, 2000, 2.4 million volunteers had downloaded the programme and around half a million were actively running them and returned a total of 437,000 years of Central Processing Unit (CPU) time in two weeks (Korpela *et al*., 2001). Although the intention of the project was also to incite public interest in the search for extra terrestrial intelligence, it was primarily aimed at utilising computing power to analyse large amounts of data.

A simple model of SETI can be described as the exchange of data between the SETI database and the volunteers' computers. This model only makes use of the computing power, while the volunteers remain as donors of their computing power. With an increasing trend towards public involvement in science, despite lack of clarity or consensus on the nature and role of public expertise, citizen science operated on the assumptions that it will increase understanding of the scientific process (Cohen, 1997 in Brossard *et al*. 2005). It is safe to state that while the non-scientists cannot be expected to replace scientists anytime in the near future, the extent to which non-scientists can learn, be trained and contribute towards scientific inquiry is yet to be explored. One such study of a citizen science project led by the Cornell Lab of Ornithology was one of the first of its kind to involve 700 participants, the results of which suggested that although involvement did not lead to scientific thinking, it did provide volunteers an avenue to engage in the habit of thinking about scientific phenomena (Trumbull *et al*., 2000). One of the interesting findings from the study was that methods of scientific observation and their instructions had to be designed for a lay observer, evolving methods to accommodate and train non-scientist to do science. Therefore, with the recent surge of interest in citizen science projects such as FoldIt[1], Quantum Moves[2] and Galaxy Zoo (GZ), the demands on the practices of scientists as a result of public participation in scientific inquiry, becomes a intriguing area of exploration.

1.2 Galaxy Zoo: A New Direction for Big Data Astronomy and Citizen Science

GZ, a citizen science initiative, was conceived with the objective to visually classify galaxies in the Sloan Digital Sky Survey (SDSS) catalogue, which would have taken a considerable amount of time for experts to classify (Lintott and Land, 2008). Since the human eye is superior in identifying the variations and differences in the galaxies photographed than any computer recognition software

[1] FoldIt is an online game-based protein folding project http://fold.it/portal/
[2] Quantum Moves is an online game-based project www.scienceathome.com

to date, the GZ Team decided to create a website through which volunteers could help classify galaxies into different shapes or categories. In July 2007, GZ went online.

Figure 1 A: Snapshot of Galaxy Zoo Website[3]

Although the initiative stemmed from similar problems of unmanageable amounts of data collected in the SETI @Home project, the solution provided was markedly different. Although a computer and the Internet was required, GZ required the volunteers to carry out the classifications. Volunteers are asked to take a tutorial online before attempting to classify galaxies and can participate in discussion forums if they encounter problems in making decisions in their classifications.

[3] Extracted on 28.04.10 from www.galaxyzoo.org/

Within the first day of the launch, the scientists received 70,000 classifications an hour, and by the end of the year 150,000 volunteers had classified 50 million galaxies (Lintott and Land, 2008).

One of the most remarkable discoveries since GZ's inception was made by a Dutch school teacher, who discovered an object, which had never been previously documented and was later named after her as Hanny's Voorwerp (Voorwerp literally translated as 'thing' or 'object' in Dutch). For a non-scientist to make a significant contribution to the field of the established scientific discipline of Astronomy[4], can be argued to be a distinguishing affordance of this new mode of science, now also known as citizen science. By 2010, GZ had reached their target of 60 million classifications and had 10 academic papers published and six more working papers in progress.

The involvement of non-scientists in Astronomy is not an entirely novel phenomenon, as amateurs have previously made discoveries by virtue of the observational nature of the discipline. However, the field has been transformed from an observational science to a science based on image processing of the phenomena observed (Mulkay,1974). These changes in the discipline of Astronomy are for example, facilitated by the new forms of technologies available for scientific research. Robotic telescopes are programmed to capture images of the night sky as set intervals, which used to be a task undertaken by Astronomers. Changes like these, as Mulkay (1974) argues, raise questions

[4] Astronomy is used interchangeably with Astrophysics in this thesis.

about shifts in kinds of expertise and knowledge that are brought about by changes that occur within the discipline.

Since GZ is a data- intensive undertaking involving non-professional volunteers in the process of scientific inquiry, it invites an exploration of an emerging context of online Big Data citizen science. The assumption here is that this mode of scientific research involving the public, scientists would have to re-examine their existing practices and perhaps create new ones. The definition of practice that I am using is that they are

"[h]istorically accumulated, knowledge-laden, emotionally freighted and given direction by what is valued by those who inhabit them", Edwards A (2010:7).

Furthermore, it would be plausible to assume in changing and creating practices suitable for citizen science, the expertise involved may be of a different nature than in scientific activities without public involvement.

The initial understanding of expertise I had before embarking upon the analysis of this research would resonate with notion put forward by Ericsson and Smith (1991:2)

"[t]the study of expertise seeks to understand the account for what distinguishes outstanding individuals in a domain from less outstanding individuals in that domain, as well as from people in general."

1.3 Constructing the Research Question

With the possibility of scientists developing a new set of domain-expertise in running citizen science initiatives, GZ provides an intriguing context for an investigation into the possible implications of new modes of scientific inquiry for the production of scientific knowledge. Given this assumption, the present study attempts to address the following overarching research question and three specific questions that arise from it.

How is citizen science shaping the practices and expertise of Galaxy Zoo scientists?

1) In what way is citizen science changing scientific practice?

2) What practices can be identified in the work of Galaxy Zoo as a successful citizen science initiative?

3) What kinds of expertise are evident in these citizen science practices and how are they developed?

1.4 Organisation of this Study

The aim of this chapter is to begin to introduce the background and context of the phenomenon that this study aims to examine and to frame the research questions which have guided the research. This section will present brief summaries of the chapters into which this thesis has been organised.

Chapter 2 presents an abridged historical account of GZ's inception and a conceptual structure at its initial stages that helped determine the areas of focus of this study. Chapter 3 presents the existing literature on the themes that are relevant to the context of the study in relation to science and society, knowledge-producing cultures in science, emerging challenges in data-intensive science and, finally, the emerging concepts in the field. Since the empirical studies directly relevant to this research are sparse at best, the aim is to present literature that helps position this study in the existing areas of research that have been deemed to be most relevant to the work taken forward in GZ.

Chapter 4 presents the process of research design. Since the main research question is exploratory in nature, the design selected had to account for complexity and uncertainty, requiring the flexibility to capture data that may arise during unprecedented events. As this is a study of evolving practices in a context that has not been empirically examined before, as far as my knowledge of published research extends, an ethnographic approach was selected as being most suitable for studying the work of those responsible for designing and developing GZ as GZ itself unfolded and revealed itself to them. The methods employed in collecting data from key sources are discussed and the process of data collection and analyses are then explained. Finally, the ethical issues relating to the nature of inquiry and data management are considered.

Following the background of GZ presented in chapter 1, chapter 5, presents an analytical account of the how GZ was organised into different teams with different responsibilities at the beginning of the data gathering phase. A second conceptual framework, co-created with the main participant of the study, is presented to allow a systemic view of the various practices and processes that comprise the GZ system.

Chapter 6, the first of the findings chapters, examines the ways in which public participation in the process of scientific inquiry demanded changes to existing stages of the translation of scientific objects when dealing with Big Data. The focus in this chapter is primarily on the influence of public volunteers on the scientific practices of the GZ scientists.

The emerging practices undertaken by the GZ scientists that support citizen science, which might not contribute directly to the scientific output, is presented in chapter 7 in order to contextualize the discussion of expertise in the chapter that follows. Chapter 8 examines the development of expertise in the domain of managing citizen science projects, by focus on the on the two principal participants in the study, who are responsible for all management decisions for GZ. By the end of of the fieldwork, they had been recognised as experts in building citizen science projects.

Lastly, chapter 9 will begin by ensuring that that research questions presented in chapter 1 are clearly answered by synthesising and consolidating the findings of this study. It is hoped that this study will contribute to knowledge in the following ways: a) by examining the knowledge production system of the most successful citizen science undertaking yet; b) reconceputalising the role of the non-professional public in science; c) identifying translation as a critical practice in citizen science; and lastly, d) following the development of expertise required to run citizen science projects.

Chapter 2

The Galaxy Zoo: A Brief History

The aim of this brief chapter is to present a historical background from the inception of Galaxy Zoo (GZ) and construct a conceptual framework of it constituent elements of GZ's knowledge production system. The conceptual framework of GZ that I have formulated is the most recent representation of the how it operates and the practices, interactions, actors, communities and tools involved. The formulation (shown in Figure 2 A) also assumes certain boundaries and categories of expertise and intelligence to be located within the structure of GZ. The activities that contribute to GZ are also indicated where there are complex interactions between the unfolding knowledge objects that are being worked on, the scientists, the non-scientists and the tools they use. These features are regarded as inherently interlinked through being brought together in activities in the practices of citizen science. This model (Figure 2A) was presented to several members of the GZ project and was recognised as a useful and original mapping of the evolving structure.

The Principal Investigator (PI)[5] of GZ arrived as a post doctoral researcher at the Department of Astrophysics in 2007 to work on Astrochemistry in distant galaxies, when his colleagues came upon a problem that required them to have thousands of classification of galaxies. The two ideas that were considered were: a) setting up a laptop in the department for passersby to help; and

[5] The PI and TL of GZ believed that their anonymity was impossible to achieve if any references to GZ were made and they granted permission to not be anonymised. However all other participants in the study are anonymised.

b) putting the problem up on the web to get more publicity and wider assistance from the public.

Inspired by Stardust @Home, an online project which took the latter approach,, within a few weeks, they had a collection of academics interested in that option. However, none of them knew how to build a website that would allow them to run the project. The PI called in a favour from a software developer in Northern Ireland, who volunteered to build the site. At this point, the interested researchers were all early career academics. The situation changed when a senior scientist from an American university got involved and gave some direction to their efforts. After designing the simplest useful task for the lay-volunteers, the site was launched on July 11, 2007, with an expected time scale for the project estimated at five years, and a few thousand volunteer members of the public volunteering, who later became known as the 'Zooites'.

The classification target of three to four years exceeded within a few days with the help of the volunteers. The GZ team recognised this to be the turning point. E-mails from volunteers were overwhelming enough for Google to block the GZ team as spammers, as five members were simultaneously sending out e-mail replies. In response to this unexpected situation, a forum was quickly built, thereby establishing an open platform for communication between the UK-based scientists and the public in December 2007.

The continued success of GZ was not without challenges as the team encountered a major database problem, which was to be rectified only by the senior scientist involved. The PI pointed out in an early interview that capturing his interest early on proved to be critical for the project. There were also problems with classifications that had been tampered with by some of the volunteers and therefore required ‘cleaning’. The cleaning process became the first major data analysis exercise for GZ and led to a paper acknowledging the bias in public classifications. The paper attracted researchers in the field, who became interested in collaborating. By early 2008, there were 20-30 scientists interested in collaborating with GZ. Until this point, the project was being run from Oxford on volunteered time and resources. It was only when the PI was approached by another academic with an offer to help, after hearing a talk on GZ, that a grant was gained to sustain the PI’s position financially and to hire a Technical Lead (TL) for GZ. As the PI assumed the full-time role of project manager, three divisions with distinctive responsibilities were formed.

By the time I started gathering data, GZ was in its third iteration. Table 2.1 presents a summarised timescale of the three version of GZ and the number of classifications received for each project.

Version	Time of Launch	Classifications
Galaxy Zoo	July 2007	50 million
Galaxy Zoo 2	February 2009	60 million
Galaxy Zoo Hubble	April 2010-September 2012	40 million

Table 2.1 Versions of Galaxy Zoo Since 2007

It should be noted that following GZ’s initial success, the team was approached by scientists from other fields to build similar projects for their scientific research.

Their work expanded under the umbrella brand of 'The Zooniverse', which currently hosts 24 projects with five retired projects[6] categorised under the fields of Space (Astronomy), Nature, Climate Science, Humanities and Biology. The two key informants in the present study, the PI and and the TL in GZ, undertook the responsibility for other Zooniverse projects with different groups of science teams. Therefore, references to Zooniverse were determined to be inevitable when discussing GZ's development and the expertise involved in taking it forward.

2.1 The Initial Structural Components and Processes of Galaxy Zoo

By employing a graphical representation of the conceptual framework of the initial stages of GZ (Figure 2 A), this section aims to elucidate how GZ functioned as a citizen science project. The representation was constructed by bringing together primary data provided by the PI and secondary data from the public domain of the GZ website, and was verified by the PI and the TL of the project. This is the first conceptual map constructed with the objective to understand inner workings of the GZ and the outer context in which operates; as a conceptual tool to inform the methodology and methods of data gathering and analysis. The knowledge production cycle is as follows. As the data (images of galaxies) are collected via a telescope (Hubble in the most recent version of GZ), they are transmitted into a database. Initially, GZ obtained its images from the Sloan Digital Sky Survey (SDSS[7]) which was publicly available through its website. The scientists working with GZ then edited the images as required by the research question being

[6] The List of Zooniverse Projects https://www.zooniverse.org/projects
[7] Sloan Digital Sky Survey http:// www.sdss.org

asked, and the edited images were placed on the GZ website for the volunteers to classify. On several occasions, there were lengthy discussions on the team mailing list regarding the colours to be used and highlighted, and the possible biases that may arise. There was a range of expertise involved in the image processing stage, which shall be described in later sections.

Once the images were on the website, the volunteers classified those galaxies or Astronomical objects, and once the required number of classifications were reached, the data was sent to a database accessed by the Science Team. This database of classifications served as material for further analysis for the scientists in GZ, which led to production of scientific papers. Members of the public were initially provided with a tutorial to familiarise them with classifying the galaxies, and for each galaxy, 30-40 classifications were taken to reduce the chances of error on the part of the volunteers. The volunteers were provided with a forum as a platform where they posted their inquiries or pointed out anomalies and unusual objects to the volunteer community and the GZ scientists. Initially, some of the scientists involved with GZ actively participated in answering those queries, the duty of which was later transferred to some of the most dedicated and active volunteers who assumed the role of moderators.

The database of classified galaxies was also available to the wider scientific community on request. However, the release and sharing of data was always negotiated within the team prior to allowing access outside the GZ team. Therefore, non-GZ scientists could also benefit from GZ's datasets once they were released, and scientific papers have been published using that data. The

GZ team tried to keep track of the papers published using their datasets, and that information was shared through the team mailing list. keeping track was also important in order to monitor whether the dataset was utilised correctly and that GZ was appropriately credited. There have been instances where some GZ scientists needed to correct the description of GZ published by non-GZ scientists in academic papers and the interpretation of the dataset; in order to safeguard the image of GZ as a valid public dataset, aggregated after data reduction, in the field of Astronomy.

With scientific papers published by researchers within and outside the GZ team, a contribution is made to the disciplinary knowledge of Astronomy as an established scientific field of inquiry, hence completing the cyclical process of production of knowledge. The newly produced knowledge can then be communicated to non-scientists or the wider public who may or may not have helped create it. (P.T.O.)

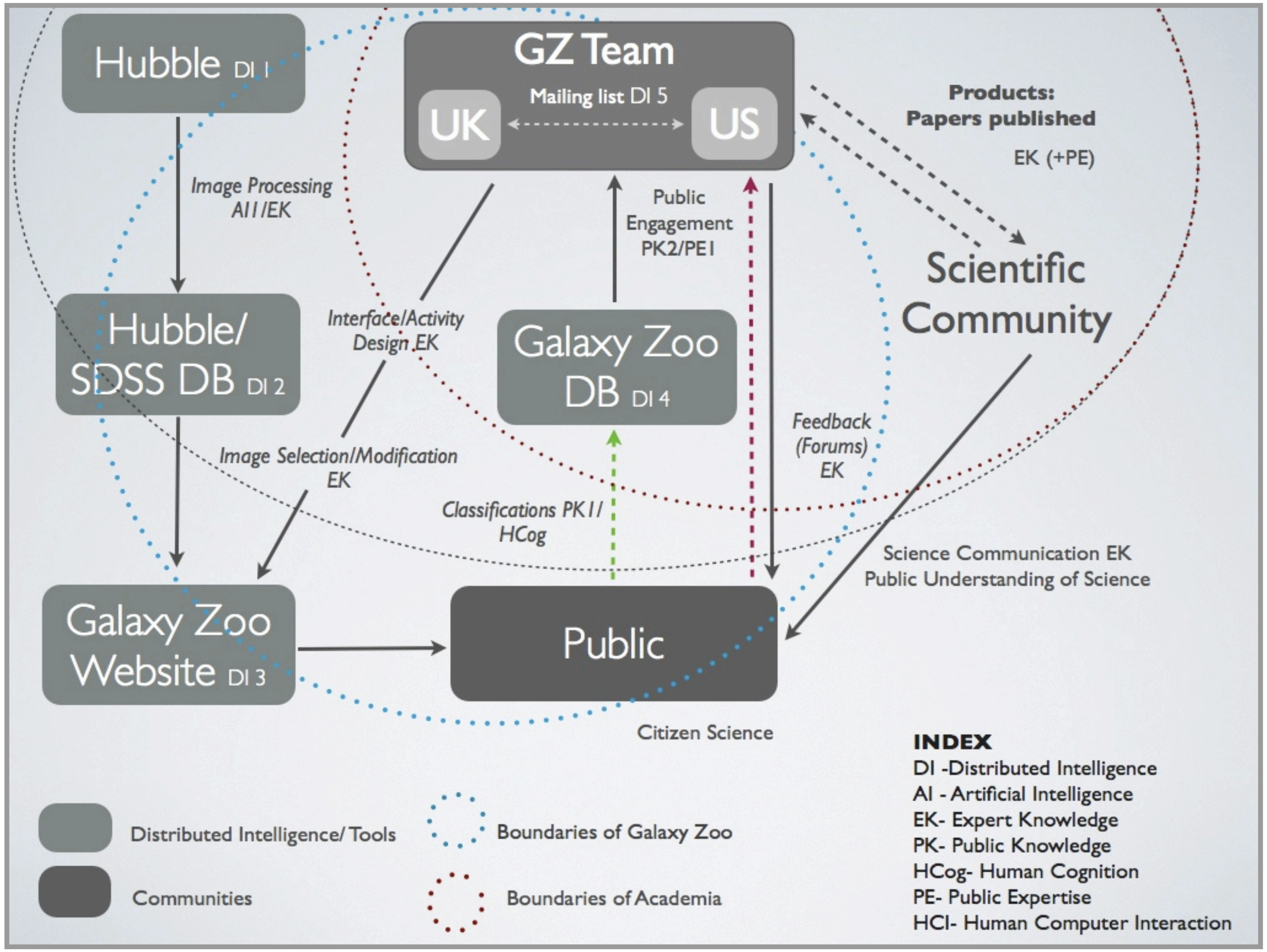


Figure 2 A: An Initial Conceptual Framework of Galaxy Zoo

As the representation illustrates (Figure 2 A), at each step it is assumed that different kinds of activities (tasks), expertise, technologies, and capabilities (skills) are operating in interaction with each other to facilitate the knowledge production function of GZ. The processes outlined so far characterise the emergent practices of citizen science as observed in GZ. These practices appear to call forth different kinds of expertise. Conceptions of expertise are therefore now discussed briefly in the following section 2.1.1 and are elaborated later in sections 3.2.6 in the context of Science and Society and different types of expertise relevant to this study are discussed in section 8.1.

The preliminary framework presented as Figure 2A, represents conceptualisations of the new tasks and capabilities facing the scientists engaged with citizen science, which create new demands on the different kinds of expertise they require. I would argue that GZ is characterised by the ordering and interplay of these different kinds of tasks, practices, expertise and capabilities which require GZ scientists to operate beyond the boundaries of academic science. The framework in Figure 2A draws on the following concepts: Pea's (1997) idea of social and material dimensions being vital in understanding how intelligence is accomplished as humans share the cognitive load through the use of artefacts; referred to as distributed intelligence. Similarly, Hutchins' (1991) idea that intelligence is not only a characteristic of human cognition, but is distributed across tools, artifacts, and rules that reduce cognitive loads on humans. Artificial Intelligence (AI) refers to the technologies that have the capability of performing tasks based on human behaviour taking into account the environment, rather than simply being programmed to function regardless of externalities (Brooks, 2001). In the context of GZ, AI refers to the use of robotic telescopes and machine learning techniques that used classification behaviour from the volunteers to improve the classification algorithms. Expert knowledge is this context refers to scientific expertise attributed to scientists which has been acquired through their training and experience in the profession of being scientists. Human cognition refers to the capability of humans to process and analyse data provided by GZ without any prior formal training, especially in classifying Astronomical objects and identifying anomalies that the volunteers were asked to carry out. The cognitive capabilities of the volunteers may be regarded as a part of distributed cognition present in GZ's knowledge production system. Human computer interaction refers to the use of computers in performing tasks for GZ that lead to scientific knowledge, and combines the distributed intelligence of human cognition and computers.

These elements in the GZ's knowledge production system are not mutually exclusive but work synergistically with each other and can be brought together under the theoretical framework of 'Collective Intelligence' (CI), put forward by Malone *et al*. (2009). They suggest that although the idea of communities working together towards a goal has been studied previously, what is needed now is a focus on how humans and technologies act more intelligently collectively than as separate entities. The framework as shown in Figure 2 A, therefore not only accounts for the participants of GZ, but also the tools and technologies that play an integral role in the success of GZ as a citizen science project. These entities operate in varying combinations at different stages of knowledge production within GZ.

2.1.1 Conceptions of Expertise: A Brief Introduction

Nowotny (2003) suggests that never before has expertise been simultaneously so indispensable and heavily contested. "The question of whose knowledge is to be recognised, translated and incorporated into action has been exacerbated under the pressure of democratisation. It has received additional exposure under the constant scrutiny of the mass media" Nowotny (2003:152). She is of the view that experts are often under the pressure of having to act as if they know the answers and are required to synthesise all the knowledge available to them and transgress the boundaries of their own domain of speciality and the limits of their own knowledge. Therefore, the notion of expertise is a problematic one.

Ericsson and Smith suggest that"[t]he study of expertise seeks to understand the account for what distinguishes outstanding individuals in a domain from less outstanding individuals in that domain, as well as from people in general" (1991:2).

Chi (2011) offers a brief summary of the conceptions of expertise that have dominated the study of expertise as an area of inquiry. She argues that the initial approach in studying expertise held the assumption that the the difference in performance of experts and novices could be attributed to super search strategies employed by the experts. As new empirical evidence began emerging, the assumption was replaced by the differences in the structure of knowledge of novices and experts. Finally, it was demonstrated that differences in the structure of knowledge led to differences in problem representation by experts and novices; And the representation corresponded to more efficient and more correct solutions by experts. While the representation of problems offers many insights, Chi (2011) highlights the challenge of teaching learners how to construct better knowledge structures to aid better problem representation.

While there have been other concepts in the study of expertise that have gained interested in the recent years, such as Ericsson *et al.*'s(1993:363) notion of deliberate practice, which attributes expert performance to "individuals' prolonged efforts to improve performance while negotiating motivational and external constraints." They propose that what was once believed to be innate talent of experts and elites are a result of deliberate practices.

Another notion that has received attention in expertise is the idea of group expertise. While it is established that groups often exhibit better performance than individuals (Barron, 2000; Pfister & Oehl, 2009; Schwartz, 1995), why group performance is higher is not understood very well (Chi, 2011). However, recent studies in the field of Collective Intelligence are beginning to contribute empirical evidence as to what might lead to higher performance in groups. For instance, .Woolley et al. (2010) reported that in two studies with 699 participants working in groups of two to five, the equality in conversational turn-taking, the average social sensitivity of members, and the proportion of females members in the group were correlated with higher group performance.

Another conception intriguing scholars studying expertise is 'adaptive expertise', which was introduced by Hatano and Inagaki (1986). They proposed by while routine experts can complete the tasks quickly and accurately, adaptive experts “have the the ability to apply meaningfully learned procedures flexibly and creatively.” (Hatano, 2003: xi) and with a deeper understanding of procedural skills, they can generalise their skills to non-routine problems. Therefore, implying that adaptive expertise might mean to acquire another perspective which Chi (2011) suggests considering adaptive expertise at all levels of expertise.

It worth keeping in mind that the understanding of what expertise and experts means vary greatly across the literature, for example, as Edwards (2005:60) suggests expertise as “a capacity to interpret the complexity of aspects of the world and have the wherewithal to respond to that complexity”, a view that “acknowledges that there are bodies of knowledge, ways of thinking, sets of values and expectations of behaviour that are associated with particular culturally derived forms of practice and that these features of practice are themselves open to change (Edwards A, 2005:60). While, Shanteau (1992:255) offers an understanding of experts as being “operationally defined as those who have been recognized within their profession as having the necessary skills and abilities to perform at the highest level” While this is not intended to presented comprehensive review of the conceptions of expertise, this section presented some of the historical trends and current notions relevant in the understanding the area of expertise

2.1.2 Conclusion

The aim of presenting the conceptual framework of this study in Figure 2 A was to briefly introduce the complexities identified within the context of GZ as an example of citizen science. Latour and Woolgar (1979) argue that even though our knowledge of the effects of science has increased, our understanding of the complex internal workings of scientific activity remains unexplored. Those internal process and structures may be what Knorr-Cetina (1999) defines as ‘epistemic cultures’; they are cultures that create and warrant knowledge.

Drawing from the major studies in this field, Latour and Woolgar (1979) have focused on the social construction of scientific knowledge by studying the processes by which scientists make sense of their observations. They propose that scientific practice involves confronting and negotiating with confusion to make order out of disorder. Ziman (2001) argues that the study of scientific processes of knowledge cannot be separated from the scientists themselves. Therefore, the problem of understanding scientific practice and the practices that support scientific inquiry is a Social Science question, which is what this study explores.

Chapter 3

A Review of the Literature

3.1 Introduction to the Literature Review

Citizen science is a recent phenomenon that has only gained momentum in the last few years and academic research in this area is gradually beginning to receive attention. The objective of this chapter is to contextualise the study within the existing literature, and in the process identify gaps in the relevant body of work, hence informing the nature and scope of this thesis. To the best of my knowledge, in-depth ethnographic research on the scientists who run and manage large-scale citizen science initiatives whether online or offline, have not been published to date. Since there are currently no empirical Social Science research studies on scientists involved in a large scale online citizen science project, this chapter introduces the principal concepts and research themes that were available at the beginning of this study and therefore, informed its design.

It should be noted that the emerging literature on the phenomenon of citizen science is heavily populated with studies on citizen scientists, that is their volunteers, rather than the professional scientists involved in citizen science. The reason for the one-sided attention of the studies in citizen science is unclear at this point. This is precisely the gap that this study aims to address by focusing the analytical lens on the scientists behind one citizen science initiative. My aim is to contribute towards an understanding of the phenomenon of citizen science by examining the practices of the scientists who conceived of the initiative that has become one of the most successful undertakings of its kind yet.

Since citizen science practices are under-examined phenomena, the opportunity to identify essential concepts and definitions of terms to guide this research is a challenge. In meeting this challenge this literature review draws on the various research fields of Science and Technology Studies (STS), Sociology of Science, Knowledge Management and Educational Research to understand the social phenomenon of the broader science practice of citizen science, its development, and complexities.

3.2 Science and Society: A Brief Summary

"Science is the belief in the ignorance of the experts."

- Richard Feynman, Physicist

There is an undeniable consensus that we live in a knowledge society (Drucker, 1992). These are societies built around knowledge-based work. Stehr (1994) suggests that scientific knowledge has increasingly become the most dominant form of knowledge. As a consequence, debates on the benefits and pitfalls of science for the public have been brewing over the last couple of decades. Recently the idea of social accountability of scientific research (Nowotny, 1999) has received much attention leading to increasing demands on the scientific community to produce results which have a direct and positive impact on the public. In the United Kingdom (UK), for example, the Higher Education for Funding Council for England (HEFCE) introduced the Research Excellence Framework as a development of the former Research Assessment Exercise, in order to determine funding distribution based on research quality, but adding a new benchmark for public accountability measured in terms of impact (HEFCE, 2010).

As the demands on science to be more transparent and become more accountable to society have increased, Latour and Woolgar (1979), Nowotny (2003) cautions that the current obsession with science can be counterproductive, without an understanding of the processes and the nature of science. To address this problem, there have been several initiatives by governments and the scientific community to increase public understanding of science. There is of course a long-standing and ongoing debate on what public understanding of science constitutes, and what level of understanding of science is considered fundamental for the public to engage in policy debates (Miller 2004; Nowtony 2003; Thomas and Durant 1987).

Governments across developed nations, especially the US, Canada, UK, Europe and Japan have been concerned with public attitudes towards science and have continuously invested in science communication programmes (National Science Foundation, 2004). However, such efforts aimed at public understanding of science are not limited to government agencies, as academic institutions are becoming more active in promoting understanding of scientific facts and research. The University of Oxford, for example, has an appointment titled 'The Simonyi Professorship Chair for the Public Understanding of Science' and the holder of this position

"[i]s to communicate science to the public without, in doing so, losing those elements of scholarship which constitute the essence of true understanding."

To explain further in Charles Simonyi's words:

"The goal is for the public to appreciate the order and beauty of the abstract and natural worlds which is there, hidden, layer-upon-layer. To share the excitement and awe that scientists feel when confronting the greatest of riddles. To have empathy for the scientists who are humbled by the grandeur of it all."

Reflecting on this statement raises many questions, such as: How is it possible, as a non-scientist, to truly understand and appreciate the natural world from a scientific world view and what kind of a role does the process of 'doing science' play or is it enough to just communicate without participation? It is at this point precisely that the role of engagement in understanding arises. If one is removed from active participation in the process of science, can one truly understand what scientists do? This question raises the further question of whether non-scientists can participate without the scientific expertise required. These questions are relevant for understanding the recent interest in the role of the public-intellectual and professional scientists in their role in knowledge societies (Collins 2002; Fuller 2004; Nowotny 1999).

3.2.1 Public and Science: Understanding, Engagement and Participation

Public engagement in science, takes us to the heart of the debate around what engagement entails and what value it might hold for science. The debate is important because it contextualises the research problem in the idea of a knowledge society rather than seeing it as an isolated phenomenon in the world of science.

The idea of engaging the public in scientific and technological changes has become popular in the arena of policy making in Europe since the late 1990s (Hagendijk, 2004). Jasanoff (2003a, 2003b), nonetheless argues that as essential as participation of the public is to a democracy, its motives and mechanisms should be carefully examined. As public participation in terms of understanding, interest and engagement is promoted and heavily funded by governments and educational organisations (Miller, 2004), the question that needs to be addressed is - why should scientists be interested in it? Poliakoff and Webb (2007) suggest that there is a noteworthy discrepancy between the scientific findings and their coverage by the media. This suggestion was supported by a survey of members of the public, which showed that seven out of ten participants believed that media sensationalises science (Wellcome Trust, 2000).

A report presented by the UK House of Lords Select Committee on Science and Technology in 2000, officially acknowledged public mistrust in science as a major concern. Scientists can no longer ignore the fact that by engaging with the public they can change negative perceptions about scientists (Hughes, 2001), which could lead to more public support for scientific activities and make science more enjoyable for the public (Greenwood and Riordan, 2001). The increasing need for scientists to present their public engagement plans when applying for research grants (Pearson, 2001), is an indication that funding councils and agencies have recognised public participation in science as an issue of importance. Nonetheless, there is resistance among scientists, as reflected in a Royal Society survey of 1485 scientists in 2006 which revealed that majority (64%) of them wanted to spend more time on research instead of engaging with the public and 20% believed that scientists who did engage in such activities were rated less highly by their peers. These results indicate the cultural conditions and barriers that exist within the scientific community and which perhaps reflect the perceptions of the public held in that community.

The scientific community is, however, not without members who actively engage with the public with varying motivations. These include Neil deGrasse Tyson in Astronomy, Richard Dawkins in Evolutionary Biology, Noam Chomsky in Linguistics, Stephen Hawking in Theoretical Physics, as particularly popular scientists who are also public figures. However, the motives for such efforts do not pass without scrutiny. For example, Miller (2003) contests, that while popular science books may appear to incite interest in science in the general public, they reinforce the dominant beliefs that the scientists want to promote, which remain unchallenged.

Although some scientists are skeptical about the value of public understanding to science, Bond et al. (2007) contend that public understanding in the age of technological advancement could possibly provide the environment that is supportive of another Kuhnian (1996) ‘scientific revolution’. As limited by their relevant expertise as the public may be, Nowtony (2003) has also argued that the public does have a role to play in scientific progress through policy, as boundaries become more fluid and science can no longer operate in isolation from other aspects of society. In her later arguments, she suggests STS have widened the notion of ‘public understanding of science’ by including the public more in the process of democratic decision making and by allowing the public a more pluralistic identity in liberal democracies (Nowotny, 2014).

In presenting relatively early and enthusiastic arguments in support of public understanding of science, Thomas and Durant (1987) approached the debate on what is meant by public understanding of science in terms of scientific literacy. They began by describing the varying notions of what is meant by 'public' and 'understanding', which are still not very often found in studies in this area. The lack of a clear definition or a common understanding of fundamental terms and the obsession over the attributes of scientific literacy, is what they argue have led scientists to overlook what is most important; how to relate to the public.

3.2.2 Science and Society: A Brief Debate

Bauer (2009) presents the evolution of paradigms of public understanding of science, which, he argues, reflects the dominance of the public deficit model up until the 1990s. That dominance was later reversed so that currently science is being questioned on its absolute authority and power of persuasion over the public as a result of crises relating to public trust and confidence in science. Although, Bauer's model conceptualises the relationship between science and the public as being purely chronological to reflect dominant trends, it can be argued that all of these paradigms exist simultaneously in the current context (2007). Bauer is particularly critical of the prevalence of the deficit model and contends that

"[a] critical public is an asset rather than a problem for the future of science" (Bauer 2008, 2009) as science increasingly operates within the commercial sphere.

To support his argument, Bauer (2007) calls on Fleck (1935). Fleck offers a conceptualisation of the relationship between science and the public which he represents as concentric spheres of science with its esoteric core extending into an exoteric periphery, where scientists not only seek social legitimacy, but also epistemic reassurance. Fleck agues that public communication is the "'elixir of life of science'' (As quoted in Bauer 2009: 237), adding that the opinion of science held by the public is not just an epiphenomenon of scientific activity, but is critical to the continuity of science. Although studies have shown that scientists are apprehensive of colleagues who engage in the public sphere (e.g. Powell and Colin, 2008). For Bauer (2009) the relationship between science and society is not solely about the distance between them, but it is also a question of quality. It is the quality of the relationship, which has received scant attention in the literature on public understanding of science, a discussion that is inevitable in the context of citizen science.

Open systems of knowledge especially in the arenas of science, have been discussed by Nowotny (2000), who also argues that contemporary scientific communities are no longer held together by paradigms (Kuhn, 1962), but by tools (e.g. methods, computers, modeling and visualisation techniques). She proposes that the epistemological core of science may:

> "[b]e neither immutable nor empty, but crowded by different and shifting scientific practices, techniques, norms and values. The production of knowledge then can no longer be clearly situated between academics and markets, but has to be reconceptualised as taking place beyond these categories - in a new space that grows out of the contextualisation and which we call the 'agora'." (Nowotny 2000: 194)

Nowotny's 'agora' can be compared with the notion of 'ba' put forward by Nonaka et al. (2000) which are spaces within the interaction of the collective, individual and the mental. Both analogies can be explored in relation to how the Internet provides that space for the public to participate in science and the democratic decision-making process. It can be argued that citizen science, being new spaces for new practices are held together by common tools and knowledge. Nowotny (2000) and Bauer (2009) both suggest that in understanding the social and the scientific as a co-evolutionary process requires focusing on their interdependent influence, which is where the existing empirical studies are lacking. By focusing on the practices of the scientists in citizen science, this study aims to address this gap and contribute to a more in depth understanding of the social and the scientific.

3.2.3 The Problems in Defining 'the Public'

The terms 'citizens', 'public', 'society' are used almost interchangeably in the literature and numerous papers have been published with the assumption that there is no need to define the 'public' (Ferber et al. 2007, Miller 2004, Rooker 2001). What is evident is that there is science and the scientists and then there are the rest who are non-scientists. Therefore, perhaps the public refers to non-scientists simply by the logic of exclusion. However, complications arise when the role of government and media are factored into the equation as the networks and relationships run in multiple directions as boundaries overlap and paths intersect (Nowotny, 2014). This complexity in defining relationships between citizens and science and technology is at the heart of contemporary society, where the public is usually presented as being passive (Irwin, 2006). Therefore, for the purpose of this study, the term public includes anyone who does not formally belong to the scientific community and has little or no expertise based on their experience with science.

3.2.4 Science and the Public: Definitional and Conceptual Challenges

In the special issue marking the 20th year of the journal on Public Understanding of Science (PUS), Nowotny (2014) characterises the original models of PUS as '"overly restrictive" and a "mistaken notion" (p. 17), which has since been widened to include a pluralistic conceptualisation of the public in the process of democratic decision making. She urges scrutiny of the current forms of public engagement citing the crowdsourcing, citizen science and the open access movement to follow how citizens engage with new technologies and new media that opens up issues for science and democracy. Jasanoff (2014), Irwin (2014) and Bauer (2014) collectively call for a renewed examination of what was initially conceptualised as PUS and has been more recently framed as 'Public Engagement in Science' in order to address the increasing complexities in relationships between citizens, politics and science; which will be discussed in light of the findings of this study in the concluding chapter.

The need for the public to understand science today clearly echoes in the literature (e.g. Bond et al. 2007, Irwin 2006, 2004; Miller 2004; Rooker 2001; Ziman 2001) and too often the studies begin and end without clarifying what 'engagement' 'participation' and 'understanding' entails, having often used the terms synonymously. Hagendijk (2004) contends that this interchangeable use of those terms reflects an inherent assumption in the expression itself that the non-scientists appreciate scientific procedures and products. He also suggests that although research in the area of public engagement has gained rapid momentum in the last two decades, perspectives on the nature of public understanding differ, as does the idea of competent scientific citizenship (Irwin 1995, 2001).

Most commonly, public understanding still appears to be understood and explained in terms of scientific literacy as a measurable indicator (Miller 1989). Miller (2004) proposes a working definition for public understanding of science with the benchmark being the ability of citizens to understand the science section of the New York Times, which raises issues such as relevance bounded by national context and agendas that are defined by a commercial corporation. From a science communication point of view, Burns et al. define PUS as:

"[t]he use of appropriate skills, media, activities and dialogue to produce one of the more of the following responses to science: awareness, enjoyment, interest, opinion-forming and understanding" (2004: 183).

An example of how the terms vary across the literature is Poliakoff and Webb's (2007) use of public engagement as being scientific communication that engages audience outside academia.

3.2.5 Public Engagement: A Typology

There appears to be a significant lack in consensus as to what public understanding definitionally or operationally entails, which is also a result of lack of theorising the phenomenon. The discussions are therefore frequently suffer with the lack of common definitions of key concepts. Rowe and Frewer (2005) contribute by presenting a typology of mechanism of public engagement in Figure 3 A.

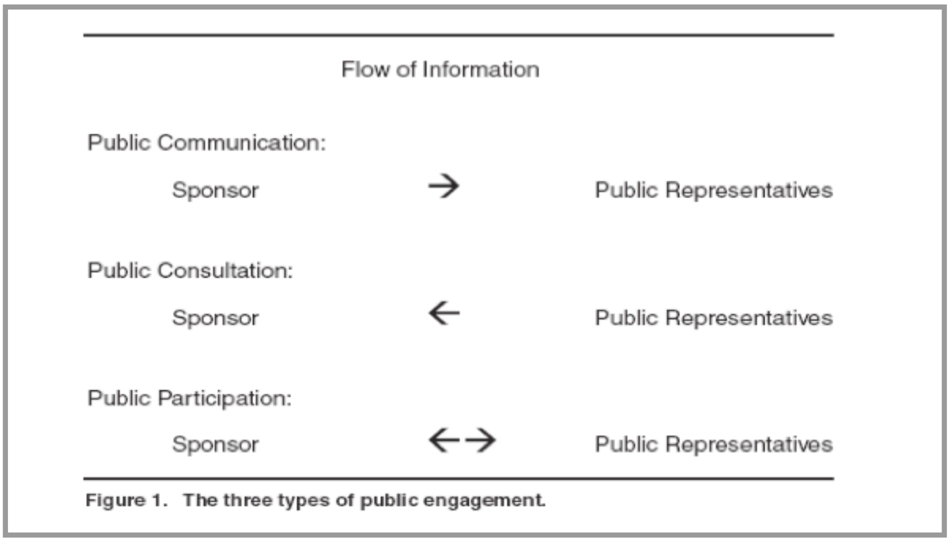


Figure 1. The three types of public engagement.

Figure 3 A: Typology of Public Engagement (Rowe and Frewer, 2005: 255)

Rowe and Frewer (2005) begin by explaining that public participation in their article relates to the public involvement in policy affairs and decision making. They admit that the key concepts in this area lack clear definitions and conceptual formulations which may have implications for engaging the public in science. Their typology is essentially based on conceptual differences in terms used for different ways in which the public is involved in decisions regarding public policy.

Public communication, as shown in Figure 3A, is unidirectional, whereby the information flows from the sponsors to public representatives. In this context, a sponsor is referred to as the party that commissions the initiative to engage the public, for example a governmental agency. In certain instances, representatives of the public may also be sponsors[8]. Feedback from the public is not sought in this mode and therefore, mechanisms to support the public in making attempts are absent. This mode of public communication may relate to generally what is referred to as public understanding of science (Miller 2004) with scientific literacy being a general indicator, where the public is the passive recipient of scientific knowledge (Ziman, 2004).

Public consultation occurs when the sponsor initiate the process of obtaining information or feedback from the public. Although, the exchange of information is bi-directional, there is a lack of formal dialogue between the two parties.

Public participation entails a dialogue which may involve negotiations between the representatives of the sponsors and the members of the public. What differentiates this mode from the previous two is that the interaction between the two parties may result in transformation of opinions and ideas held by both or either parties through acts of negotiation.

All of the three concepts in combination are then referred to as public engagement in this typology, which creates a definitional distance between the terms communication, consultation, participation and the one that encapsulates all three: engagement.

[8] Sponsor refers to the party initiating the engagement activity and are usually governmental agencies. However, representatives of the public may be involved.

Rowe and Frewer (2005) suggest that the need to theorise public engagement lies in understanding the key terms in participation, which has direct implications for conducting research and implementing such initiatives in practice. Their typology, however, only accounts for involving the public in policy making based on scientific facts or theories that have been produced by the scientific community and does not account for citizen science. They also acknowledge the lack of empirical and theoretical research required to examine contingencies and develop different mechanisms for characterising the initiatives, a contribution that this study makes.

3.2.6 The Role of the Public in Science: A Matter of Expertise

In their seminal thesis, Gibbons et al. (1994) suggest that new scientific knowledge must be socially robust, arguing for transparency and the co-production of scientific knowledge through closer links with scientists who work outside the university, as well as a stronger focus on public involvement more generally. Gibbons (1999) contends that the role of technical and scientific expertise in societies with a high-reliance on industrialisation is undergoing changes as expertise becomes fragmented between institutions and professions. He proposes collective narratives that account for the uncertainty and complexity arising from such a fragmentation of expertise:

"Since expertise now has to bring together knowledge that is itself distributed, contextualized and heterogeneous, it cannot arise at one specific site, or out of the views of one scientific discipline or group of highly respected researchers. Rather it must emerge from bringing together the many different 'knowledge dimensions' involved. Its authority depends on the way in which such a collective group is linked, often in a self organized way." (Gibbons 1999:C83)

Nowotny (2003) too has indicated the need for a new agreement between science and society, with science being expected to produce reliable knowledge and society bestowing their trust on science with that responsibility. However, the need for this fundamental involvement does not necessarily imply readiness on the part of the public. Hence, the critical question of possible ways of contributing to science arises. She argues that for the expertise to move from a quest for reliable knowledge to a socially robust one, requires a certain transgressive competence, where new kinds of expertise are needed (Nowtony, 2000). Kerr et al. (2007) emphasise that the context in which expertise is institutionalised has to be examined and also suggest that considerable effort needs to be directed towards developing a dialogue between professionals/ experts and the public.

While offering the distinction between broad categories of expertise, Collins and Evans (2002) question the value afforded to scientists' experience and knowledge over that of others in that context. They suggest that the most important boundaries in those circumstances may be those between different kinds of experts and groups of specialists rather than between experts and non-experts. They propose ways in which different kinds of expertise should contribute towards decision making. Therefore, the debate on expertise in relation to public involvement in scientific enquiry becomes critical in understanding the phenomenon in question. This is precisely where Collins and Evans' (2002) idea of different forms of expertise, arising in their research programme on the Sociology of Science, becomes relevant in the context of public engagement in science.

As illustrated in Figure 3B, Collins and Evans (2002) contend that the understanding of the divide between experts and non-experts has evolved in three stages:

a) Wave one, where the distinction is clearly drawn between experts and non-experts.
b) Wave two, where the boundaries between experts (scientists) and the public disappeared as democratisation of policy making surged in popularity.
c) Wave three, whereby the boundaries are now no longer between experts and the rest, but between different kinds of experts and the rest.

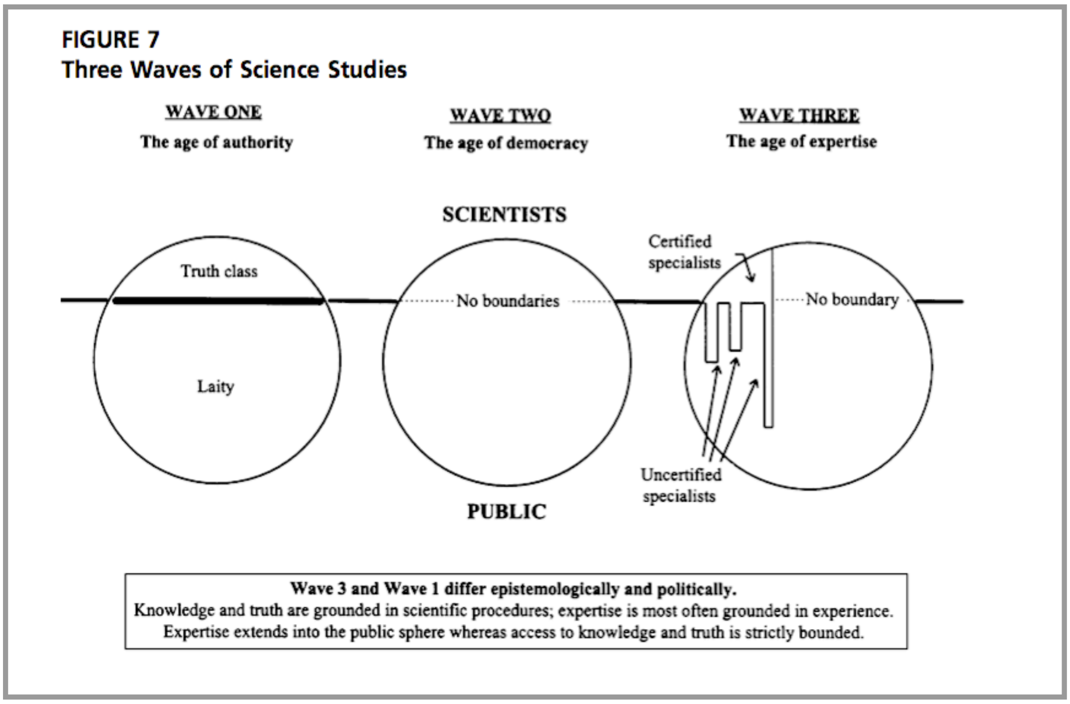


Figure 3B: Waves of Science Studies (Collins and Evans, 2002: 250)

It is in the paradigm of Wave Three, that they draw from their experience as sociologists of science, theorising three kinds of expertise:

i) No Expertise: This is a level of scientific expertise of the sociologists who sets out for fieldwork, which is not sufficient to carry out any level of analysis of the scientific work;

ii) Interactional Expertise: This degree of expertise of the researcher that affords interactions with the participants of the study and conduct sociological analysis of the phenomenon studied.

iii) Contributory Expertise: This degree of expertise allows for one to make contributions to the scientific discipline that is being researched.

The authors contend that interactional expertise does not lead to one having contributory expertise and the former is not necessary in order to achieve the latter. By using the example of Wynne's (1996) study of the relationship between sheep farmers and scientists after the Chernobyl disaster led to contamination of Cumbrian fells, they illustrate how the farmers did not require contributory expertise in science to share their knowledge of the local ecology as long as the scientists had interactional expertise to relate to the farmers.

The relationships between these kinds of expertise are one of the foci of inquiry in the study of Galaxy Zoo (GZ). As GZ is modeled around the idea of public participation in scientific inquiry of galaxy formations, amateurs have not only exponentially reduced the speed and time taken to classify objects of scientific scrutiny, but they have also made discoveries. Consequently, non-scientists and different kinds of scientists engage through mediated environments to gain experience and perhaps develop expertise.

With an increasing trend of public involvement in science, in spite of lack of clarity over the nature and role of public expertise, Cohen (1997) has argued that much of what he believes to be the zeitgeist of the last two decades supports citizen science, on the assumption that it will increase understanding of the scientific process. It is safe to state that the non-scientists cannot be expected to replace scientists in any manner, the extent to which non-scientists can learn, be trained and contribute towards scientific inquiry is yet to be explored. A citizen science project led by the Cornell Lab of Ornithology was one of the first of its kind to involve 700 participants, the results of which suggested that although involvement did not lead to scientific thinking, it did provide the volunteers with an avenue to engage in the habit of thinking about scientific phenomena. One of the interesting findings was that methods of scientific observation and their instructions had to be designed for a lay observer, pointing to how citizen science may impact on scientific processes and the expertise among scientists that is needed to make these adaptations.

3.2.7 Engagement and Beyond

Citizen science offers a new mode of coproduction which takes engagement right at the core of the scientific process. It invites the public into science in contrast to communicating to the public about science. The definitions and activities of citizen science may vary across the myriad of initiatives which directly involve the public in the process of scientific inquiry as explored in chapter 1, but what emerges as common is that citizen science strives for the true meaning of coproduction of scientific knowledge. Citizen science as a mode calls for public engagement in science as opposed to public engagement with science. With growing enthusiasm for citizen science and crowdsourcinginitiatives, Nowotny (2014) brings to attention, the need for examining current forms of public engagement in science, which give rise to intricate issues ranging from open access to the inclusiveness arising from crowdsourcing supported by rapid social technologies and new media that shape political and social realities (Castells et al., 2009). Therefore, perhaps as helpful as the terminology of public engagement has been, citizen science as a phenomenon will open avenues and questions for conceptual and theoretical possibilities in understanding a more symbiotic relationship between science and society, with technology-enabled platforms for coproduction of knowledge that go beyond notions of engagement that has been conceptualised thus far. This is perhaps only the beginning of a whole different reconfiguration of our social and political reality that includes the public in the epistemic cultures of science.

3.3 Epistemic Cultures in Science

The notion of epistemic cultures proposed by Knorr-Cetina (1999) was briefly introduced in chapter 1 as a central concept to this study. In her description of epistemic cultures, epistemic objects and subjects are intricately bound within epistemic practice in scientific cultures. She defines epistemic cultures as

"cultures that create and warrant knowledge, and the premier knowledge institution throughout the world is still, science" (Knorr-Cetina, 1999:1).

She argues that culture is a permanent feature of the ways in which humans evolve, and that science is pursued by groups of specialists and experts who are separated by institutional boundaries that comprise all levels of education with varying characteristics that define scientific cultures. The diversity within scientific cultures in its epistemic machinery (machines of knowledge construction), is what she believes creates a disunity in science, which Hacking (1992), and Gallison and Stump (1996) have articulated. Using the notion of culture to denote the variable patterns and dynamics of expert practice, she asserts that a study of such practices refers to investigating scientists at work rather than to the history of ideas, the structure of theories or institutional settings in science.

Her concept of knowledge object is characterised by their distinctiveness from everyday objects, since they are defined by an unfolding open-ended structure. They have a signifying property that leads the epistemic subject to engage in the unfolding of the object. Although epistemic objects can be bounded in material form, they are constantly producing meaning and function. The idea of a knowledge object in Knorr-Cetina's (2007) work therefore focuses on creativity and motivation in professional practice, and provides evidence of the externalisation of learning as practitioners follow the objects and work on them. She argues that the complexities of these objects are found in their lack of completeness, which invites constant engagement on the part of the practitioner. These epistemic cultures are distinct from disciplinary cultures as they move away from being bounded by disciplines to form specific cultures of expertise employing what she calls epistemic machinery or tools, which have technical, symbolic and social dimensions of intricate expert systems. Therefore, these dimensions have to be studied in relation to each other; establishing the link between practices, knowledge objects and tools used for the production of scientific knowledge. She asserts the criticality of epistemic cultures and epistemic objects in understanding scientific cultures, as she contends that without an examination of expert-object relations, epistemic environments cannot be understood.

Knorr-Cetina's (1999, 2007, 2008) notion of epistemic object and technical object have proven to be critical in the STS research on experiments and laboratory studies. In the studies of e-Science in different disciplines, Wouters and Beaulieu (2006) among many others propose those notions might need revisiting in the context of e-Science, as new infrastructures become knowledge objects in their own right. However, notions of organisational perspectives might become more important here since the infrastructures span across varying scales that do not extend to a global scale. Empirical studies of organisational learning (Argote, Beckman, and Epple, 1990) have indicated that while they have the ability to learn, they can also forget, with loss of personnel further significantly contributing to that phenomenon (Carley, 1992). Including organisational perspectives in investigating epistemic cultures in e-Science may be an avenue that researchers can pursue. Such an approach might benefit from analysing the epistemic objects as pursued not only by individuals, but also by organisations; it may offer a systemic view which might directly benefit design, implementation and sustainability of new knowledge infrastructures for science in the fourth paradigm, which is discussed in further detail in section 3.4.

Latour and Woolgar (1979) argue that even though our knowledge of effects of science has increased, our understanding of the complex internal workings of scientific activity remains unexplored. This lack of understanding of the processes and cultures of scientific activity has, as we have seen, led to concerns about the trustworthiness of scientific knowledge. This skepticism has arisen, in part at least, as a result of what Whitely (1972) refers to as 'black boxism' in his early call for the scrutiny of scientific knowledge. Irwin (1995), in response to this lack of trust, argues that science should be self-critical in the current scenario of growing public skepticism regarding its connection to progress. Nonetheless, these gaps are being filled by social scientists, for example, those internal process and structures may be what Knorr-Cetina (1999) define as epistemic cultures; as cultures that create and warrant knowledge.

Hine (2006) observes that Social Studies of Science as a field has developed approaches to the exploration of emerging processes of knowledge production as they occur through spatial, material and social arrangements. These efforts in examining the inner workings of science have constituted a field, which has come to be known as the sociology of science, or scientific knowledge according to Collins (1983). In Britain, between 1969 and 1981, Collins argues, there were only six independent prominent contributors to this field. The field has subsequently grown, with Collins continuing to make significant contributions. Nonetheless, the empirical literature remains relatively thin, despite key contributions from for example Latour and Woolgar (1979) and Knorr-Cetina (1999).

Latour and Woolgar (1979) focused on the social construction of scientific knowledge by examining the processes by which scientists make sense of the observations they make. They propose that scientific practice involves confronting and negotiating with confusion to make order out of disorder. While they conducted a seminal anthropological study of how facts were constructed in a laboratory at the Salk Institute over a period of two years, Knorr-Cetina (1999) studied a group of physicists at CERN focusing on the machineries of knowledge creation, and Collins (1975) examined the replication of experiments in the field of Physics. Ziman (2001) argues that the study of scientific processes of knowledge cannot be separated from the scientists themselves. Therefore, the problem of understanding how scientific knowledge is produced is as much a social question as is the question of how scientists work together. While the seminal studies in STS set out to explore scientific practices, thereby unravelling the complexities of the laboratory settings, the present study is primarily concerned with practices that allow scientists to work with the public, rather than practices that follow a scientific protocol. Hence, the critique and examination of STS studies here is not directly relevant to the current thesis. However, what is of importance is that the scientific practices, the object of enquiry for STS scholars, and practices for science, the object of enquiry for this research study, characterise epistemic cultures. Therefore, there are two distinct categories of expertise involved- scientific expertise and expertise in activities that support science.

3.3.1 Practices in Epistemic Cultures

Before moving forward in exploring the emerging range of studies that examine the practices that inhabit the epistemic cultures of science in the current context of Big Data (or data deluge), the meaning of practice in this study should be clarified. Schatzki (2001:11) outlines the challenges inherent in defining the concept of practice as:

"Given this multiplicity of impulses, issues, and oppositions, it is not surprising that there is no unified practice approach. Most thinkers who theorize practices conceive of them, minimally, as arrays of activity. Not only, however, do their conceptions of activity and what connects activities vary, but some theorists define practices as the skills, or tacit knowledges and presuppositions, that underpin activities (e.g., Turner 1994; Dreyfus 1991). Most theorists, moreover, above all those in philosophy and the traditional social sciences, identify the activities involved as those of persons: practices are arrays of human activity."

Schatzki (2001) notes how all practice theorists acknowledge that activity depends on shared skills or understandings, but he also highlights that thinking about practice involves conflicting conceptions, intuitions and research strategies. Barnes (2001:25) similarly argues that central to social systems are "ongoing, self-reproducing arrays of shared practices" signifying practice as collective action. The notion of practice that has informed the present study is also collectively oriented and is Edwards A's (2010:7) characterisation of practices as

"[h]istorically accumulated, knowledge-laden, emotionally freighted and given direction by what is valued by those who inhabit them."

This definition pays attention to both the historical origins of a practice and the motives and purposes of those who create practices through their actions within them.

As explained in chapter 1, the research question formulated for this study is concerned with emergent practices within the epistemic cultures of a citizen science initiative. Most of the empirical and theoretical literature on practice focuses on recurring or established sets of practices. The current study examines the practices that emerge as a result of scientists engaging in citizen science to produce scientific knowledge. In studying knowledge-producing social environments, the more precise notion of objectual practices proposed by Knorr-Cetina (2001) offers the most appropriate focus since it particularly attends to the pursuit of the knowledge objects that creates a new practice. To avoid considerable overlap of literature, this particular approach to practice is discussed in further detail in chapter 7.

In the present study, attention is given to practices in order to create a framing for an analysis of the expertise involved in running a citizen science project. Although the literature on the nature of established expertise and its acquisition is rich with theoretical and empirical work such as the contributions by Dreyfus et al. (1987), Dreyfus and Dreyfus (2005), Ericsson (1991), and Chi et al., (2014), the focus of this study is on new kinds of emerging expertise among professional scientists. Therefore, bringing together the notion of practice put forward by Edwards A (2010) along with the concept of objectual practice proposed by Knorr-Cetina (2001) creates a set of conceptual tools for examining the processes by which practices are shaped in relation to their historicity and the purposes values of the subjects who are in continuous pursuit of knowledge objects. Knorr-Cetina (2001) argues that the scrutiny of expert-object relationships is indispensable to an understanding of epistemic environments. As knowledge-centered work is characterised by a constant shift between routine procedures and new and creative responses, this consolidated notion of practice helps tackle the overarching research question, as it examines practice and expertise in epistemic cultures operating in the emerging field of Big Data citizen science.

In the following sections, the most recent challenge faced by scientific research disciplines, one that concerns the rapid creation and management of large amounts of data, will be described and examined. As a result, old practices are being modified and new practices emerge within the epistemic cultures of science, which provide an overview of the environment in which GZ operates.

3.4 The Fourth Paradigm: The Changing Landscape of Data-Intensive Science

The increasing prevalence of data-intensive computing in science was coined as the fourth paradigm of science by Gray (2007)[9] He suggests the following evolution of paradigms in science: i) that a thousand years ago, science was empirical as it was based on describing natural phenomena; ii) the last few hundred years saw the dominance of theoretical science with Kepler's Laws and Newton's Laws of Motion and so on; iii) the last few decades have been marked by an increase in computational modeling through simulations; and iv) the last few years have seen the unified approach of previous paradigms whereby data is captured or generated by instruments, processed by software, stored in databases and analysed by scientists using statistics and data management tools. He argues that the significant differences in methods and technologies between data-intensive and computational science are a warrant for the *fourth paradigm* in science. Among many challenges of the 'fourth paradigm', he discusses the prevalence of "terrible" data management tools (2009: xxiv); the kinds of changes in practices needed to allow scientists to manage large amounts of complex data at different stages of their research; and how scholarly

[9] Jim Gray was a pioneer in modern database technology.

communication would need to evolve through changes in peer review mechanisms and publication of data. Emphasising the need for an object-oriented method, which is a kind of database management system in which information is represented as objects that have attributes assigned to them, in defining objects and units in all fields, he warns of the endless debate on ontology and semantics that may follow. He illustrates his argument with a successful operational example of the Life Sciences Search Engine called Entrez.[1]

> "[c]reated by the National Center for Biotechnology Information for the NLM. Entrez allows searches across PubMed Central, which is the literature, but they also have phylogeny data, they have nucleotide sequences, they have protein sequences and their 3-D structures, and then they have GenBank. It is really a very impressive system. They have also built the PubChem database and a lot of other things. This is all an example of the data and the literature interoperating. You can be looking at an article, go to the gene data, follow the gene to the disease, go back to the literature, and so on." (2009: xxx)

Linked data and design of systems that enable modular connections or pathways to accessing and understanding data are what he sees as the way forward. Discussing the implications of developments such as these, Bell (2009) suggests that the pervasiveness and longevity of data in the 21st century is markedly distinct from previous eras of scientific investigation. Today not only do new instruments continually capture data, but computer models are generating worlds of information that are available for analysis and are becoming more open to public access. The permanence of data brings many challenges to science including the management of data and information as demonstrated by the identification of the emerging professionals referred as "data scientists" in a report published by the National Science Foundation (2005). These data scientists include professionals skilled in managing collections of digital data, such as Computer Scientists, database and software engineers, curators and annotators, archivists and librarians. The drive for developing data science is being pursued by national governments as reflected by the 2014 Budget announcement that allocated £42 million in funds to lead the UK in Big Data and algorithm research, with a total of £222 million allocated to science and engineering projects[10].

In the UK, the term “e-Science” is the prevalent label as it was introduced by the Director- General of Research Councils in 2001

“to encapsulate the technologies needed to support the collaborative, multidisciplinary research that was emerging in many fields of science” (Hey, 2005: vi).

The terms e-Research and e-Science have been used interchangeably in the literature depending on the origin of publication[11]. In 2002, a national level programme called ‘The UK e-Science Programme’ was initiated with the vision for creating computational infrastructure for scientific research, and later became known as a ‘Grid’ infrastructure or technologies (de Roure et al. 2003; Foster et al. 2001). Some of the important characteristics of e-Science collaborations were intended to be access to large databases, large-scale computing and visualisation technologies for individual scientists and institutions (Hey and Trefethen, 2002). As the term e-Research denotes, e-Science has permeated outside of Natural and Physical Sciences into the Social Sciences and the Humanities (Jirotka, 2013). E-Research has evolved into computing infrastructures designed for academics to be able to support research

[10] Source: https://www.gov.uk/government/news/plans-for-world-class-research-centre-in-the-uk
[11] The discussion on the distinction between e-Science and e-Research (which can be found in Jirotka *et al.* 2006) was not considered to be directly relevant to the review and has therefore, not been discussed further.

collaborations of distributed and multidisciplinary nature (Borgman, 2007). Nonetheless there is still work to be done on these systems. Employing the analogy of a 'data iceberg', Gray (2007) problematises the disproportionate amount of data that are curated or published compared to the amount of data collected and calls for tools that would support the entire research cycle, from data capture and curation to analysis and visualisation.

3.4.1 Knowledge infrastructures: A Response to the Fourth Paradigm

Edwards P. *et al*. (2013) suggest that in the last 20 years, systems of knowledge production, sharing and discussion have gone through considerable transformations which have been associated with the proliferation of Internet, social media, open source software that have changed the basic mechanics of knowledge production. They further state that as a result, new practices such as crowdsourcing, citizen science, open access and Massive Open Online Courses (MOOCs) have emerged, which have not only changed the way data is shared, but also the ways in which the public engages with the data. They contend that these phenomena have been influenced by deliberate strategies by policymakers, and funders have a role to play in the emergence of these developments. They cite the NSF's programmes such as Knowledge and Distributed Intelligence in the late 1990s, Information Technology Research in the early 2000s and Humans and Social Dynamics in the late 2000s that have propelled experimentation in modes of knowledge production and distribution.
Edwards P (2010:17) defines Knowledge Infrastructures (KI) as:

"[r]obust networks of people, artefacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds."

While this definition includes individuals, organisations, routines, shared norms and practices, Edwards P. *et al*. (2013) identify characteristics of KI as modular, multi-layered and indicating a 'rough-cut' nature, suggesting that are not fully coherent systems. They further elaborate KI's as being

"[e]cologies or complex adaptive systems; they consist of numerous systems, each with unique origins and goals, which are made to interoperate by means of standard, socket layers, social practices, norms, and individual behaviours that smooth out the connections among them. This adaptive process is continuous, as individual elements change and new ones are introduced - and it is not necessarily always successful" (Edwards P. *et al*., 2013: 5).

While significant changes in knowledge systems are occurring, they suggest caution against losing sight of questions about the complex processes by which older forms of knowledge organisations mutually adapt with those that are emerging.

As emerging forms of collaborative and collective discovery and knowledge production through crowdsourcing, wikis, shared scientific workflows and citizen science emerge across many disciplines (De Roure *et al*. 2011; Shilton 2009; Shirky 2009), Woolgar and Coopmans (2006) suggest that approaches in STS of e-Science are only in their initial stages. Woolgar and Coopmans (2006) are, nonetheless, optimistic about the potential that the field holds in carrying out in-depth examinations of the developments of notions of data, networks and accountability. They propose that the STS perspective of virtual witnessing is

"[a] key idea in understanding the early emergence of criteria of adequacy in experiments and demonstrations at the birth of modern science" (Woolgar and Coopmans, 2006: 2).

3.4.2 Evolving Practices in e-Science: Emerging Findings from Empirical Studies

This section discusses some of the studies that have examined the epistemic cultures of the emerging field of e-Science, and hence, contribute to the understanding of how practices within science are evolving while the use of computer and web technologies proliferate across scientific disciplines. For instance, Hine's (2006) examination of the adoption of a 'computerisation movement' in the branch of biology called systematics, revealed that while the community of practitioners in that field were hesitant in accepting or approving the idea completely, efforts being made towards computerisation itself added to a debate regarding the practices and goals of the discipline as a whole.

Using Knorr-Cetina's (1999) notion of discrete epistemic cultures in science, Wouters and Beaulieu (2006) explore the case of the academic field of women's studies through which they question whether the specific needs of the discipline are being overlooked as a result of the drive to implement e-Science. They argue that understanding what the disciplines require and the motives for driving changes towards a computational paradigm have to be questioned and examined. Their arguments usefully highlight the challenges presented by disciplinarity within a generalised notion of the deployment of knowledge infrastructures (Haythornthwaite, 2006; Merz 2006; Fry, 2006).

Vann and Bowker (2006), as an alternative, propose the term 'epistemic IT' to signify the information technologies used for scientific knowledge production. They present the vision of a kind of interdisciplinarity by exploring a case study about the production of an e-Science infrastructure in the US. They focus not only on the IT design and use, but also on the decisions made about skills, commitment and performance of the scientists in the initiative. Their study reveals differences in interests of the communities that built the infrastructure demanded by the grant, and the communities that had intellectual concerns over the discipline.

Through an ethnographic case study of the collaborative practices of Theoretical Particle physicists, Merz (2006) suggests that epistemic cultures vary depending on their adoption of communicational technologies, and therefore a unified phenomenon of e-Science adoption must not be expected. By embedding the use of communication technologies in practices bound by disciplines and building on Galison and Stump's (1996) notion of 'disunity in science', she suggests the term 'disunity of e-Science' to refer to the uneven development of the use and adoption of new knowledge infrastructures. She argues that Theoretical Particle Physics is a 'thinking science' (Merz and Knorr-Cetina, 1997), where building models and construction are what defines the work of these scientists, which predominantly consist of using writing materials, writing surfaces and computers for processing and which differentiates it from many other disciplines. Drawing on an empirical study, Merz (2006) also proposes that, since the use of digital infrastructures varies across scientific disciplines, successful models may not be transferable or exportable from one field to another. She posits that the specifics of embeddedness is related to how successfully an infrastructure can be exported.

Reflecting on studies of distributed, interdisciplinary and collaborative work in science teams, Haythornthwaite *et al*. (2006) suggests that working in such teams require that the tacit ways of working have to be made explicit, which is a considerable challenge. They illustrate with examples of changes in attitudes required to work across disciplinary boundaries, for example, citing a case in Systems Biology which not only brings together the fields of Engineering, Computer Science and Biology, but requires a shift in perspective from a single gene or protein to a more holistic approach to living organisms. This revelation is an important contribution to the field, as ten years earlier Star and Ruhlerder (1996) had found that system developers did not anticipate the need for assisting domain scientists with the transition in research culture, and that such lack of awareness led to difficulties in the adoption of infrastructures that were developed.

Differences in ways of working in areas which are close neighbours conceptually, also need to be recognised. Sundberg's (2010) interview-based study of Astrophysicists and Meteorologists in various career stages in Sweden found that although both are Physics-based disciplines that employ simulations and numerical experiments, they differ in many ways. For instance, while observing the practice of different sets of simulation skills in the two fields, she found that in meteorology, scientists often used simulation output with observed datasets, blurring the lines between the two kinds of data; whereas that particular practice was not observed amongst Astrophysicists.

The use of technologies and the behaviour related to more universal research processes, such as how a document is read, is also changing. Evans (2008) found that scientists exhibit more skim reading when reading a document online. For example, while searching for documents online rather than browsing through print materials, readers exhibit behaviours such as jumping from one link to another. Although these actions might enable keeping up to date with the prevailing developments, Evans (2008) warns that these kinds of behaviour "[a]ccelerate consensus and narrow the range of findings and ideas built upon" (Evans, 2008: 395).

In a study commissioned by the Institute of Physics in the UK, Meyer *et al.* (2011) examine the information-practices of scientists, focusing on areas such as information retrieval, information management, data analysis, dissemination, collaboration and generating new research questions. They surveyed and interviewed 79 participants with ages ranging from 22 to 73, with an average age of 39, 80% of whom had doctoral degrees. The disciplines they sampled include: Particles Physics, Gamma Ray Astrophysics, Nuclear Physics, Earth Science, Nanoscience and the Zooniverse as a citizen science-based research field. They characterise Particle Physics to be the field most studied by Social Scientists of the fields, contributing to their leading role in developing computational information technologies, including the World Wide Web, email and pre-print facilities such as arXiv. Gamma Ray Astrophysics is marked by unexpected phenomena that occur without warning and last only a few seconds, which requires quick alerting mechanisms for the bursts to be studied. Nuclear Physicists in the UK have been required to participate in international collaborations as a result of lack of facilities to conduct their research in the UK since 1993 (Meyer *et al*., 2011), and differs from the previously mentioned fields in the practical applications of its research in areas such as nuclear weapons and nuclear power. Earth Science, being an interdisciplinary field, combines several disciplines such as Climate Science, Geology, Geophysics, Seismology, Volcanology and Hydrology to name a few. There, Meyer *et al*. (2011) have found personal contacts to be the principal method of keeping up to date with new information. The use of social media appears to be as important a tool for research as it is in informing the public. Lastly, they distinguish Zooniverse as markedly different, as the scientists in this case, have to interact with the public to sustain the project and therefore their social media practices are most active.

Meyer *et al.*'s (2011) study reveals bespoke software and tool building as dominant practices within these sub-disciplines and software programming skills with data becoming a prerequisite for research studies. However, there are differences evident between the different sub-disciplinary fields. While only a few participants agreed that the new technology enabled new scientific questions, it was generally reported that due to the development of technology over the years, some questions that could not have been answered before can now be addressed. Email lists play an important role in communicating with peers and experts for Particle physicists and members of the Zooniverse team; while other participants continued to rely mostly on face-to-face communication. With the exception of the Zooniverse scientists, most of the participants were slow in adopting Web 2.0 technologies. The development of in-house software was also most prominent among Zooniverse scientists; whereas it was least likely among chemists. In relation to online research dissemination, only half of the participants did so regularly; but only a few never published their results online. The authors conclude that although computing has been pervasive in the academic fields of Physics, scientists are not uncritical users. Rather, as Lamb and King (2003) suggest, they are ‘critical social actors’ who engage with the technologies that are available. They point towards the importance of this finding, as it suggests that the failure of technology adoption is not always a failure on the part of the users in understanding how to use technology, but the users judge the technology to be inappropriate for their research needs.

In terms of the nature of collaboration, Meyer *et al.* (2011) found that while chemistry exhibits smaller collaborations as the equipment does not require large teams to build and maintain them, the Zooniverse is also an example of collaborations of a large scale which do not require large investments in infrastructure. This finding could be attributed to the participation of the public through a web-based platform that helps in processing the data, also data itself is open and available through public databases. Consequently, generating and collecting the data does not contribute towards the cost of a Zooniverse project. Based on the case of the Zooniverse, Meyer *et al.* (2011) allude to possibilities for scientific research offered through citizen science, with its use of relatively simple web-based technologies to harness the cognitive power of many volunteers, in ways which are yet unparalleled.

One of the few ethnographic studies in e-Science, is of the SENSEable City Lab (SCL), a research initiative at the Massachusetts Institute of Technology (MIT). The study investigates the ways in which digital technologies are changing human lives on an urban scale. The collaboration over the past several years has had 350 members representing disciplines from Architecture, Mechanical Engineering, Theology and Space Science (Simeone and Ratti, 2011). They suggest that with a context-driven, problem-based approach, the initiative embodies what Gibbons *et al.* (1994) propose to be Mode 2 knowledge production. The preliminary findings of the study that spanned four months in its initial phase, focuses on the organisational culture. The findings indicate that the structure of the organisation of the lab emerges from the interweaving functions of the teams, with a constant process of horizontal and vertical integration. The flexibility of the organisational structure in the transdisciplinary lab is facilitated by the lack of order imposed from the top. Teams were encouraged to be responsive to the changes and the collaborators appeared to be aware of their contributions to the lab. Simeone (2013) later employs the notion of 'strategic ambiguity' proposed by March and Olsen (1976) to reflect on its potential as a dialogic strategy, which accounts for the differences among members of a transdisciplinary organisation. Strategic ambiguity is a :

"[s]trategy for suspending rational imperatives toward consistency [that helps organization] explore alternative ideas of possible purposes and alternative concepts of behavioural consistency" (March and Olsen, 1976: 77).

As a managerial approach, strategic ambiguity calls for a deliberate use of inconsistent communication that omits significant contextual cues which permits space for multiple interpretations, allowing members some freedom of interpretation and action based on their own viewpoints (Eisenberg, 2007). This is the beginning of a rapidly growing area of Science Studies, which will require drawing on old concepts and creating new ones towards understanding how practices are changing in the novel setting of e-Science.

Olson *et al*. (2009) contend that the major hurdles for such collaborations are issues around tacit knowledge and the lack of common ground. However, collaborations over a long geographical distance are increasing all across science under those restraints and technologies, and practices are evolving in response to those obstacles (Rosner, 2012; Vertesi 2012). Cummings et al. (2008) asserts that the pervasiveness of communication technologies such as Twitter, Skype, Google Hangouts and Youtube demand a renewal in how we understand scientific knowledge and knowledge production.

With the changes in new technologies and methods in science, what qualifies as knowledge and how to treat those new forms of knowledge and methods is also being recognised (Edwards P. *et al*. 2013). Giere (1999) offers an example of discussions between scientists and philosophers regarding the treatment of simulation models and questions whether the understanding of validation and verification methods apply to both theoretical and experimental models. Anderson (2008) and Hey *et al*. (2009) similarly propose that what counts as knowledge is being questioned and changing.

3.4 3 Challenges in Data Sharing and Re-Use in Science

As data serves science as its 'life blood' (Jirotka *et al*. 2013), the potential of e-Science lies in the shared access it offers and the opportunities for the reuse of data by scientists in collaborations across disciplinary and geographical boundaries. Edwards P *et al.* (2013) acknowledge the growing excitement engendered by data mining and sharing across scientific disciplines. However, for scientific use, other argue that the meaning of data is tied to the details that entail how, when and where they were created (Ribes and Jackson 2013; Burton and Jackson 2012; Bechhofer *et al.* 2010). There are therefore questions about the kind of solutions and practices necessary for ensuring that data is reusable and shareable between scientists, which need to be addressed (Edwards P *et al.* 2013). Existing practices have not only set limits on the output of knowledge in the forms of printed media but also on determining who are valid participants in the research process.

Jirotka *et al*. (2013) and Zimmerman *et al.* (2009) argue that an understanding of the context in which data was created is critical to its appropriate use. They speculate that perhaps the need for supporting the range and scale of interdisciplinary work was not anticipated, and these require accounting for the context of data production, capture, use and reuse. Hartswood *et al.* (2005) argue that prior to data sharing and reuse, the data requires work; while Edwards P *et al*. (2013) propose notions of obligation in relation to how data should be embodied in the practices and values of the individuals, technical systems and institutions involved. Offering an example of the notion of sustainability of knowledge infrastructures, they contend that the long-term purposes of the conservation and preservation of data and associated data-related practices must be considered. They suggest that sustaining knowledge infrastructures requires historical and contemporary studies to examine how such structures change, how they break down and the factors that enable or prevent the systems to endure through time. The authors assert that, given the varying layers and scales of time and space, human collectives and data that characterise the infrastructures, there are considerable critical challenges in designing, using and maintaining infrastructures that will be robust in nature. They suggest that the struggle for agreement on standards and ontologies that have to consider tensions between a vision of universality against a desire for change, would be fundamental.

Clearly, it is not only Computer Science and Computing Technologies that science has to work with; the rise in multi-, inter- and trans disciplinary collaborations also offer an opportunity to re-examine the nature of scholarship in the current digital age (Edwards P *et al*., 2013; Borgman *et al*., 2012). It is not only the technical dimensions of the challenge that is critical, but ethical considerations have to be recognised. Lessig's (1999) caution over the social consequences of a techno-centric approach, whereby technologies can enable socially unacceptable behaviour and prevent acceptable and legal behavior, still holds true. Borgman *et al.* (2012) similarly propose that amidst the complexity of the technical issues, other social and legal issues such as managing trust and intellectual property must not be overlooked.

Yet another challenge to the building and using of complex e-Science data systems is highlighted by Edwards P *et al*. (2013). This challenge relates to Kuhn's (1962) notion of 'incommensurability'. They argue that an object-oriented solution[12] perspective in data sharing between disciplines is theoretically improbable, as the fundamental ontologies will differ and also change as disciplinary fields evolve. The notion of 'incommensurability' can be illustrated with the example that 'mass' in Newtonian Physics fundamentally differs from the understanding of "mass" in Einsteinian Physics (Kuhn 1962).

Edwards P *et al*. (2013) note that a new cadre of metadata practices and professions need to be trained in order to address such issues, and since the field of practice to support such infrastructures is relatively new, they suggest considerable challenges within and among different disciplines. This is where social science methods may be helpful. Qualitative empirical research in the areas of Sociology of Science, and STS has been mainly ethnographic in nature (Knorr-Cetina 1999; Latour and Woolgar 1979), and has clearly been established as one of the dominant methodologies in studying scientific cultures (Marcus, 1995).

Edwards P *et al.* (2013) speculate that new knowledge infrastructures would not only help find answers to existing research questions, but will make way for new questions. As illustrated earlier, the study by Meyer *et al.* (2011) found no evidence of generating new questions, but they did find that "more of the same" was enabled by the new technologies. Therefore, further research into the ability

[12] "If the world actually corresponded to the hopeful vision of data-sharing proponents, one could simply treat each discipline's outputs as an "object" in an object-oriented database (to use a computing analogy). Discipline X could simply plug discipline Y's outputs into its own inputs. One could thus capitalize on the virtues of object-orientation: it would not matter what changed within the discipline, since the outputs would always be the same. Unfortunately, this is unlikely — perhaps even impossible — for both theoretical and practical reasons (Borgman *et al*.,2012)."

of new infrastructures to help scientists ask new and different questions than the existing ones, is yet to be seen. However, there are examples of fields that owe their development heavily to technologies, such as artificial intelligence which evolved directly from interactions with computers in the 1950s (Edwards P, 1996).

Citing Kuhn's (1962) "paradigms," and Lakatos' (1970) 'research programmes' as conceptualisations of this shift, Edwards P *et al.* (2013) argue that the new knowledge infrastructures mark a shift in the history of science and production of human knowledge. They suggest that these infrastructures not only help map the geography of human knowledge, but also shape the geography itself. They present examples to support their argument: a) Web-based citizen science led by initiatives such as GZ and FoldIt; and b) the Cornell Laboratory of Ornithology's eBird program. These examples are often cited for their new strategies for scalable scientific practice and educating and communicating with the public (Kelling *et al.* 2012); while also raising questions about the expertise required in public participation in the production of data for science. Edwards *P et al.* (2013) anticipate that scientists will need to find how to keep the public engaged in a sustainable way. They suggest this demand might require reworking some of the established traditions of public engagement, some of which started decades and even centuries ago in order to distinguish the expert work of credentialed scientists from amateur contributors.

3.5 Emerging Challenges and Concepts in Understanding Epistemic Cultures of e-Science

Following the discussion on the multi-faceted challenges facing e-Science, it is clear that science is undergoing significant changes. According to a cross-disciplinary citation analysis conducted by vanLeeuwen and Tijssen (2000) from the period of 1985-1995, more than two-thirds of citations came from cross disciplinary boundaries, with fields like medicine being increasingly compared to other fields like Astronomy. Cummings and Kielser (2011:3) suggest that science has

"[u]ndergone major organizational changes over the past century and has embraced new ways of structuring incentives (e.g., million dollar prizes), collaborative relationships (e.g., virtual scientific networks), project governance (e.g., open source projects), scientific participation (e.g., citizen science) and knowledge dissemination (e.g., publicly accessible journals)."

In terms of data sharing, Kowalczyk and Shankar (2011) suggest that while stakeholders have shown an interest in making data re-usable through development of repositories, metadata schema, standardised ontologies etc., there is no clear research on consolidating what is known and what is not. They discuss the various overlaps of interest between different shareholders who at the level of policy are concerned about sharing data, to the practicalities of ethical training, which is yet another challenge for science.

The history and sociology of STS have contributed to the understanding of knowledge sharing through construction of social worlds and conceptual frameworks with ideas such as "boundary objects," "trading zones," and "actor networks," and developed an understanding of how such sharing works in practice (Callon and Latour, 1981; Galison 1996, 1997; Latour 1983, 2004, 2005; Latour and Woolgar 1979). In addition, studies from sociology of standards have contributed to an understanding of how large communities with heterogenous stakeholders reach agreements on shared practices, norms and technical systems (Bartky 1989; Blanchette 2012; and Sundberg 2011).

Jirotka *et al.* (2013) proposes that literature from the design community might enable the transformation of such concepts into practice in a scalable manner. He argues that socio-technical phenomena are not limited to a single domain of research or approach, citing an example of how climate change simultaneously requires action from individuals, state policy and is a matter of economic reorganisation and technological innovation. His outline for conceptual challenges to e-Science ends with the assertion that learning from the peculiarities of disciplines while formulating conceptual and theoretical frameworks can transcend disciplinary boundaries. He illustrates his point by suggesting that theories of system and tool design along with theories of collaborative socio-technical innovation have the potential of wider applicability across research disciplines and collaborative knowledge work.

While research in STS has studied scientific collaboration for several decades, Jirotka *et al.* (2013) argues that the field lacks interaction with design and implications that Computer Supported Cooperative Work (CSCW) offers. Similarly, while technical fields have been engaged in e-Science, they lack theories and methods to understand practices that inform design of not only simple tools, but also large scale socio-technical systems. By explicating the history that CSCW has in combining various disciplines such as Computer Science, Organisational Science, Information and Social Sciences, Jirotka *et al.* (2013) suggests that the large undertaking of designing new knowledge infrastructures in order to understand and respond to the complexities of the current data-intensive paradigm science is underway. Similarly, Hine (2006) argues that STS is an established field that offers critique and analysis of scientific and technical developments that contribute to understanding design and its process. However, she warns of the anti-science perception that STS has developed among some practitioners of science, but is more hopeful in the area of technology design where STS scholars have found acceptance. Nevertheless, she advocates that discussions of e-Science should be seen as an opportunity to constructively engage science with policy.

Pentland (1995:20) captures the paradox of the e-Science in the following excerpt:

"[t]he very systems that are meant to increase our information processing capabilities, thereby increasing understanding, may have the opposite effect by restricting the range of our inquiry and experience, effectively putting us in a kind of epistemological box. Whether information systems enhance or dull our senses is a difficult question to answer, but it is clearly an important question to ask".

Therefore, as Big Data science, e-Science, knowledge infrastructures and cyberinfrastructures draw the discussion around scientific data from fields varying from STS to Artificial Intelligence to Social Computing, it is clear that the complexity involved in mapping and understanding the challenges in the fourth paradigm are only beginning to be understood.

One of the emerging perspectives in addressing that challenge is the rise of design thinking in epistemic cultures, specifically in how knowledge infrastructures of science are shaped, with the potential to contribute to the understanding of knowledge object - subject or expert relation as it unfolds, and demands attention to the process by which the object, the subject and the system itself is deliberately shaped. It provides an analytical tool to STS researchers in examining the evolution of a knowledge production system as a result of individual and social forces within science acting in response to the current challenges arising from rapid technological advances in creating, capturing, storing, sharing, analysing, releasing, reusing and archiving data. Therefore, the design perspective has the potential to contribute to our understanding of the complexities involved in the practices and expertise that emerge and evolve within an epistemic culture in a data-intensive era.

Chapter 4

Research Design

This chapter presents and provides the rationale for the research design and the analytic processes undertaken in order to address the overarching research question:

How is citizen science shaping the practices and expertise of Galaxy Zoo scientists?

1) In what way is citizen science changing scientific practice?

2) What practices can be identified in the work of Galaxy Zoo as a successful citizen science initiative?

3) What kinds of expertise are evident in these citizen science practices and how are they developed?

Since the research questions are exploratory in nature and the future of Galaxy Zoo (GZ) was uncertain in terms of its directions, the research design had to be able to capture this complexity and uncertainty as events unfolded. As explained in chapter 1, this work is a study of evolving practices in a context that has not, to my knowledge or that of the GZ participants, been empirically studied before. For these reasons, a qualitative approach which attended to the participants' context-embedded actions and decisions was considered most suitable for studying relevant phenomena as they unfolded and revealed themselves.

The key feature in initial decision-making about research design was therefore what Rossman and Rallis (1998) describe as the flexibility of much qualitative work. The research design also draws on what Bhaskar (1978, 1989) calls "transcendental realism". He suggests that social phenomena exist not only in the mind but also objectively in the world and have reasonably stable relationships among them. A transcendental realist approach is in line with my belief, that although the constructs that underlie social and individual life may not be visible, their invisibility does not reject a claim to their existence.

This position is also central to the epistemological view that the pursuit of enquiry should have value to the participants. In brief, for research to have value in practice, a view derived from the concept of phronesis put forward by Aristotle, the world has to exist, not just in our minds, but it has to exist operationally. Therefore, a phronetic epistemology, evident in, for example, Flyvberg's (2001) idea of phronetic Social Science is accepted as a way to gain some purchase on reality, even if we do not have complete access to it. The concepts employed and tested in the research become tools to help participants to approximate reality. Rejecting a view that reality cannot be at least partially represented and understood leads to the conclusion that the research carried out is of no value beyond intellectual pursuit. The latter is a position I disagree with, mainly on the grounds of lack of accountability and validity for the research.

There have been attempts to simplify the relationship between research paradigms and methodologies. For instance, Hatch (2002) suggests five major paradigms with corresponding methodologies; while Guba and Lincoln (1994) suggest the implications of the choice between four paradigms- positivism, postpositivistm, critical theory and constructivism. The long list of paradigms and methodologies in qualitative research (Robson, 2002; Hatch, 2002; Miles and Huberman, 1994) will not be discussed here. Instead, I shall attempt to explore the nature and value of examining the world through research paradigms. Kuhn (1996) proposed the idea of paradigms as perspectives passed on in history through which the world could be viewed. Kuhn (1962) also argued that the dominance of clearly established paradigms in 'mature science'[13] distinguishes these areas of research from newer fields. Beecher and Trowler (1989) echo this view in their argument that disciplinary overlaps in these newer domains may result in lack of consensus over paradigms, which could explain the number of positions taken in social and educational research where a multidisciplinary nature of inquiry is accepted.

Nonetheless, Guba and Lincoln (1994) suggest that paradigm issues are crucial and no researcher should engage in inquiry without being clear about what paradigm guides his or her approach. Here they are supported by Robson (2000) and Hatch (2002), but not all researchers agree. Seale (1999) is among those who consider the burden of fulfilling some philosophical or methodological scheme a considerable threat to quality of the research. I find this debate helpful as, as will become evident, part of the challenge of the present study has been to develop a methodological approach, which is sufficiently flexible to capture the shifting and complex inter-relationships that constitute citizen science in GZ.

Perhaps because of increased subtleties in the work of science and the

[13] "Successive transition from one paradigm to another via revolution is the usual developmental pattern of mature science." (Kuhn, 1962:12)

recognition of human engagement even in the most objective of enquiries (Knorr-Cetina, 1999), the ontological and epistemological debate in the purest of sciences such as Physics is gathering momentum in relation to the nature of the relationship between the inquirer and the inquired (Guba and Lincon, 1994). For example, one of the leading Theoretical Physicists, David Bohm, argued that indivisibility exists between the observer and the observed. The observer effect in the field of Physics can be explained as the change in behaviour of the phenomenon observed as a result of observation by an observer. This move could potentially intensify the debate in the study of social phenomena and may inspire an interdisciplinary debate between the ‘hard’ and the ‘soft’ sciences, which would open exciting avenues in human understanding of reality and how human epistemology evolves. I concur with Seale’s (1999) suggestion of acknowledging the value of lessons that can be learnt from a variety of approaches. He urges for a sense of community of social researchers who respect strengths of varying positions and recognise the value of developing research skills taken from different approaches. Huberman and Miles (2002), nonetheless, offer some basic yet wise counsel by suggesting that the best research can do is be thorough and explicit, since vague descriptions are of little practical use to other researchers.

At this point, as a researcher, I see myself as a constructivist, who aims at working with the best conceptual tools available to gain some grip on the field I am examining. However, I find the paradigm debates fascinating as they have particularly encouraged a critical approach to prescriptive texts on how to do research. As researchers, we have a responsibility not only to engage in intellectual inquiry and debate on the nature of our inquiry, but also to the participants and institutions that make empirical research a possibility. It is this responsibility to the research participants that has perhaps been one of the strongest methodological learning points for me in the present study.

4.1 Methodology

In the sections that follow, I outline the methodological approach I took in pursuing the research questions from February 2010 to April in the life of GZ. I then outline the research design, discuss the analytic process and consider the ethical implications of the study, including my own position as a researcher in the study.

As already explained, GZ is an online citizen science initiative collaboratively run and managed by academics affiliated with various institutions in geographically diverse locations mostly across the UK and the US. They communicated primarily via e-mails and other modes of online communication technologies and very rarely had face to face interactions, with the exception of the core management team. Exploring the processes and structures within GZ that produce scientific knowledge, and therefore the expertise needed, required understanding the practices that participants inhabited, the activities that occur with them, the actions participants take and their interconnectedness and purposes. In effect, I was examining the activities in emergent practices and trying to capture them. The features of ethnography outlined by Hammersley and Atkinson (2007:3) fit aptly with the nature of this research, as they allow a flexible research design that permits a focus and objects of inquiry to emerge with an increasing depth of understanding the focal case in context. As they summarise the process, an ethnographic study has the following characteristics:

"a) it involves studying people's actions and accounts in the contexts they occur;
b) data are gathered from a range of sources, however, participant observation and informal observations are dominant;
c) 'unstructured' nature of data collection as the research design is not detailed and fixed and the categories for interpreting data emerge out of the process of data analysis;
d) the focus is usually on a few cases of fairly small scale in order to facilitate in depth study, and
e) the data analysis involves interpretation of meanings, actions and practices of the institutions, which result in production of accounts through verbal descriptions, explanation and theories" (Hammersley and Atkinson, 2007:3)

Some of the most important qualitative empirical research in the areas of Sociology of Science and Science and Technology Studies (STS) has been ethnographic in nature (Knorr-Cetina, 1999; Latour and Woolgar, 1979) and this approach has clearly been established as one of the dominant methodologies in studying scientific cultures (Marcus, 1995). After an initial period of familiarisation with how the GZ team was set up, in light of the guiding research question, it was determined that an ethnographic study should be undertaken. This decision was operationalised by ultimately following the work of the Principal Investigator (PI) and later the Technical Lead (TL) (who took over the PI's responsibility for coordinating the work at Oxford while the PI was placed in Chicago for a year as a result of funding conditions). The ethnographic approach undertaken in this study followed guidelines put forward by Geertz (2000) and Hammersley and Atkinson (2007) and adapted them to the context of e-Science as suggested by Hine (2007) in order to take into account the inherent spatial complexity involved.

Accessing negotiations between the PI, the TL and their colleagues involved understanding the organisation of GZ's work, which led to the development of a preliminary conceptual framework for the study. This process began by identifying the key informants, namely the PI and the TL of the project, as the functions of all the teams involved were led by them. I needed to get close to them and try to see GZ from their perspectives. The idea of co-presence as put forward by Beaulieu (2004) provided a conceptual rethinking of what an ethnography entangled in communication technologies could entail. It was established quite early that online communication technologies were an indispensable element of how GZ's members organised themselves. Therefore, formulating methods to account for the events occurring through those channels was a significant adaptation to the mode of ethnography suggested for example by Geertz (2000).

4.1.1 Co-presence: Reconceptualising Ethnography

"To discover who people think they are, what they think they are doing, and to what end they think they are doing it, it is necessary to gain a working familiarity with the frames of meaning within which they enact their lives. This does not involve feeling anyone's feelings, or thinking anyone else's thoughts, simple impossibilities. Nor does it involve going native, an impractical idea, inevitably bogus. It involves learning how, as being from elsewhere with a world of one's own, to live with them" (Geertz, 2000: 15).

Hine (2007) recognises some more recent challenges, arguing that, while ethnography is a well established way of exploring the details of the processes of knowledge production, some adaptations could be necessitated by the spatial complexity of e-Science, some of which have been discussed in chapter 3. Although the focus of the present study is not the e-Science characteristics of

GZ, borrowing an approach developed in that field allows for exploring an emerging scientific practice. Beaulieu (2010) undertakes the challenges of examining science as it happens across different sites and proposes co-presence as an 'epistemic strategy' that pays attention to knowledge production in non-lab based settings that can account for infrastructure, textuality and mediation, suggesting that such a strategy helps the ethnographer as a participant-observer, scholar and author and that co-presence as an approach to doing fieldwork that generates new prospects for the study of knowledge production. The argument is that co-presence enables STS as a field to develop ethnographies of highly mediated, distributed or non-lab-based contexts. .

One of the challenges that Beaulieu and Simakova (2006) discuss for such studies is the formulating of strategies for identifying field sites and exploring the connections between them. For example, a hyperlink in a database might need to be followed just as much as a presentation at a conference. This approach differs in manifold ways from many earlier examples of ethnography, as attention is given not only to pre-determined field sites, but also to the mobility and connections inherent in the continuous move from online to lab settings, focusing on how various forms of communication are employed to contextualise one another (Leander and McKim, 2003). Leander and McKim (2003) also argue that tracing and tracking relevant boundaries and connections is an ethnographic puzzle in such studies, rather than a question to be solved before the ethnography begins.

The present ethnographic study combined observation and interviews within an emergent system of practices through e-mail list participation, exploration of GZ's online presence, and analysis of expectations of information and communication technologies as portrayed in the grant proposals to funding bodies. Beaulieu (2005) also argues that ethnographic studies of this nature should make connections between expectations of practice and the actual practices, as what scientists do online might have implications for how online activities are transformed and transported to other domains. It is the process of multiple transformations, as practices and expertise extend across domain boundaries, and their implications for science as a profession that became the GZ activity that was most fruitfully examined in the present study.

According to Beaulieu *et al.* (2007), fieldwork in such contexts is carried out by following actors, allowing the researcher to identify the activities involved in developing artifacts or knowledge in the setting. Therefore, a triad is assumed between the lab setting or context of use, the ethnographic account, and the constitution of an object of study for the research. In the present study the objects of study became the practices and expertise within GZ.

The research design is also informed by the notion of Vygotskian 'sense-making', which allows developing expertise in practice to be examined through analysing the changing relationships between the context and the actor. The approach taken in the design of this study considered a focus on the sense-making and actions of the PI and TL to be a way of examining how they were interpreting and responding to the demands they faced as GZ developed and how their responses shaped those practices. The use of sense-making here is therefore not related to the Vygotskian concern with how personal sense connects with public meaning. Instead it allows attention to what was being interpreted and why certain aspects of phenomena were regarded as important by the two key participants. Edwards A et al. 2002) indicates that in practices where what is seen as expertise is not prescriptive, which is the case in this study of GZ scientists, participants' own externalisations of meaning will have influence on how the meaning is shaped and understood by the participants themselves. This is an argument whichI would suggest is congruent with the notion of Flyvberg's (2001) phronetic Social Science introduced earlier in this chapter. An alternative terminology derived from Flyvberg's (2001) work would be phronetic design, which places ethics and epistemology in design of research, the parallels for which were revealed in the design decisions relating to GZ tools for their volunteers in chapters 7 and 8.

In pursuing these objects of study I therefore also draw on a cultural historical approach which emphasises how ideas in use are revealed through actions in activities, which are themselves embedded in practices. Observations, including analyses of texts that aim at communicating, can then be followed by interviews about actions and activities to allow researchers to probe personal sense-making in the course of the activities examined. As Edwards A (2010) observes, one of the advantages of this approach, derived from the work of Vygotsky, is that it allows the researcher to maintain some grip on the dynamic changes that occur when people act in and on the practices they inhabit. This approach is therefore particularly appropriate for tracking emergent practices and the expertise employed in them.

4.2 Research Design

The data gathering period began from February 2010 and was carried our regularly till June 2011, however, follow up meetings and interviews were concluded in April 2013. In addition, the negotiation of access took an additional period of four months from the first meeting with the PI in October 2009, until the GZ team consented to my research proposal in February 2010. The first contact with the PI, who subsequently became the main gatekeeper for this research, was made after attending a public lecture he presented in Oxford in September 2009. The initial exchange with him was followed up by an e-mail expressing interest in studying how citizen science projects worked. The initial concepts and ideas that I had in examining the communication between scientists and the public complemented the nature and the context of GZ. I took some time to familiarise myself with the GZ website and attempted some classifications to understand the tasks that were given to the volunteers. The following few meetings with the PI involved explaining what kind of research I had envisioned, through which I began to understand what the PI's role entailed. During the time of these discussions, GZ was in its second year of running the second version, GZ 2 and in preparation of launching the third iterations of GZ - GZ Hubble.

As I have already indicated, the two key informants identified as actors to follow as they pursued their enquiries through citizen science, were the PI[14] and the TL. The PI joined the Department of Astrophysics at the University of Oxford as a post doctoral researcher in Astrochemistry, after completing his PhD. He had a

[14] As I have already indicated, both the PI and TL agreed that complete anonymity was neither possible nor desirable.

history of being involved in science communication from a very early age when he was in school as a contributor to an amateur Astronomy magazine. This interest of his developed with more time being spent in communicating science which also led to him becoming a regular presenter on BBC's Sky at Night. Although he had experience in science communication for several years, he had never been responsible for managing a project prior to GZ. He clearly had wide interests and was seen as an excellent manager and communicator, which he believed made him more of a 'generalist' than a 'specialist.'

The TL on the other hand had no prior experience in science communication and joined GZ as the lead software developer, having worked at a private research institute as a programmer after being awarded his doctorate in Astrochemistry. He was to assume the role of the TL leading the technical front of GZ. He was recruited as the second salaried full-time professional in January 2009. At the beginning of the fieldwork, his expertise was mainly in building software for GZ. However, with time and in response to the demands of the nature of work at GZ, his work in GZ extended beyond simply development as he began presenting GZ's work to various audiences and began to think of new ways of programming to transform how research is carried out in the field of Astronomy. The PI described him as a 'game changer'. He brought in new perspectives on possibilities for GZ on the technical front, making the examination of the distributed intelligence elements in GZ more responsive to GZ's science needs and to the behavior of the citizen scientists with whom they collaborated. It was fascinating to follow the work of these two key informants who came from similar academic backgrounds with significantly different professional experiences as they gradually formed the Core Team to run the most successful citizen science project to date.

4.2.1 Negotiating Access with the Galaxy Zoo Team

At the time when the initial contact was being made, the GZ Team had approximately 20-30 members located across various institutions with the team mailing list being their primary medium of communication. The membership and affiliation to GZ did not have any clear guidelines, therefore collaborators who were a part of the mailing list were the ones accounted for. After several discussions with the PI about what the nature and scope of this study could potentially shape itself to be, a proposal was drafted and discussed with him for approval. After providing very helpful feedback he forwarded it to all members through the mailing list with an additional note that if anyone had any objections to participating in the research they should contact the PI or request to opt out. This mechanism was suggested by the PI as it was the general rule that the team had established among themselves for any matter that required consent. The PI gave the team members a week to opt out of the research and at the end of the allocated period, he confirmed that I could go ahead with the research, which involved firstly allowing access to the team mailing list. The members on the mailing list were made aware of this fact. Having the PI as the gatekeeper was critical to negotiating access with the rest of the team, not just in terms of the time taken, but also since his willingness to not only participate in the research, but to take up the responsibility to get others involved indicated his own interest and involvement. He was also the only member of the team who personally knew all the members on the team, unlike many other members who had never actually met each other and only had known each other through the interactions in e-mails sent via the mailing list. Over the period of the first 16 months of fieldwork, I met with the PI at regular intervals of two weeks to a month, depending on his busy travel schedule. Table 4.1 presents an overview of the data collected for this study.

Source	Description	Total no.	Participants/ Informants	Dates
Interviews	Semi-structured/Unstructured	44 (24 with PI, 13 with TL)	PI, TL and GZ members	February 2010 to April 2013
Conceptual Maps	Constructed by the PI and TL	4	PI and TL	February 2010, December 2010, June 2011
Documents				
	Publicly Available			
	GZ Website	(Periodically monitored for changes)		February 2010 to April 2013
	GZ Blog Posts	120		February 2010 to April 2013
	Peer-reviewed papers	20		Published between 2008 to 2013
	Internally Available (Access granted by the PI and TL)			
	Science Team Mailing List	632 email threads		February 2010 to April 2013
	Funding Applications	4		February 2010 to June 2011
	Internal Reports	9		April 2010 to May 2010
	Presentations by PI	15		Dated January 2008 to March 2010
Observation	Events Organised by GZ	2	PI, TL and GZ members	22-23 April 2010; 4th April 2011
Social Media Platforms	GZ's Tweets			Periodically monitored from February 2010 to April 2013

(All Publicly published)	GZ's Google Hangout	1	GZ Science Team members	February 2, 2013
	PI and TL's Blogs		PI and TL	Periodically monitored from February 2010 to March 2013
	Twitter Accounts		PI and TL	TL (Between Jan 2009 to April 2013) PI (Between September 2008 to April 2013)

Table 4.1 Overview of the Data Collected

Participants	Role in the GZ Team	Date of Interviews
GZ PI	Co-founder and Principal Investigator of GZ (US/US)	February 2010 - April 2013
GZ TL	Technical Lead of GZ (UK/US)	March 2010 - October 2012
GZ Team Member 1 - GZ1	Co-Founder of GZ and Science Team member (US)	12th March 2010
GZ Team Member 2	CSA Board Member and Science Team member (US)	22nd April 2010
GZ Team Member 3	Education Research Team member (US)	23rd April 2010
GZ Team Member 4	Education Research Lead (US)	17th December 2010
GZ Team Member 5	Education Research Team Member (US)	17th December 2010
GZ Team Member 6	GZ Team Scientist (UK)	4th April 2011

Table 4.2 List of Participants Interviewed

During the period in which my access was being negotiated, several meetings with the PI were held. These meetings contributed to my gaining a sense of GZ history in order to comprehend the roles and the processes which had played a part in shaping how GZ operated. It became clear very early on in the data gathering process that the Core Team members of GZ did not wish to formalise the GZ structure and how they operated as an organisation. Their intention was to remain structurally flexible and open in the ways they worked and in contrast to the current modes of scientific practice, which they understood to be rigid. Their ambition was not only to produce scientific results, but also to experiment with new practices, especially through citizen science in an academic research context. Knowing that GZ deliberately lacked a formal structure had implications for the questions I could ask and the research design, which had to allow for capturing data on emerging practices and expertise.

A conceptual framework of GZ was created two months into the field work to represent the practices, interactions, actors, tools and stakeholders (See Figure 2A). The representation exercise also assumed certain boundaries and categories of experience and distributed intelligence to be located with the structure of GZ. The first version of the conceptual framework (Figure 2 A) was a helpful exercise in locating the complex interactions between the knowledge objects being worked on and scientists and the public, and the tools that they used, which were found to be inherently interlinked through the activities that constituted the emerging practices of citizen science in GZ. The model was presented to several team members of GZ and was recognised as a useful and original mapping of the evolving structure. Therefore, preliminary analysis had already begun as a result of interactions with the PI during the initial stages of data collection.

All the data gathered helped me orient myself with how GZ worked and see the changes in the negotiations within the GZ team. However,, as the study progressed, I became focused on the practices and expertise being developed and exercised by the Core Team and therefore, the decision to not use interview data from the other team members was made since it would have been a deviation from the focus of the study. Please refer to Appendix A for a detailed timeline of the research.

4.2.2 Identifying the Areas of Interest and Informants: Locating the Object of Study

The conceptual framework presented as Figure 2 A, and which is discussed in more detail in chapter 2, became a helpful representational tool not only to provide a complete picture of the context in which GZ operates in, but also the areas of GZ activities that later became the focus of the research. These were the core scientific practices and accessing distributed expertise through a focus on the work of the PI and the TL. Once the areas of interest were determined, the next decision was locating the relevant data, which meant identifying the significant informants beyond the PI and the TL and finding sources for secondary data (for example, the GZ blog and proposals to funding organisations). The Core Team comprised the PI and the TL who at the time, were both based in Oxford. However, the PI relocated to the partner institution in the United States (US) at the end of June 2010 for a period of one year as required by the US funding agency, the TL took over leadership responsibilities at Oxford. Although the nature of the PI's work remained the same in terms of managing the project, there was a change in the work environment (from academia to a public research and education institution). Upon the PI's return to Oxford[15] in June 2011, the TL took up the permanent position of Director of Citizen Science at the US institution, a post created entirely as a result of collaboration with GZ. Their roles, therefore, were often interchangeable apart from their core areas of expertise. This merging of aspects of expertise at the core of GZ activities was recognised several months into the fieldwork, and allowed for reframing the interview questions to explore the nature of that expertise in the tasks they carried out.

With co-presence as an ethnographic strategy (Beaulieu, 2010), it was determined that since GZ exists as an organisation only online while the team members are dispersed 'behind the scenes', it was essential to follow the social media presence of the PI and TL, which was publicly available. This involved monitoring their official blog, mass e-mails sent out to volunteers and Twitter feeds. In order to continuously keep track of the data, an account on the GZ website was created and a subscription to the blog posts was requested keeping me informed of the developments as they happened. The information collected in this way became helpful cues in constructing interview questions for the participants and to verify the developments that were reported by the participants. In terms of secondary data, documents generated by the team members for internal use (i.e. publication rules), grant applications and presentation slides for a variety of audiences were retrieved from the PI and the TL on a regular basis with intervals of approximately two-three months. Details on tracking

[15] After the completion of his mandatory one year at the Adler Planetarium as required by the National Science Foundation education research grant.

documentary and online data in order to capture the knowledge objects (Knorr-Cetina, 1999) they were working on will be discussed in further detail in section 4.3.

4.2.3 Designing for Uncertainty

Several opportunities arose during the data collection period to interview other members of the team. These included two US-based Education Leads and several members of the science and development team from both the US and UK. These meetings usually occurred during their visits to Oxford to meet the Core Team members of GZ and my two trips to the partner US institution, one in April 2010 and the other in December 2010. On several occasions, the PI alerted me to the opportunity and was willing to make introductions whenever he deemed it would be interesting for the research to have the other team members involved. It was established with the PI that I checked with him before I made first contact with any of the members and he helped me significantly with second introductions by e-mail and in person. Apart from one member of the team, I was allowed to communicate with the team members without his permission or approval on the content of the communication. It was also suggested that I should attend two events: a) the first meeting with their advisory panel at the US partner institution in April 2010; and b) a conference organised at Oxford for other scientists who were interested in using new online methods and technologies to do science in April 2011.

In particular, I was interested in whether further developments were influenced by personal interaction and whether differences were visible when compared with the interactions through online platforms. Twitter feeds during the conference were monitored to follow conversations surrounding the event which ran parallel to the event itself. The Google Alerts system was set up for GZ related news to monitor how the project was being perceived in the media, especially when GZ team members were required to be involved in an article being presented or were required to defend their work. Therefore, monitoring parallel sources of online internal and publicly available data and participation in GZ team's events continually informed the research design throughout the fieldwork process.

4.3 Methods

The methods selected were:

A) Interviews: The extensive literature on interviews (Punch, 1998; Kvale, 1996; Rubin and Rubin, 1995; Robson 1993) suggest they are the most effective way of accessing how people make sense of what they do. Accordingly, I used semi-structured and unstructured interviews to allow me to keep a focus on activities in practices of the participants while allowing enough flexibility to follow emergent themes. Interviewing different key informants also allowed me to identify varying understanding of the developments of GZ. The structure of the interviews changed over time based on new data from the other sources used. In April 2010, I began interviewing the TL, who became a key informant after the PI was relocated to Chicago. Most of the interviews with the rest of the team were selected according to the precise object of study being followed at the time. For example, I interviewed a few GZ team members immediately after the GZ team events in order to gauge their expectations of the impact of what they had heard. In total, 44 interviews were carried out, 24 with the PI, 13 each with the TL and the rest of the GZ Team; with a total of 59 hours of recorded interviews. All the interviews with the PI and the TL were transcribed and notes were taken from the interviews with the other GZ members. During some of the GZ-organised events, I had the opportunity to have informal conversations with the team members as a result of my participants in those which were not used as data for this research, but inspired some of the questions that were followed up with the PI and the TL. My participation in the two key events I attended were on the approval of the PI and he was made aware of the team members I had spoken to ensure transparency in the research process with my key informants.

B) Documentary sources: As e-mails were the most commonly used method of communication between the team members, there was an archive of e-mails from the inception of GZ in 2007, which was a rich source of data. Rossman and Rallis (1998) argue that examining documents is an unobtrusive way of data collection and affords the portrayal of the values and beliefs of participants and the community, which might not be revealed by other means. However, Laws *et al.* (2003) warn researchers to carefully examine the authors of the documents to be analysed as they often merely reflect the views of those who produced them. In this study, documents formed a critical source of interview questions, themes and identifying participants to be interviewed for a deeper understanding of phenomena. The documents used in this study can be categorised as follows

a) Publicly available:

i) GZ Website: The website is where GZ is represented as an organisation to the public and this is the portal where the public is involved in classifying galaxies.

ii) GZ Forum: This online forum is where volunteers discuss various topics of interest in Astronomy and support each other in answering queries about classifications. Some of the scientists of GZ also participate in answering more difficult queries.

iii) GZ Blog: The blog acts as an online newsletter to the public on the new developments relating to GZ website, findings and other news.

iv) Academic papers: Published on the findings made based on the classifications obtained through the GZ website.

b) Access restricted to Team members

v) Mailing list archives: The mailing list is the primary method of communication within the team and is only open to team members and the PI

vi) Funding applications for different agencies such as the National Science Foundation (NSF), National Aeronautics and Space Agency (NASA) and Leverhulme Trust have been obtained which provide a summarised formal overview of GZ's aims and how the initiative is presented to funding bodies. Comparing these documents reveal how GZ's representation varies corresponding to the different kinds of funding bodies. It is also one of the very few documents that require GZ to explain its formal structure.

vii) Presentations made by PI and TL to various members of the public, scientific societies and at academic conferences not only helps to trace the developments of GZ's work over time, but uncovers varying representations of knowledge that are worthy of comparison.

viii) Other documents such as publication rules and formal meetings' reports.

Essentially, in addition to the publicly available data, I was allowed access to all the documents created within the team and available to all the team members. The sheer amount of data was indeed overwhelming, and therefore the principle of multi-sited ethnography based on operational modes of following the object of study - the practices, expertise and distributed intelligence as GZ team members pursued their own knowledge objects, helped determine which documents would be most relevant for that instance of inquiry.

C) Observation: The third method employed in data gathering was 'naturalistic' observation of participants, sites and practices (Punch, 1998). Here events and situations were observed as they unfolded. Concepts and categories emerged in discussion and these were noted. I did not attempt to undertake a detailed discourse analysis of either talk or text. Since GZ exists because of the affordance that communication technologies allow, there were only rare occasions when members met in person, therefore opportunities of observing them as a team were mostly through analysis of the mailing list. However, as I have already indicated, I was able to attend one major meeting in the US in April 2010 and one in Oxford in April 2011. At these meetings I collected papers, documents and made field notes.

There was an unexpected turn of events at the first advisory meetings in the US institution in April 2010 that was relevant for data collection. Due to the eruption of the Eyjafjallajökull volcano in Iceland, flights around Europe were suspended, which resulted in the Oxford-based Core Team and some UK-based GZ members being unable to attend those meetings. There was a state of mild panic and chaos at the host institutions as this was a very important meeting for the Zooniverse projects in terms of providing progress reports of their work in education and development. These reports could determine the extension of future funding contracts. Since I was attending an academic workshop in New York immediately before the event took place, I was left being the only participant from Oxford present at those meetings for the first two days. The US-based Education Lead (EL) requested my help in running the online presentation management software for one of the parallel sessions of the meetings in which the PI, the TL and the Science Lead (SL) were supposed to be presenting and attending. Therefore, I found myself assuming a role of a participant observer: rather than being a non-participant, I became someone who was actively helping run the meetings on a technical front. This sudden turn of events was an opportunity for me to be introduced to the advisory team and the rest of the GZ team members as someone who was not only 'studying the GZ team' but also the only 'representative' from Oxford who would be monitoring the online meeting management system.

Therefore, not only were there unexpected events and demands on the GZ team, I found myself in situations during the data collection process where I had to adapt to unprecedented events that changed my role as a researcher in the process and enriched my experience of being in the field.

4.4 Analysing the Data

This is a study of emerging practices and expertise, which was evolving in response to changing demands and unfamiliar and often unprecedented situations within GZ and the analysis undertaken here has been a process of successive attempts at making sense of each data source and their inter-relationships.

The study is, therefore, characterised by continuous analysis and periodic reviews of the conceptual framework designed to identify and understand the knowledge object[16], following Knorr-Cetina's (1999, 2003) approach in analysis of STS cases) I was following. As I have already indicated, the knowledge objects for me in the present study are the practices and expertise exhibited in activities of the GZ scientists. The next two chapters discuss some of those activities and practices within GZ in detail in order to identify the expertise that was being employed.

4.4.1 Identifying the Knowledge Objects For Analysis

For the purpose of clarity, I need to distinguish between the knowledge objects or objects of enquiry for me as a researcher, i.e. expertise in activities in practices, and those that took forward the work of participants in GZ. As a researcher I needed to be able to recognise the knowledge work they were doing

[16] The knowledge object referred in this chapter is different from those discussed in the next chapter. In this chapter, it refers to the practices and expertise within GZ which is what this research examines. Whereas, the knowledge object in GZ may refer to the Astronomical objects or phenomena to be scientifically investigated.

as they pursued their own objects of enquiry in GZ.

The scale and ambition of the study of GZ grew over time, as I began to recognise the expertise that was being developed as participants worked on the knowledge objects of GZ. Initially my intention had been simply to understand how GZ functioned as a citizen science project in terms of the structure and processes. After several meetings and interviews with the PI, it became very clear that in GZ there was no formal or established way of organising tasks, people and the resources; there were no rules. The GZ team managed their research in response to the situations they appeared to be in. Such serendipity is therefore necessarily reflected in the way this study has taken shape. The feedback between research design and analysis in this study became a regular research activity with periodic reviews of my conceptualisations of the work of GZ. This iterative process enabled a responsive understanding of the knowledge objects being pursued by GZ as they continually unfolded in front of me while I tried to access the intentions and preoccupations of the GZ participants.

As the study progressed I a) narrowed my focus while trying to remain alert to changing demands being made on the people I was following and b) focused on how they worked with those demands and the influence on the development of expertise in the team. At different points in the process I had to conceptualise what was going on in order to ascertain some kind of stable analytic platform, from which I could survey what was happening, identify focus points for future work and move on. Analytic notes were made throughout the data collection process and were discussed regularly with my supervisor, and, occasionally with the PI while discussing the innovative practices they had to develop while tackling new challenges and how their expertise was being shaped with a deeper understanding of the varied expertise of their volunteers (see Appendix B and C). Those discussions were critical in the overall analysis of the data and have defined how this study proceeded. Recordings of those meetings were made, which enabled reflective analysis at different times and was of significant help in keeping track of the knowledge object of the research study. Therefore, my understanding of the knowledge object for this study was also being constantly updated and refined during the data gathering process.

4.4.2 The Analytic Process: A Flexible Focus

The analytic process in the present study placed the researcher's interpretations at the centre. As I worked with the complex intertwined data over time I found the following definition of analysis helpful:

"The isolation of what is more elementary from what is more complex by whatever method." Dictionary of Philosophy and Psychology, 1925

This maxim was adhered to from the beginning of the study. As shown in Figure 2A, the first significant analysis was carried out while creating a conceptual framework in March 2010[17], by separating components of the structure with the data at hand, and the processes that bound the system together. This exercise was aimed at isolating the constituent elements of GZ's knowledge production

[17] This conceptual framework was presented to early career and senior academics at the workshop on "Transdisciplinary Methodologies" at the NSF Web Research Workshop at Rensselaer Polytechnic Institute in New York in April 2010. I was the only doctoral student there and the only presenter. It resulted in some useful feedback, such as how the items in the index could serve as codes for analysis.

system in order to determine the direction of the research inquiry and the knowledge object of interest for this study.

Conceptual models like Figure 4B became reflective platforms, which provided jumping off points for the next stage of inquiry. The analysis of the knowledge objects of the team and of the research as they unfolded, provided pointers for directions that led to a deeper understanding of those objects, reflecting what Knorr-Cetina (1999) proposes in her ethnographic study of High Energy Physicists at CERN (and discussed in chapter 3).

As chapter 6 will show, there were continuous processes of translation (by which I mean the activity of transforming or modifying a GZ knowledge object) going on within GZ. I had initially intended to follow one set of processes of translation i.e. simply one GZ's knowledge object as it moved over time. However, I soon discovered that would mean that I would perhaps need to ignore other data that was relevant to the analysis on translation activities that came my way through the serendipity of GZ. What became increasingly intriguing to me were the activities involved in the translation process (for example, designing the primary task from the raw data into a form that can be uploaded on to the GZ site), and the need to build up a broader picture of what the activities were and the expertise involved in the process. That insight showed me that I needed to follow key actors and their activities, rather than simply one process, and track what they were trying to accomplish in those activities. This insight confirmed that the cultural-historical emphasis on following actions in activities in practices would be fruitful. The key informants were being extremely responsive to what was happening around them and the unexpected demands that were made on them by GZ..

The need to examine translation activities, i.e. what happened as GZ knowledge objects, such as data, moved across work settings, was also reinforced by three events; a) observation of the PI's role in the first annual advisory meeting in Chicago in April 2010; b) the co-conceptualisation exercise conducted with the PI to list and place his responsibilities and roles in GZ; and c) observing his assumed role during the conference organised by GZ team in Oxford later in 2010. Therefore, following his work in GZ as he facilitated, mediated and recognised the need for translation activities through different methods such as interviews, blog posts, internal e-mails and Twitter posts and a detailed analysis of his practices and expertise became crucial. The detail will be examined in chapters 7 and 8, which focus on the expertise being developed in the core GZ team.

The TL, had a more defined role in the development team when he first joined GZ two years into the project. As already explained, the decision to follow his work was made as his role began intersecting with the PI's and with his increasing involvement in managing the activities in the Zooniverse.. The changes in his work and his growing expertise has been fascinating to follow over the period and will be discussed in chapter 8.

The evolution of roles, responsibilities and expertise within the Core Team was unexpected, and reflected unpredictable developments in GZ. Changes in its pace, the environment surrounding it, the progress being made and the situations that required instant solutions were continuous. The lack of structure, a deliberate decision on the part of the GZ team, allowed for the responsiveness with which they could solve problems and pursue new ideas. This was analytically challenging for me since the knowledge object that my participants were pursuing were also not clear to them and it was possible for them to construct and follow any number of directions. The literature on research methods that was potentially most relevant in this kind of context was, as I have already indicated, from e-Science (Beaulieu 2005, 2010; Hine 2006;).

However, as useful as their work has been to my conceptualisation of the study, GZ does share similarities with e-Science projects, but it also differs in nature in many aspects. The present study was conducted in a highly mobile environment where new ideas were being developed and influencing established methodologies and techniques of data analysis within Astrophysics. I could not, therefore, rely entirely on suggestions put forward for an e-Science methodology and analytic process. The study needed an approach to design and analysis that could particularly capture the dynamism of the context in which the knowledge object was being followed. Jirotka *et al*. (2013) recognises that while approaches to early e-Science was directed towards the process of research as workflow, those conducting more detailed workplace studies have to have the ability to reveal a diverse set of practices and expertise that transcend boundaries of disciplinary practices, which poses challenges to fieldwork.

This need is recognised by Rossman and Rallis (1998) who suggest that open-ended flexibly shaped studies, of the kind I began to develop, allow for the analytic direction to emerge as the research progresses. Their argument is that highly organised data gathering and schemes of analysis often do not reveal the unusual and pose limits on flexibility, hindering the serendipitous - “the paradox that can lead to insights”. Nonetheless, GZ was found to be quite complex and fast changing. I found that I needed to limit the open-endedness of the design and use periodic points of analytic reflection to bring some stability to the enquiry and to shape the next phase of both data collection and analysis. To that extent analysis was on-going and the reflective platforms and ensuing discussions was centrally important to the analytic process at every stage of the study; a design that was deliberate, not only following phronetic principles, but also giving validity to my interpretations.

The analysis of the data gathered for this study began with ongoing data reduction by editing, segmenting and summarising the data as an when the data was collected from various sources and methods, for e.g. interviews were transcribed as they were conducted and the data was reduced by what Kvale (1996) refers to as .‘meaning condensation’ (Kvale, 1996), which requires the researcher to compress long statements to represent the main gist of what said or reported. This process was followed by coding, which is essentially categorising data as concepts and themes emerge from the data. According to Punch (1998), coding calls for tagging of data that beings with the descriptive and low inference ones of higher inference. Along with coding, relationships between and the patterns within the themes and concepts from the data were continually identified and refined, a process referred to as memoing (Punch, 1998). These analytical processes call for continual interpretation and requires finding meaning beyond the verbatim (Kvale, 1996); the notes taken during the interviews and the analytical notes taken throughout the period of data gathering were particularly helpful during this stage. While coding, memoing and interpreting the data are carried out simultaneously, these analytical activities were carried out in regular iterations as more data was collected, which Punch (1998) suggests refines the analysis. The following stage of displaying the data entailed assembling information in an organised manner to draw conclusions (Miles and Huberman, 1984) and conceptual maps that were co-constructed with the PI and the TL at periodic intervals were an integral part of the analytical process during the data gathering process and again, after the data gathering was compete. The final stage called for organising and integrating piles of memos to achieve coherence and meaningful data presentation (Punch, 1998). At this stage, memos from data gathered from all sources and methods were consolidated and compared for consistency in interpretations that were presented as findings.

As Rossman and Rallis (1996) suggest, the process of analysis begins the time of formulation of the main research questions. The analytical notes that were made periodically provided pointers for reflection after every two or three interviews with the PI and the TL, which led to changes and improvement in focus of the interview data and other sources of data analysed. Therefore, the pursuit of the the knowledge object (Knot-Cetina, 2001) of this study was carried our by continuous analytical process throughout the research period, by iterative consolidating and refining emergent concepts that helped answer the main research questions.

4.4.3 An Example of Following a Knowledge Object in the Research Study: The Role of Citizen Expertise in Science

As I followed the PI and the TL and their preoccupations to elicit understandings of their expertise, other highly relevant knowledge objects emerged for me as a researcher and I needed to limit these explorations. However, the role of citizens in citizen science was so relevant to the demands on the expertise of the GZ team that it was judged a legitimate focus. The analysis of the role that the public volunteers (Zooites) play in the scientific process that was designed for their participation was carried out at various periods, as the need for such a focus became more evident in the data. At the same time as the expertise of the PI and TL as a knowledge object for the study was being identified, an early interview (with the PI) partially reproduced below, provided an analytical direction. This insight helped the conceptual framing of the role of the public in GZ, which I shall later argue as epistemic subjects:

13 February 2010

"If we are not careful, you just break the boundary between experts and non-experts. How do you distinguish between what's actual science and what's...science. That is one of the challenges we have got, I think." - PI

As I was reflecting on this excerpt from the PI's interview, I was provided with an internal document (section presented below in Figure 4A) by the Education Team of GZ prior to the first annual advisory meeting in April 2010, made mandatory by their biggest funding agency. It outlined in successive stages, the generic categories of activities within a scientific process and how the public may be involved in it.

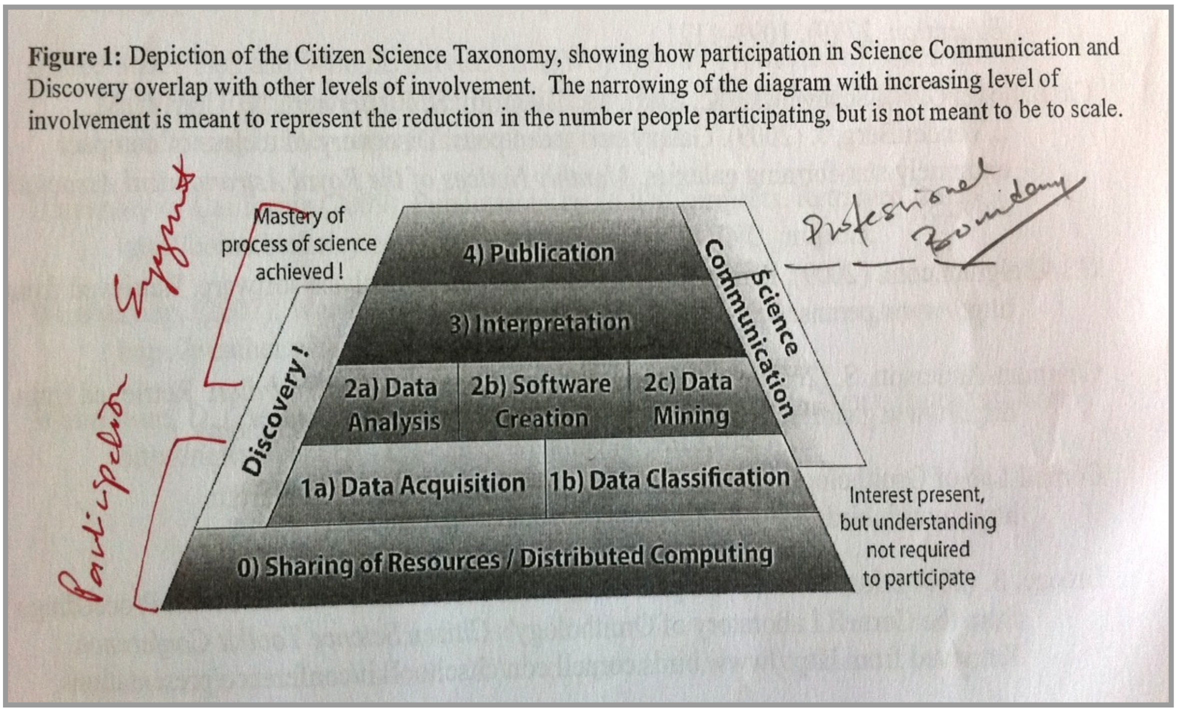


Figure 4 A: Snapshot of the Citizen Science Taxonomy

(Presented in an Internal Document by the GZ Education Team for the First Advisory Meetings in April 2010)

The intersection of the challenges relating to expertise and the model of scientific expertise outlined by the GZ Education Team led to the following form of analysis, which not only positioned the scientific tasks on scales of levels of engagement and understanding, but helped make visible the boundaries of public expertise (as identified in the first conceptual framework of GZ in March 2010 in Figure 2 A).

In parallel to my analysis, discussion on how UBRETs (User-Based Research Enabling Toolkits) should be designed on the GZ platform was underway in the internal documents and within the Core Team. These GZ discussions brought attention to the question of how far the public can go in terms of the scientific process. The continuous discussions over parallel sources of data on the varying expertise of the volunteers in the scientific process sparked my rethinking of public as epistemic subjects, a topic which is discussed in more detail in chapter 6.

As Rossman and Rallis (1998) observe, during the process of immersion in the field, the researcher's task is to use field notes, interviews, images, the noting of thoughts and emotions, in order to identify recurring ideas, salient themes and patterns of belief that helps one respond to the research question. Multiple data sources of these kinds provided me with simultaneous pointers to how the knowledge object for the research was unfolding. Above all, I found that the practice of generating analytical notes and reflecting on them throughout the data gathering process enabled my appreciation for the complexity inherent in the process of researching dynamic entities. The quality of engagement that I experienced was parallel to that of my participants who were in pursuit of their own knowledge object, which resonated strongly with the notion of engrossment proposed by Knorr-Cetina (2001), which she contends is the emotional basis for research work.

4.4.4 Synthesis

From the very beginning of the study, it was always an aspiration to present the complete picture of what makes GZ work by making visible the practices, expertise, tools and the environment it operates in i.e. a synthesised analysis.

What I mean by synthesis is reflected in the following definitions:

"The explanation or reconstruction is often then exhibited in a corresponding process of synthesis"[18]

"[i]n philosophy, the combination of parts, or elements, in order to form a more complete view or system. The coherent whole that results is considered to show the truth more completely than would a mere collection of parts"[19]

These maxims suggest that synthesis complements the process of analysis by establishing relationships between the elements of a system that had been previously isolated into separable entities for further clarity in order to understand their nature. In this study, the concept of collective intelligence played a defining role in selecting the analytical approach. The concept, discussed in chapter 1, is borrowed from Malone *et al*.'s (2009) work in organisational studies suggests an approach to research that aims to understand how humans in all their capacities and the technologies they use can perform more 'intelligently' as a whole, rather than seeing technology as merely an addition to their individual capacities. This approach allowed me to keep the bigger picture in mind during the analytical process, which helped me to remain aware of the various elements, forces and entities that might have played a mediating or central role in what I was then focusing on.

Chapters 5 and 6 present the constituent entities (organisations, technologies and stakeholders) and the activities that are GZ's responses to them in order to run a successful citizen science project. It then highlights translation as a crucial practice in the movement of GZ knowledge object along the process of

[18] http://plato.stanford.edu/entries/analysis/
[19] http://www.britannica.com/EBchecked/topic/578628/synthesis

contributing to science. The task of chapter 5 is therefore, to isolate the elements and processes of GZ making them more visible analytically separable objects. The next chapter in which findings are presented (chapter 6), examines the knowledge object for the research study (activities in practices and expertise in GZ) as they unfold in the work of the Core Team, the PI and the TL. This chapter is informed by the analysis carried out for chapter 5, but augmented by a particular analytic focus on what the translation processes are and what is demanded of the GZ team as they enable and take part in them. Here the analytic concepts of expertise and working relationally with others, outlined in chapter 7 and 8, will be particularly important as ways into the data. In chapter 9, the isolated elements and activities within GZ are brought together under a collective intelligence framework for a holistic understanding of the complexities of GZ work, revealing how GZ is characterised by the dynamism that drives it forward in response to public engagement. In this way the answers to the research question have been organised through a continuous analytical process driven by attention to the knowledge object of the GZ team in order to reveal with some clarity the knowledge object of the study.

4.5 Ethics: Researcher Reflexivity and Responsibility

The role of the researcher in how the study was shaped during the process of conducting the research was a continuous concern for me. Beaulieu (2010) suggests that the epistemic power of the ethnographer is derived from the intertwined activities of defining the ethnographic field, pursuing fieldwork and writing. Mortensen and Walker (2002) hold the view that the contiguity of the ethnographer's activities and of the field, demands an awareness and attention to the mutuality of the field relations, which requires ethnographers to posses particular skills to modulate those relations. From this perspective, which is one I have also come to assume, the researcher has considerable power and responsibility.

It is also unimaginable as a researcher today to not use the Internet. While online spaces offer new opportunities for studies in the field of Social Sciences, Schroeder and Bailenson (2008) suggest that this poses new challenges that are different from the previous modes of research. Enyon *et al.* (2009) emphasise the need to rethink the ethics of online research while suggesting the importance of the context of the study, what the study aims to achieve and the methods that are used, to remain unchanged. There are, however, issues that can be more challenging in online settings, such as the difficulty of determining whether the potential participant is competent of giving informed consent (Kraut *et al.,* 2004). There are also issues around privacy of the participants as illuminated through various cases such as re-identification of individuals through analysis of benign data fields such as zip code (Sweeney, 2002) or movie ratings (Narayan and Shmatikov, 2008). Buchanan and Zimmer (2013) further contend that the rise of sharing of information on social networking sites has compounded the complexity of privacy issues in Internet research.

I was extremely fortunate in the way the study was embraced by the GZ Team, particularly the PI and the TL. Not only did they agree to be interviewed on a regular basis at my request, they also communicated information about my research to the rest of the team in positive and encouraging ways, as I found out when I met other team members at GZ events. On several occasions, when I was concerned about whether reflecting on the activities and bigger goals of GZ was of any help to them, both the PI's and the TL's response to the concern was of encouragement, as they not only enjoyed some of our meetings, but in the activity of recalling and reflecting on their practices, they were reminded of the bigger picture of GZ by stepping back from day to day tasks. After a few months of interviewing the PI and the TL, they started referring to the interviews as 'therapy sessions' with each other and other members of the GZ team, which I believe was a positive description. Analytical data collection exercises such as the PI's mapping of all the activities in GZ (See Figure 4 B) is an example of an activity which was found to be very useful for the PI, who wanted to have a clean version of the map once it was ready.

In addition, in April 2010 the advisory committee had suggested to the GZ team, that they should document their activities and progress in reports and visual representations. Consequently, my analysis and input was regarded as helpful to the GZ team in meeting this suggestion. The final analysis of activities, practices and expertise (Figure 4B) was presented to the PI in time for their next scheduled meeting in late 2011.

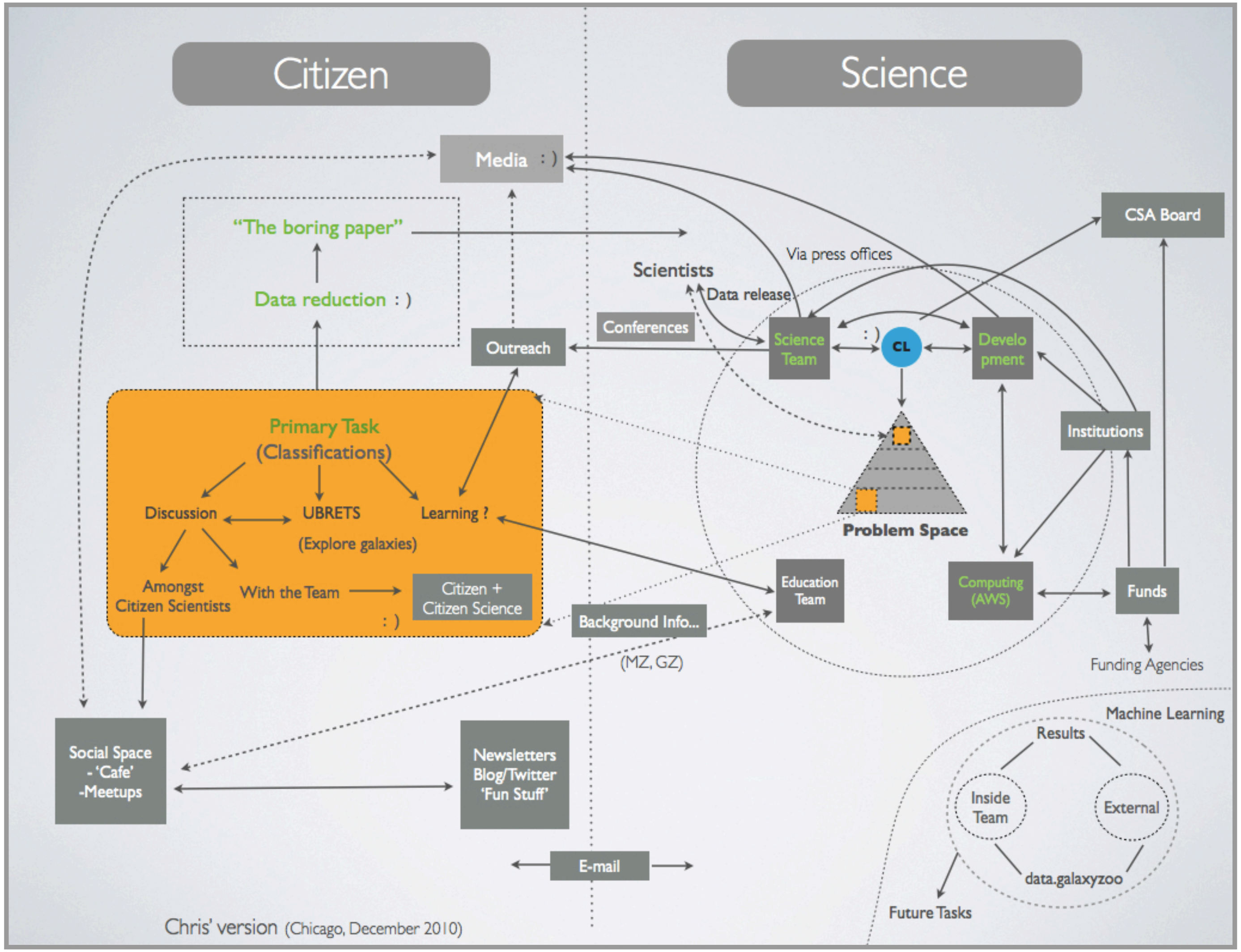


Figure 4 B: The Principal Investigator's Mapping of Galaxy Zoo's Activities

My concern was to do good Social Science and be useful to the team. I am very much of the view taken by Flyvberg (2001: 4) that

> “[t]he goal [of social scientists] is to help restore social science to its classical position as a practical, intellectual activity at clarifying the problems, risks, and possibilities we have as humans and societies, and at contributing to social and political praxis”.

The anonymity of the participants was impossible to achieve without losing many of the characteristics that define the project. This has been discussed with the PI, who is the principal gatekeeper, who pointed out that anonymity strips the study of its value. The problem of anonymity can be illustrated with a simple example: “The participant was the PI of a successful online citizen science project in the field of Astronomy.” Removal of any of the terms, not only distorts what the project stands for, it also takes away the key distinguishing traits without which GZ cannot be defined. Especially, since GZ is the most successful online citizen science project so far, it really does not take effort to trace the ‘anonymised’ project or PI.

The three findings chapters not only serve as a historical account of GZ’s evolution, but also an analysis of what configurations and practices were operational and how they led to certain decisions and events. It was established that having to think about the project as a whole has been helpful to the PI in how he understands how the project works.

While Busher and Clarke (1990) contend that total ethical research is impossible to achieve, there is a strong consensus in the research community over the need for careful attention to ethical considerations in the research process. This research was carried out under the British Educational Research Association (2004) guidelines and approval for research was obtained from the Central University Research Ethics Committee (CUREC) of the University of Oxford.

The following are some of the ethical concerns that were reflected upon during the study:

a) Process of data collection: Prior to the interviews and observations, informed consent was obtained from all participants explaining the purpose of the study, their right to anonymity and withdrawal from the study at anytime. The responses of all participants were anonymised i.e. names were not attached and quotes were checked and agreed before the final presentation. As already indicated, the value of anonymising the PI's responses was problematic. His view was that his contributions should not be anonymised and he has been very vocal about instances, which should not be quoted. Therefore, the level of disclosure was determined by discussion with the PI on sensitive issues.

b) Feedback from the participants: The Core Team of GZ have been very supportive and two of them have explicitly expressed their interest in the nature of this study as this is an unusual undertaking for them and the field of science. After several interviews and meetings, they found it helpful to talk about the bigger picture of GZ, which they admitted to at times have taken for granted and ignored, as day to day activities took most of their time. Some of them found reflecting on the questions very helpful as they reminded them of their overall objectives occasionally.

c) Dissemination of data: It is my responsibility to represent the account of work in GZ in ways that do not undermine participants' status as professionals. I discussed the findings with the PI, TL and other team members and gave them the opportunity to comment on my over all account as well as tackling the accuracy of their specific contributions.

4.6 Challenges

There were numerous challenges encountered from the beginning of the study, which made the study into a continuous learning process in many respects. Since a largely successful online citizen science based project has not been studied through a Social Science lens to date, this was seen as an opportunity to design a study drawing from the strengths of several research areas such as Organisational Studies, STS, e-Science studies and Innovation Studies. Formulating the conceptual framework at the beginning of the research was driven by the concepts that were of interest, and during the later stages of data collection these concepts were questioned, critiqued and refined through theoretical and empirical analysis.

Following the innovative practices of a scientific undertaking required a considerable amount of flexibility in data collection primarily because no events were predictable and were shaped by forces and factors that would not have been foreseen. Also following actors using an ethnographic approach called for periodic interviewing with the same participant from the first interview in February

2010 to the last follow up interview in April 2013, which requires for preparing questions of topics of enquiry that the participant is willing to engage in. Serendipitously, as the fieldwork began, GZ was developing as a successful project, which resulted in practices, decisions, technologies and structural changes to evolve, since there was a possibility that the initiative would not yield success or could no longer be sustained.

The other considerable challenge was to develop basic domain-specific knowledge in the field of Astronomy in order to have uninterrupted conversations with the participants. Therefore, it was necessary to learn the basic science behind GZ and the technologies that it operates on. It was also necessary to consult with colleagues and contacts in the field of Astrophysics to understand the implications of GZ's work on the field itself. I had the fortune of being able to get clarifications and judgments on the developments that GZ was making in its own field, which was outside my area of expertise.

It was a constant process of negotiating with myself on how much I needed to know for the purpose of this research. Although it may be easy to worry about the lack of domain knowledge, the position does privilege a perspective which is different than that of the participants, which was the perspective taken by the PI when asked about this concern. This stance of a non-expert on occasions inspired questions that required the participants to reflect on their own understanding of what they do. Therefore, it was necessary to build 'relational expertise' as the capacity to take their standpoint in understanding the motives in their practices (Edwards A, 2010), and be able to engage in discussions about those practices. This is one of the several parallels between the approach in relating with others, to my participants for this study and to volunteers and non-scientists for the PI and the TL. To help build the relational expertise necessary for this study, I took upon myself to attend events such as talks, seminars, workshops and conferences which developed my understanding of the field of Astronomy and Astrophysics research, which was challenging and thoroughly enjoyable in its own right, in addition to what it meant for this study (see Appendix C). The openness and enthusiasm of the participants, it may be plausible that the personal interest and basic knowledge on my part played a role in the level of access I was allowed as it is a reflection of my commitment to understanding their work.

As Beaulieu (2010) recognises, a challenge in mediating settings is that ethnographers need to align themselves to actors to gain access to what she calls the ‘backstage of infrastructures and mine of log files’. The researcher requires some form of guidance to identify what those resources are and how they are being used by participants. In the present study, the PI was indispensable in grasping the sheer network of contacts, diverse types of data managed and the work in progress of GZ. Nevertheless, the challenging nature of this research has predominantly been the vast array of data available in the backstage and the need to navigate a pathway through it. Here, my regular analytic framings proved invaluable again in refocusing and redirecting attention to the object of my study.

4.7 Concluding Points

As new forms of communication and media support new modes of organisations, scientific practices are moving from traditional lab-based sciences towards disciplines with computer-mediated practices at their core (Hine 2007, Knorr-Cetina 1999). Simultaneously, society is demanding science to be more transparent and engage in communicating what scientific knowledge and practice entails (Nowotny 1999, 2003). This research is an exploration of an emerging mode of scientific community which thrives on communication media to not only sustain itself, but uses it to involve the public in production of scientific knowledge. The combination of those features and practices make GZ a fascinating and worthy case for research. GZ can be considered as what Frickel and Gross (2005) define as Scientific/Intellectual Movements (SIMs), which are collective efforts to pursue research projects while resisting forces from within their community to form critical mechanisms for generating new ideas and knowledge. Not only does this study aim to conceptualise and trace the developments of such organisations, but as a result of this undertaking, it aims to contribute towards newer methodological directions, which as Hine (2007) argues have a capacity to take us to exploration of new theoretical orientations that makes them topics in their own right.

Chapter 5

An Analytic Account of the Development of Galaxy Zoo: Constructing the Epistemic Architecture

5.1 Introduction to the Knowledge Production System of Galaxy Zoo

The aim of this chapter is to present an analytic account of the development of Galaxy Zoo (GZ) from its initial configuration of a few postdoctoral researchers operating within a single university department to a multi-project citizen science collaboration comprising researchers across numerous institutions working with more than 400,000 volunteers from around the world. The process of development of GZ from a single project to the team being recognised as the experts in building citizen science projects reveals the complexity in the interactions between the scientists, technologies, institutions and the public.

In chapter 2, a descriptive account of the origins of GZ was presented along with the initial conceptual map where its constituent components were identified which established the background for this chapter. Building on that background description, in order to understand how GZ has evolved since its launch in 2007, this chapter shows how its undefined structure began to emerge in terms of the organisation of teams shaped by their roles and responsibilities, with intersecting activities and practices. Each team's primary responsibility and purpose are outlined and compared through illustrative examples of the various activities that members of different teams engaged in. In doing so, the few rules that were established as experiences of working with each other grew, are also highlighted. The analysis of the roles and responsibilities of teams will reveal that although different teams focus on different aspects of GZ knowledge objects, it is the Core Team that is responsible for making the most important decisions about how GZ proceeds in conducting citizen science. The expertise that the Core Team has developed through their innovative practices making its citizen science work will be presented in detail in chapter 8.

5.2 The Work of Galaxy Zoo

At the beginning of the data gathering process, GZ comprised four teams: a) Science, b) Development and c) Education and finally, d) the Core Team, which was primarily responsible for project management. The Core Team consisted of the Principal Investigator (PI) of the project; the Science Lead (SL); the Technical Lead (TL); and the Education Lead (EL), however only the PI and the TL were employed full-time to do GZ work. While the PI and the TL were based in Oxford at the time fieldwork began, the SL was based in another city in the UK and the EL in the US-based partner institution which led the grant for the education research in GZ.

The Zooniverse is also overseen by the Core Team and is the collection of citizen science projects which were built after the initial success of GZ. By 2013 there were 10 live projects under the Zooniverse and the website claimed that it is:

"[t]he home to the Internet's largest, most popular and most successful citizen science projects".

All of these projects were managed and developed, but with different science teams and principal investigators. However, this study focuses primarily on GZ and only draws upon the wider Core Team activities when relevant to GZ.

In order to understand the ongoing developments of GZ during the fieldwork year, in December 2010, seven months after the production and validation of Figure 2A I worked with the PI to co-construct a model of the system which is shown in Figure 4B. The purpose of this exercise was to reveal the activities that were required to run GZ successfully and identify the GZ members involved in it. The PI was the only GZ member who knew everyone else involved; while and his position as project manager from GZ's inception offered him a holistic view of the project. The description outlined in Figure 4B is based on GZ Hubble, the third iteration of the first version of GZ and there are some significant changes since Figure 2 A was produced in April 2010. These changes include how the teams were organised and the practices that were needed for GZ to maintain its status as a successful citizen science project.

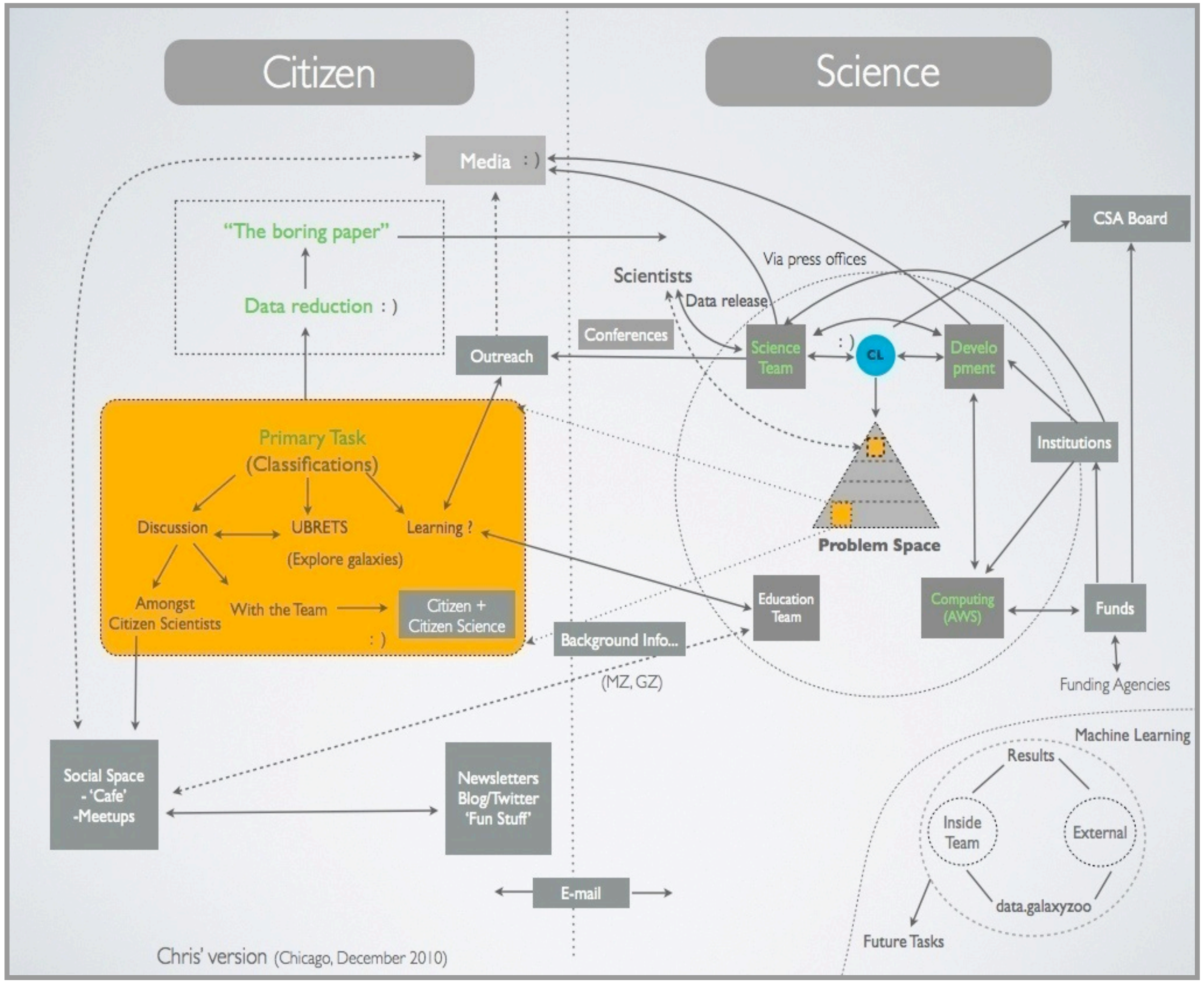


Figure 4 B: The Principal Investigator's Mapping of Galaxy Zoo's Activities.

Key: CSA Board: Citizen Science Alliance Board; UBRETs: User-Based Research Enabling Toolkit

To begin to capture the development of GZ, let us start with the right hand side of Figure 4B and the scientific demands of GZ. The PI of GZ would usually be approached with a project idea which would then be discussed with the Development Team to determine the level of difficulty in building the project. That level of difficulty is understood by the GZ team in terms of producing the primary task, which is the basic activity the volunteers are asked to carry out. The primary task is the most basic question that can be asked to the volunteers, which would eventually lead to data that can be used to publish a scientific paper. The triangle on the right-hand side of the diagram represents the hierarchy of tasks in scientific process, whereby the top-level activity is to produce a scientific paper.

Prior to determining if the project was worth building, the Science Team would identify one of the papers they would be writing. All projects would be required to produce a scientific paper, even in the absences of new results a negative paper must be written. It was one of the main criteria by which a project idea was determined to be worthy of building by the core GZ team. A project might be built on an existing model; for example, the Milky Way project was adopted from Moon Zoo, an existing project which was built after GZ 2. If the existing resources were insufficient, then a grant proposal for additional funding would be submitted. Most frequently, new projects depended on the availability of time and resources of existing GZ members.

Once the primary task was completed by the volunteers, the data would be fed back into the project database and would undergo the process of data reduction by a member of the Science Team. For instance, if 10 volunteers have classified a galaxy and seven have determined it to be elliptical, the bias would be measured and an expert evaluation made of that classification. Most projects, including GZ have discussion forums. Several unexpected outcomes have come about as a result of having these forums for the citizen scientists, as engagement in a forum exceeds the demands of performing the basic primary task. Scientists have benefited from these discussions as well. For instance, a discussion thread on overlapping galaxies which was being discussed by the volunteers caught the attention of a GZ member, who later wrote a paper on the topic; although this topic was not part of the primary task that had been identified by the Science Team.

The Education Team in GZ had a strong interest in the left-hand side of Figure 4 B and carried out research on whether or not any learning occurred as a result of public engagement in the project. Questions about whether attitudes towards science were changing were also addressed, amongst many other questions about the benefits for the volunteers. The GZ Development and Education Teams had built what they called User-Based Research Enabling Toolkits (UBRETs), which allowed the users to discover more about the galaxies that they had classified. The project had received funding from organisations such as the National Science Foundation (NSF) in the US for its outreach and science education potential. In addition, a completely unexpected use of the discussion space by the volunteers was its use as a social space for personal interactions, which were completely separate from the primary task. However, there were also volunteers who simply sustained a serious interest in the scientific aspects of the project. An example of the latter category was a volunteer who built his own site to explore a new object, which was previously discovered by GZ volunteers, and designed his own primary task to study the mergers of galaxies.

Back at the right hand side of Figure 4B, the Science Team also served several functions in the working of GZ. The interaction between the Science Team and non-GZ scientists was established through the academic papers published. Members of the Science Team gained and lost membership on the basis of their commitment to public scientific papers from GZ data. Therefore, the Science Team was the most dynamic and permeable in terms of the flow of its members. Conferences were also another means by which GZ scientists interacted with other scientists and were spaces where new collaborations would potentially form. Although the commitment towards science communication in the public sphere varied between members, the Science Team was responsible not only for producing papers, but also communicating about GZ within the scientific community and the public. The Science Team also provided inputs in the discussion forums of GZ. Their participation in the discussion space not only inspired the second version of GZ (also known as GZ 2, which emerged in 2009), but also led to the increased difficulty in the primary task which was prompted by the discussion between the volunteers and the scientists. The effort put in by the Science Team in outreach activities were not anticipated by the PI. Members of the Science Team were active on social media platforms such as GZ blogs, Twitter and GZ newsletters which were specifically targeted towards GZ volunteers.

Funding agencies and institutions also played a critical role in supporting the conditions for GZ to be a successful project. Academic and research institutions not only offered financial support for the project, but they also hosted the Science and the Development Teams through their affiliations as professional scientists. The funding bodies and the research institutions therefore provided, according to the PI, the professional infrastructure of science, which is what distinguished the work of science undertaken in the public sphere from scientific activity.

One of the other bodies fundamental to GZ and the Zooniverse was the Citizen Science Alliance (CSA) Board, which was responsible for signing off funding applications and approving new projects. It consisted of the directors of all three teams - Science, Development and Education. Another major component required was computing. This was hosted by Amazon Web Services, and was initially supported by Johns Hopkins University for the first version of GZ. The role of media was seen by the PI as indispensable in the success of citizen science as exemplified by GZ. Their role ranged from promoting the launch of the GZ in 2007 on its very first day to having regular media coverage of the science developments, for which the Zooites were called upon to share their experiences. The media not only helped the GZ team to communicate with the public, but also drew the attention of scientists who were potentially interested in collaboration; serving functions in both the science and the public spheres.

The aim of this section was to uncover the interactions between the different components in the context of GZ, revealing the changes in ways of organising since its inception. The following section will describe the ways in which the main teams, also referred to as divisions, developed and the principles and very few rules that governed their practices.

5.2.1 The Management of Galaxy Zoo

The way in which the project was managed was interesting to follow owing to the conspicuous lack of emphasis on a planned design structure by the PI, who also functioned as the project manager. The PI described his role as managing "...the energies of a huge number of people and resources." He was the only member of the team who had a role in every team and had been involved since the inception of the project. As the project grew bigger in volunteer size and in the tools and functionalities on the website, his responsibility to provide financial support for the members. This work involved coordinating people and resources, writing grants and managing the financial resources and it increased significantly over time. He confirmed that there had never been a formal discussion about the roles in GZ and that there had been a conscious decision to avoid conflict and rigidity in the organisation's structure.

13 February 2010

"I mean, I never emailed and said I'm going to call myself a PI, I would never use that title. I'm almost certain that if you search the archive, there is no way that the title would come up. The exception to all of this is when you get to there is a whole Zooniverse citizen science, because there was an agreement between institutions at the stage, you have start to defining titles." - PI

Therefore, GZ roles only appeared officially when interacting with the environments external to the GZ project. As one of the founding members of GZ expressed:

12 March 2010

"I think we are a very horizontal collaboration, some people are very hierarchical, you know [the PI] is doing a fantastic job organising and putting things together"

– GZ Science Team member and Co-Founder

One of the defining attributes of how the project has been managed is by finding the right people to get the job done on a short-term basis. There were no strict and heavily protected boundaries that prohibited or inhibited participating and collaborating in the GZ project.

13 February 2010

“It’s not like an expanding company. It’s more like who do we have and what can we do.” - PI

An example of that is trying to get GZ Hubble to launch. The PI admitted that they had been trying to push the launch for quite some time but had not found anyone who had the time and the enthusiasm to put the data together. He found it natural in an academic environment to configure the group in response to what is needed. However, the point at which he sensed a significant change in scale of the project was the moment when the first full-time developer, the TL, began receiving financial remuneration for this work with GZ and recognised that he did not want to waste thousands of people’s time, and he became aware of himself becoming “more professional.” Therefore, with the awareness of increased accountability towards others, the feeling of becoming a professional team leader arose for the PI.

5. 2.2 The Science Team

The Science Team consisted mostly of early career scientists who intended to use the dataset from GZ to produce results and publish papers. They might have taken part in the different stages of the knowledge production process, but their primary interest was to pursue their scientific inquiries. In order to get the science done, firstly, they had to define the research they wanted to pursue and get involved in the discussion on constructing the primary task, which occurred mainly through the team mailing list. The discussions on decision trees involving the paths that the volunteers take, was an example of the negotiations that occurred in terms of designing a complicated primary task which was instrumental in how useful the dataset would be for the scientists. This activity required them to examine a part of the scientific process from the perspective of the untrained individual, where the chances of potential biases and errors are more likely when compared to a professional scientist.

One of the most contentious issues within the Science Team was managing the science questions that the members were interested in asking. An example of this is during the GZ 1, a number of members had expressed interest in working on red spiral galaxies, but not all agreed. There seemed no way for the situation to be resolved, when one of the members decided to write the paper without the team's full agreement. In GZ 2, a system was established to bid on science questions in advance, before launching the site. A document with 33 science projects was complied with a description and the list of people committed to the results. Despite this change, by the end of the data collection period, the team still had not established a successful method for allocating the science projects to be undertaken.

In addition to carrying out the research they committed to, several members invested time and effort to communicate with the volunteers through the official

GZ blog, official GZ and personal Twitter accounts and official GZ newsletters. Not all scientists were engaged in such activities, but there were various ways in which they were involved in engaging with the public other than just communicating through social media channels. Some of them attended volunteers' meet-ups in different locations in the UK. There were instances, as already indicated, of scientists following volunteer discussions which subsequently led to scientific papers or the discovery of new Astronomical objects. Therefore, volunteers were valued by some scientists in the Science Team as indicators of interesting questions which might have been overlooked by the scientists themselves.

5.2.3 The Development Team

As mentioned previously, the first version of GZ did not have a Development Team and favours were called in to build the first website. However, once the PI received a grant that allowed for a full-time developer to be hired, the TL at the time of data gathering was brought into the GZ team to build the second version of GZ. The PI strongly believed that the TL had the expertise and the vision to advance the initiative in ways he himself would not have conceived of. The TL's primary responsibility was to provide technical solutions for the ideas that the Science Team presented and negotiate with them what was possible to execute in terms of building a project, given their existing resources. GZ was also known to lead the way in the sub-field of Astroinformatics, which is essentially about using computing technology to build better tools to do science in Astronomy and this was one of the Development Team's secondary goals. The TL's interest in machine learning techniques was also critical in partnering up with Artificial

Intelligence (AI) experts to explore improved ways of how citizen science could be advanced in its technical methods.

5.2.4 The Education Team

The primary source of funding for GZ came from education and outreach grants, an innovative solution by the PI. By arguing that science education and communication was a central aim of GZ, they were successful in receiving grants that were directed towards scientific research while engaging in outreach and education activities. The Education Team, based at the US partner institution which was a planetarium, carried out research into understanding user behaviour and learning outcomes from the GZ projects. However, from several interviews with the PI, TL and the EL, it was found that the work between education research and the development of GZ was not as interlinked as one would have anticipated it to be, which raised questions about how well the expertise of the teams had been managed. The challenges in liaising to move the different teams towards the goals of GZ will be discussed in further detail in chapters 7 and 8 where the core practices and developing expertise of the Science and the Core Teams are identified and examined.

The previous sections of the chapter aimed to introduce and explain how GZ functions, the processes it supports and the teams it had organised itself into. The knowledge production cycle was described with its major components and activities made visible. The next section of this chapter brings together the various elements and processes of knowledge production under a single conceptual framework called *epistemic architecture*.

The first version of GZ only had one team, one primary task, some data reduction and a paper planned; and was hosted by a university server. There was no Development Team, no planned efforts for outreach and education and no funding bodies involved. The ways in which the interaction with the volunteers occurred was unexpected and so was the involvement of the scientists in the project. What determined how things were accomplished and how things turned out, depended on the time the PI, and later his other collaborators, had available to respond to the demands they faced. The consequent complete lack of strategy and planning is what characterises GZ's epistemic architecture as being flexible and responsive to the prevailing conditions in the process of scientific inquiry.

5.3 Constructing the Epistemic Architecture of Galaxy Zoo

Drawing from the findings of the study, this section brings together the essential components of GZ, conceptualised as practices, expertise, rules, principles, stakeholders and distributed intelligence (DI) which are interlinked in a complex organisation to propose a term – "epistemic architecture". In particular, the theoretical framework of collective intelligence (Malone *et al*., 2009) is applied to GZ in dissecting its critical elements and revealing its complexities. This process is necessary in order to construct a synthesised view of GZ's epistemic architecture that has continuously evolved, at least in part, as (i) a result of numerous forces external to the boundaries of GZ; and perhaps even more importantly, as (ii) a result of the actions of the actors operating at the core, who have been responsible for the decisions that have shaped the knowledge production system of citizen science in response to those forces. A critical function of conceptualising the epistemic architecture is therefore, to position the findings of the study within a framework that can reveal the elements that are the focus of the study and those that are not, in order for the complexity of GZ to be recognised and acknowledged.

The idea of an epistemic architecture is, in brief, an attempt to understand GZ as an innovative enterprise within the field of science, which successfully functions as a result of emergent design and serendipity, supported by the continually evolving skills and expertise of participants, the introduction of technologies and attention to how the volunteers acts as epistemic subjects. Collective Intelligence is therefore, a central concept in an understanding of the particular epistemic architecture of GZ. It is premised on the phenomenon of amplified intelligence as an outcome of humans and technological tools working together, rather than an outcome that is a simple aggregation of their individual efforts combined.

Epistemic Architecture in GZ, where collective intelligence is an outcome, is a term that describes the structure of how technologies, practices and expertise are positioned within a system and the relationships between the system's constituent elements i.e. entities and processes. It is the continually changing, dynamically evolving blueprint for a system of knowledge production. It draws on the theories of distributed intelligence and cognition (Pea, 1998; Hutchins, 1999) in terms of recognising conceptually distinct elements, agents and processes within a system that can be represented as the architecture of that system of knowledge production.

The rationale for putting forward the notion of an epistemic architecture is that it provides a big picture or a holistic view of the system, which allows us to focus on the elements and processes selected for this study while acknowledging the remaining elements that may not be the focus of the study - due to constraints of scope and time – but which are important to understanding the knowledge production system in its entirety. It is an exercise to construct the blueprint of the system which will allow the selection of the few elements which will be the objects of deeper scrutiny in the following chapters.

I highlight epistemic architecture as one of the central concepts in this study of citizen science, as it is the design of that architecture that allows for collective intelligence to be achieved. The epistemic architecture is the structure of a knowledge production system or structure of a system of knowing. The structure describes the system by identifying its constituting entities and elements drawing on the idea of distributed intelligence and cognition (Pea, 1998; Hutchins 1999) and identifying the relationships between those elements as processes that evolve with interactions within the system and with the external environment.

By epistemic, I mean, of, or related to, knowledge production. The term 'architecture' is much more problematic to define with precision. However, since this thesis is not situated in the field of architecture, a rudimentary general understanding of the term is employed. The Oxford English Dictionary defines architecture simply as "the complex or carefully designed structure of something[20]", and to add in computer terminology, "the conceptual structure and

[20] Source: http://www.oxforddictionaries.com/definition/english/architecture

logical organization of a computer or computer-based system". Perhaps, architecture implies an architect whose intentions are translated on to the structure that has a purpose. Therefore, in simple terms, epistemic architecture refers to the structures and organisation of energies that support the process of knowledge production in any defined system. It is proposed that the epistemic architecture can be examined through identifying and revealing the various planes of categorically similar features which exist and function in interaction with each other. Therefore, even if we place them in categories, it does not mean they could exist independently of an element in another layer, for example, Science Communication would be increasingly challenging for the GZ project if it weren't for the various social media platforms and the Internet. The categorical divisions serve an analytical function that in reality are more complex and it is crucial to acknowledge that dynamic.

5.3.1 Features of Galaxy Zoo's Epistemic Architecture

A) The Organisational Plane: This feature can be conceptualised at various levels of an organisation. From individual agents (such as the PI and TL) to groups (Science Team) and organisational entity (Universities, Funding bodies) and environment (Media, Public, Politics). The organisational entities involved in GZ at its inception were only at a group level, which increased in numbers as the project evolved to be citizen science initiative, including distinctive research teams, funding bodies and media agencies. The organisational layer of the epistemic architecture has dictated the freedom and the constraints under which GZ operates, and these shall be discussed in detail in chapters 7 and 8.

B) The Technical Plane: This feature refers to the selection, positioning and organisation of technologies (hardware and software) that the actors design, modify and interact through and with. In the case of GZ, the emphasis would be on the software architecture, as GZ is primarily an Internet-enabled entity. For instance, servers, programming languages, databases and protocols form the most critical elements of GZ's technical architecture (see Appendix E) and selecting the appropriate technologies and testing and refining them in the context of citizen science is a distinctive kind of domain expertise. These technologies allow for the primary tasks to be designed and realised on the website and to be subsequently carried out by the volunteers. All the data management, which drives the epistemic activities of the GZ team can be conceptualised to be carried out on this plane; e.g. the annotations (galaxy classifications) made by volunteers on the website, that need to be stored and accessed for data reduction by the Science Team. Even the algorithms for machine learning that they developed with the collaborators in the field of Artificial Intelligence, eventually have implications for the task refinement form an extended part of the technical architecture that characterises GZ. Communication between the scientists and the volunteers is also largely supported by web-enabled technologies through various media - from e-mails at the inception of GZ to most recently, Google Hangouts. It should be noted that the drive towards citizen science was propelled by data deluge in the field of Astronomy and operating in the age of Big Data, the technical architecture is not merely a plane that supports the epistemic activities, but is a vital feature that drives how science advances and evolves.

C) The Communications Plane: This feature refers to the interactions between the human agents in GZ activities; this could be within an organisational level or between different levels. The forms of communications examined in this study are primarily the non-exhaustive interactions between the GZ members, communications between the Core Team members and external organisations and individuals and finally, the publicly published GZ scientists' official communication with their volunteers. I am aware that there are many instances of communications and channels that were beyond the scope of this study, such as the interactions between the volunteers and the interactions between the team members at team meetings and conferences. The communications that were examined were largely facilitated by the various kinds of expertise involved, mostly relational expertise (Edwards A, 2009) introduced in chapter 4, as it is the expertise where communications between two parties is of primary concern since that is what leads to achieving joint action of shared purposes. The dynamics of the practice of communications and relational expertise is cyclical, in that relational expertise requires communication and that in turn reinforces the expertise or puts expertise into practice. The nature and purpose of communications between the GZ scientists is largely dictated by the domain expertise with a relational element, where it is assumed that the very membership is exclusive to those who exhibit domain expertise in Astronomy as a profession. The communication between the volunteers and the GZ scientists is largely supported by relational expertise, as the volunteers are assumed to have little or no domain expertise, but do have the cognitive capacities to carry out the primary tasks.

D) The Epistemological Plane: The epistemological layer of the architecture refers to the recognised latent and active (i.e. demonstrated) knowledge base in the form of human expertise and epistemic practices. However, the inclusion of the role of artificially intelligent agents in epistemic practices cannot be ruled out, since machine learning as a field is continually advancing machine capabilities that were once known to be unique to human cognition. In the context of GZ, it mainly refers to the levels and categories of knowledge operating in order to run GZ and achieve its primary goal of advancing domain knowledge in Astrophysics and in that process, developing other kinds of expertise that are critical in running a citizen science initiative operating at the intersections of science and the public. As posited before, communication can be dictated by the perceived expertise of the parties involved. Chapter 6 addresses the epistemological plane of this architecture by first examining the epistemic practice of translation, and chapters 7 and 8 reveal the emergent practices and expertise – other than domain expertise of Astronomy – that support the building of domain expertise of running citizen science initiatives. The epistemological layer reveals that the expertise exhibited across the system through practices are a result of pursuing the epistemic object, whether it is understanding an Astrophysical phenomenon such as a galaxy merger or organising the decision tree for the primary task. An analysis that follows the unfolding relationship between the epistemic subject and the knowledge object as put forward by Knorr-Cetina (1999) is proposed to lead towards an understanding of the epistemological plane of the epistemic architecture of a knowledge producing system.

E) Ethical Plane: There needs to be a distinction made in examining a design without ethical consideration even though the result of it might be perceived as being ethical or an ethical framework is imposed post hoc. Negotiating what constitutes ethical practice between the Science Team and the Core Team will be discussed in chapter 7, pertaining to decision to not ask the volunteers to perform tasks that may be considered as a waste of their time in the absence of a strong scientific justification. The emergence of the ethical concerns across the architecture was reinforced by the Core Team over time and is a reflection of the relational element of their collective expertise. Although the TL was the most vocal advocate of ethics of citizen science practice, the implications of that concern could be felt across the system as ethical principles allowed or prohibited practices to become a regular feature of the system. The examination of how ethical concerns emerged and were deliberated on will be explored in further detail in chapter 8.

The epistemic architecture described above continued to evolve over the period of data gathering. Recognising the dynamism of the system, characterised by complexity at different planes of the architecture, was possible due to the ongoing analyses that was undertaken in the present study. The initial architecture was minimal, meeting a simple purpose of receiving reliable galaxy classifications at a much faster rate. Although that remained the primary purpose of GZ, the complexity that came with unfolding of multiple areas of attention in experimenting with the ways in which citizen science works, was a deliberate decision. The GZ Core Team recognised the possibilities that might be realised as a result of pursuing that epistemic object. The epistemic architecture at GZ's conception had evolved significantly by the time GZ Hubble was launched and by then other projects had been launched under the Zooniverse brand based on the success of GZ.

5.3. 2 The Significance of Epistemic Architecture as a Conceptual Framework

It should be noted that I am not proposing a prescriptive blueprint to represent the epistemic architecture of a system. As a conceptual framework, the epistemic architecture is both an analytical tool and a knowledge object that represents the structure of a system which is relevant to the purpose of the study and the research question. It is a synthesised view of GZ that allows scrutiny of a system exhibiting the relationships within the system that can be easily ignored if not identified and referenced. Therefore, it serves as a reference to the system as a whole to aid the understanding of our epistemic object and becomes an artifact of the distributed cognition of the researcher and the reader.

The focus of analysis for this study is primarily on the practices of citizen science within GZ and the expertise of the GZ Core Team responsible for building the project and sustaining its success. Therefore, the awareness of the various planes within which they have to act is critical in understanding how the team worked and how they achieved scientific knowledge production with public involvement. Chapter 6 will focus on translation as an essential practice in citizen science while chapter 7 focuses on how the practices of the Science Team are shaped and developed in response to their interactions with other stakeholders to meet their goals. Lastly, chapter 8 will identify the expertise of the Core Team in building citizen science projects while identifying the design principles that shaped the development of their individual and joint expertise.

Chapter 6

Translation as an Epistemic Practice in Galaxy Zoo

This chapter aims to tease out the translation activities carried out by the Galaxy Zoo (GZ) scientists at different stages of knowledge production in GZ. Translation in this context signifies activities which require scientists to either a) change the way an object or an idea is represented and communicated to others and/or b) interpret the ideas and objects communicated to them by others. The participation of the public in scientific inquiry, as citizen science allows, calls for more translation activities to be carried out by the scientists to ensure that the goals of knowledge production, considered to be valid by the scientific community, are met. The translation activities carried out by GZ scientists may differ qualitatively and/or quantitatively with the public engaging in the scientific process of enquiry. Therefore, translation becomes the kind of practice that is critical in the production of knowledge through citizen science. Practice, in this context, is understood according to Scribner and Cole's definition as:

"a recurrent goal-directed sequence of activities using a particular technology and particular systems of knowledge" (1981: 235).

As one aim of this study is to examine how citizen science creates the need for a reanalysis of some aspects of scientific practice, this chapter will focus on the practice of translation.

In the following sections of this chapter, translation activities within GZ will be identified in the stages they occur as illustrated by the transformation that an object of enquiry goes through as a result of those activities. The object in this context refers to

"[t]he signifying force of partial objects (of epistemic objects in general) which resides in the pointers they provide to possible further explorations. In this sense, these objects are meaning- producing and practice-generating, they provide for the concatenation and constructive extension of practice" (Knorr-Cetina 2001:189).

In doing so, the role of the public volunteers is established as what I have termed as 'epistemic subjects', as entities who act upon the transitory objects of enquiry in the epistemic architecture of GZ. Evidence from the data collected will be presented to reveal how epistemic subjects play a role in knowledge production within the activity of translation. The function of volunteers acting as epistemic subjects supported by the by the epistemic architecture of GZ and the analysis presented in this chapter refers to the epistemological plane proposed in section 5.3.1 of chapter 5 and the translation practices are an integral part of the scientific process leading up to the production of new scientific knowledge.

6.1 Stages of Translation in Science

In order to explicate the ways in which translation practices differ in citizen science from those in more traditional modes of scientific enquiry, the stages of translation will be compared with a model put forward by Callon *et al.* (2009). They propose that the work of scientists necessitates three different stages as follows: translation 1: reducing the 'big world'/the macrocosm to the 'small world'/the microcosm of the laboratory; translation 2:

"[f]ormation and setting to work of a restricted research group that, relying on a strong concentration of instruments, abilities, devises and explores simplified objects." (Callon *et al*., 2009:48).

This stage is essentially, the analysis and interpretation of the macrocosm fabricated in the laboratory; and translation 3: returning the results from the laboratory to the big world. They suggest that Translation (with a capital T) is composed of the three stages of translation (t1, t2 and t3) which eventually result in some form of reconfiguration of the macrocosm from one state to another as illustrated in Figure 6 A (Callon *et al.,* 2009: 69). T2 takes place in laboratories, a territory marked as 'secluded research' as argued by the authors.

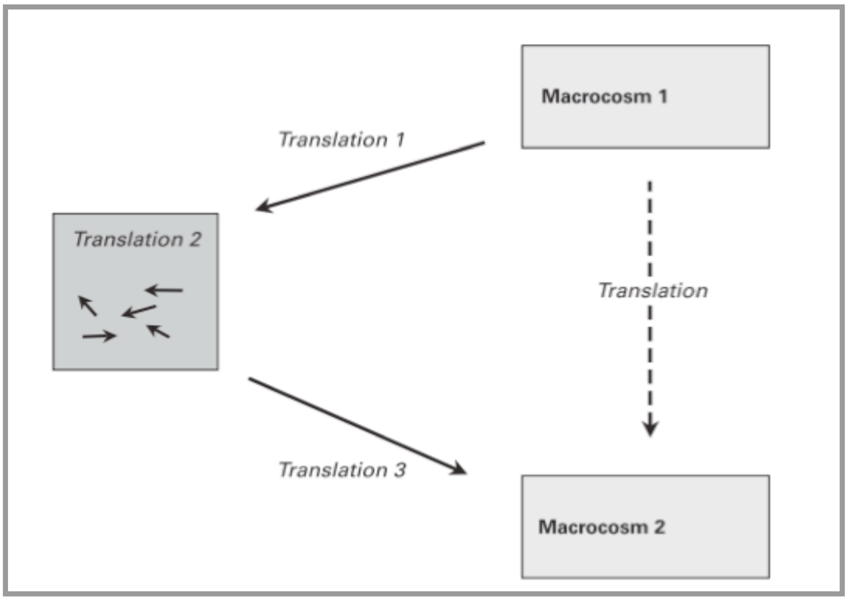


Figure 6 A: Stages of translation by Callon *et al*. (2009: 69)

Callon *et al.*'s (2009) stages of Translation in science provides a foundation for comparison with the stages that have emerged from this study. However, before we proceed, it must be clarified that the figure above presents a generic model of science presented by authors who view translation through a democratisation of science model, whereas this study investigates the practice from the scientists' point of view with a focus on what citizen science means for the practice of science. Therefore, there is a difference in perspective in terms of understanding how translation occurs in science which we must be aware of. In examining the different forms of translation activities in the practices in GZ, I do not propose a definitive number of stages of translation, but suggest that the participation of the public in the scientific process requires qualitatively and quantitatively additional stages of translation activities in the process of knowledge production than those previously conceptualised. A brief summary of the translation stages is presented in Table 6.1 below.

Translation Stage	**Activities undertaken by scientists (unless otherwise stated)**	**Callon *et al.* (2009)**
T1	Image capture of object	t1
T2	Image processed of task design	t2
T3	Annotating the object (performed by the public)	-
T4	Data reduction	t2
T5	Interpretation of data into results	t2
T6	Turning public curiosities into scientific inquiry	-
T7	Publication of peer-review academic paper	t3

Table 6.1: Stages of Translation in Galaxy Zoo

When these stages are compared to those offered by Callon *et al*. (2009), they appear to be expanding the processes involved in t2, the ‘secluded research’ phase. It is in these processes of translation that citizen science in GZ can be seen to have an impact. As Table 6.1 shows, the expanded t2 in this study are the stages that account for public engagement in citizen science. Involving the non-scientist public into the scientific practice of creating data sets required opening up the boundaries of science on the one hand and managing the involvement of the non-scientist public on the other. Such practices may be considered disruptive, in the sense that they do not conform to the established and accepted norms of such work and permeate through the boundaries of scientific practice. Through those practices, GZ challenged boundaries within and outside of its own professional/expert sphere and it is between and within those boundaries that the translation activities occur. The particular focus here will be the stages of translation that took place as data, as knowledge objects, move through the epistemic architecture of GZ.

6.2 Stages of Translation in Knowledge Production in GZ

In this section, I examine the translation processes outlined in Table 6.1. The Astronomical objects (Object1....n) have been numbered to indicate deliberate alterations made in any form as a part of the decision making involved in the scientific process, which then led to the object being transformed from one state to another, and does not bear or represent mathematical logic. The numbering is purposefully employed purely for representational clarity with an emphasis on elucidating the continuous transitory state of the object of scientific enquiry.

I) Object0 to Object1 - Image capture (T1)

Translation here refers to the capturing of digital images through the Hubble Telescope, which the GZ team did not have control over its production. However, the GZ Science Team members were able to request telescope time at various locations around the world, which could be allocated to them if they had specific requirements for observing a particular object or activity (e.g. galaxy mergers).

Most the images used in GZ were from publicly accessible catalogues of images such as the Sloan Digital Sky Survey (SDSS)[21] for the first version of GZ. This process, as an example, bears similarities with translation 1 proposed by Callon *et al.* (2009), whereby, a 'real world' object is transported to the lab as representation of the object:

"We will call this movement that starts out from the big world in order to arrive at the lab, and which replaces a complex and enigmatic reality with a simpler and more manipulable reality, which nevertheless remains representative, translation 1." (Callon *et al.,* 2009: 50)

This stage of translation was not frequently carried out by the GZ Science Team and they sourced their images from the Hubble database. At this stage, public involvement was not a frequent occurrence. However, evidence shows that when the GZ scientists became interested in the objects discovered by the volunteers (e.g. the Voorwepjes and Peas in Figures 6B and 6C), they requested observation time for those objects. Therefore, occasional traces of the volunteer-directed scientific activities were found. The public may, therefore, act as indicators for directions that scientists could decide to recognise and pursue, hence influencing the decision of which Object 0 could be translated in Object 1.[22]

> XMM-Newton time granted to observe the Voorwerpjes!
>
> Quick note to let you know that we've been granted time on XMM-Newton to observe three of the "top" Voorwerpjes. This follows the proposal we submitted earlier this year. The allocation is for priority "C" which means that they will take our observations if they fit into the schedule, but there is no guarantee.

Figure 6 B: GZ Blog post from 2 December, 2010

[21] Source: www.sdss.org

[22] For example, when a new Astronomical object, later named Voorwepjes was discovered by the volunteers, the Science Team decided to request a telescope to further investigate this object.

Radio Peas

Working with scientists in India, we have been awarded time on the Giant Metrewave Radio Telescope (GMRT) to study the radio properties of theGreen Pea galaxies discovered by Galaxy Zoo users. We hope to use this telescope to detect the first signs of radio emission from the Peas, establishing them as a new class of radio sources.

Why do we want to search for radio signals from the Peas? The radio emission comes from remnant supernovae which can accelerate relativistic electrons that emit synchrotron radiation. So when we are detecting star forming galaxies in radio emission, we are finding signatures from these supernovae, which tell us about the stars that live (or lived) in the galaxy. Therefore, using the radio emission we can trace recent star formation activity in the galaxy.

We are particularly interested in these Green Peas, because they are the closest analogues to a class of vigorously star forming galaxies found in the early universe (known as Lyman Break Galaxies). These galaxies behaved very differently from star forming galaxies in the present day universe, and can help us to understand how galaxies formed in the early universe. Because Lyman Break Galaxies are so far away, Astronomers have not yet been able to detect radio emission from any of these galaxies individually. In contrast, the Peas are much closer and we have a good chance of being able to directly detect them in radio emission. Detecting this radio emission, and determining whether r not the radio emission from the Peas is like that in nearby star forming galaxies will help us to understand the nature of star formation in the youngest galaxies.

Figure 6 C: GZ Blog post from 8th of September, 2010

II) Object1 to Object2 - Image Processing and Task Design (T2)

Once the GZ team was granted permission to access and use the images, they were required to make decisions regarding how the images were to be presented to the volunteers on the website, as the classifiers in this case were not professionally trained scientists. An example of an e-mail discussion thread on the internal team mailing list regarding this kind of translation is presented below. The excerpts from the team mailing list indicate the different ways in which the public enter the conversation between scientists regarding the translating of the images compared to conversations between the scientists themselves.

15 July 2010

“The first set are 'faded' to almost, but not quite, grayscale by reducing their saturation. My reason for not simply converting them to true grayscale is because then it is quite obvious what has been done to the image, and the classifier will potentially be alerted to the fact that they are being tested. By fading, and retaining a little colour information the image is much harder to distinguish from a 'normal' image, and so users’ behaviour should be less affected.” - Science Lead

“I'm not sure I understand how the single-filter COSMOS images get their colour. I am very unsure about the idea of enhancing the images this way and would prefer to simply show the monochrome images. We can always blog about WHY the COSMOS images aren't in colour and people will understand. Am happy to be argued out of this position though.” - GZ Scientist 1

“Although I can see both sides of the argument here, I'm in favour of using the colour COSMOS images. If our goal is to get classifications for the COSMOS images that are as close as possible to the results from the other HST surveys then it makes sense to make images that look as close to the others as possible, which means colour. The technique that S1’s using has been used by others, is at least consistent with what proper colour images would look like. If there's a particular concern, then I'm not against including a small number of monochrome images as well, so that we can measure any bias.” - PI

(Extracted from the internal GZ team e-mail. Subject: “Colour-bias test images in GZ: Hubble”)

It was the responsibility of the SI as well as GZ’s Lead Scientist to oversee whether the images uploaded on the GZ website were aligned with the task design that would enable the data reduction to be carried out without major problems. This was the objective of the translation carried out at this stage from object 1 to object 2. The decision followed a discussion on the merits and risks of employing techniques for altering images that would be presented to GZ volunteers. The Science Lead and PI, who were more familiar with user behaviour suggested a higher similarity to the rest of the images on the website in order to minimise the chances of undesirable user behaviour. The PI emphasised that the decision for such a translation had to be made while keeping the main goal of the task design in mind. It should be noted here that the translation activities of image processing and task design are done entirely with the public audience in consideration. What is also evident is the emphasis on the PI’s acknowledgement of both sides of the argument in this discussion, which he establishes prior to putting forward his suggestion and his rationale for it; this is an indication of PI’s emerging relational expertise in managing the science-related decisions within the team which will be explored in further detail in chapters 7 and 8.

T2 in this study is comparable to translation 2 in Callon *et al.*’s (2009) terms. They suggest that

“[s]ecluded research carries out a first translation that reduces the macrocosm and transports it into the microcosm of the lab. Instead of being at grips with abundant, complex, and heterogeneous phenomena, researchers work on simple objects that they can manipulate at leisure, surrounded by their instrument and their libraries. The translation leads to a change of scale that is also several of the relations of opposing forces. Through objections thus expressed, the entities end up acquiring an objective existence. We call this process of articulation and objectification, translation 2." (Callon *et al*., 2009: 59)

However, evidence from GZ suggests that the act of translation of Object 1 to Object 2 for the public audience as opposed to trained scientists, should be a distinctive category of translation; therefore, translation 2 (T2) in this study differs from Callon *et al*.’s (2009) translation 2 (t2).

III) Object 2 (Asset) - Object 3 (Annotated Asset) T3

Once the decision regarding task design and imaging was made, Object 2 was ready to be uploaded on the GZ website and was then referred to as an asset, which "represents the objects that is served up to the users in the classification interface[23]". These assets were then annotated by the users by performing the task designed by the Science Team. An annotation is the response volunteers give to a question or is marking a shape on an image. An annotated asset, therefore, could be represented as Object 3, a changed state from the previous form as a result of public interaction (e.g. classification) with the object at the third stage of translation 3 (T3). This was the only translation activity carried out entirely by the public through the primary task designed by the GZ Science Team. The discussions involving the design decisions are examined in greater detail in chapter 7.

IV) Data Reduction T4 and Interpretation T5

Object 3, the annotated asset, was directed to the GZ database for 'cleaning up' or data reduction. This is where the transformation of Object3 into Object4 occurred through translation 4 (T4). Only then was the corrected database ready to be interpreted by the members of the Science Team who had committed to writing a paper. Dissemination of Object 4 could occur in two ways, a) through a public data release, where scientists outside of GZ could use the data in their scientific inquiry and b) the interpretation of the data in papers written by the GZ

[23] Source: Zooniverse API Overview 1.0 document

Science Team. In the first scenario, Object 4 was released outside of the GZ boundaries without further translation work. This is what Callon et al. (2009: 65) argue to be t3: “the possibility of return to the world is decided in the course of translation 1, in the reduction and transportation. But it is in translation 3 that the alliances sealed by translation 1 are revealed, and their solidity and viability tested”. In the latter case, the state of object 4 was altered by the scientists’ interpretation of it (T5) and was subsequently fed back into the scientific community as presented in a peer-reviewed paper, which in this context would be through translation 7 (See Table 6.1) or t3.

The stages of translation just outlined fall within the intended structure of the original GZ where the completion of the primary task was most critical to the goals of the GZ team. The intended structure (GZ 1) included all stages of translations, but not T6. The interaction with the public demanded additional forms and stages of translation from the GZ team, beyond the intended structure that was in place for just the primary task to be carried out through annotations.

V) Translation 6 (T6): Transforming Public Curiosities into Scientific Inquiry

As the discussion forum was introduced in response to unexpected amounts of e-mails from volunteers of the first GZ project, an emergent structure, which also had to be technically supported, followed in the form of a dialogical space, which allowed for interactions among the volunteers and between the public and the GZ team members. Therefore, the GZ website, originally designed with the intended structure, later accommodated the emergent spaces of interaction, which manifested into the intended structure for the subsequent versions of GZ and other projects. This development recognised how volunteers might influence the epistemic architecture of GZ. In this instance, what was influenced was the technical architecture. This particular change was significant in expanding the GZ scientists' eventual understanding of what the GZ volunteers could potentially contribute and led to T6. These developments can be analysed on the technical and communications planes, which later influenced the epistemological planes in the epistemic architecture presented in section 5.3.1 of chapter 5.

The discussion forums built into the GZ website resulted not only in an increased interest in the knowledge objects presented to the volunteers during the primary task, but their persistent curiosity caught the attention of some GZ scientists. One example, already mentioned, was the discovery of a new Astronomical object, known as Hanny's Voorwerp, by a Dutch school teacher, which gained a lot of publicity in the media as an amateur citizen scientist contributing to scientific discovery. Another notable example was the discovery of the 'green peas', which started out as a small thread on the forum but later resulted in self-directed research by the volunteers. Within a few weeks, the volunteers (with varying degrees of domain-expertise in Astronomy) organised a search in the SDSS database for similar objects and the team ultimately submitted a scientific paper with some of the volunteers listed as co-authors.

However, both these examples occurred prior to the fieldwork done for this study. An example of how volunteers alerted the GZ team members towards new enquiries was found on the team mailing list in a thread titled "Fun Object" which started on 29th of January, 2010. The PI was alerted via email from one of the forum moderators to a rather unusual configuration of objects, which was then brought to the attention of the GZ Science Team. The thread not only had one of highest number of responses from the Science Team, but several members further followed the enquiry in searching for the original image from different catalogues of the Astronomical object.

Sustaining volunteer interest, however, required GZ Science Team time and effort in maintaining the relationship with the volunteers through a variety of channels of communication. The need to direct effort for the purpose of continuous volunteer involvement had been noted in an internal report prepared by the Education Team on the behaviour of GZ 2 users. The report examined the overall usage of the website and concluded that involvement in the project appears to diminish over time. The report suggested that occasional major events[24] and regular special events were required to maintain the volunteer population over an extended period of time and to incite interest for further engagement with GZ.

Some of the ways in which the GZ team retained the volunteers were through direct engagement in the discussion forums. One example was a discussion thread started by a GZ scientist asking volunteers to share the ways in which their involvement in GZ had changed their lives. There were also regular official GZ blog posts explaining the process involved in producing scientific knowledge. One example of this was a blog post: How to Handle Hubble images[25], which

[24] The use of the term events refers to occasions, announcements, contest through the GZ website or public media.
[25] GZ Blog port titled: How to handle Hubble images

describes how images of galaxies are processed to be ready for science. A total of 576 online blog posts were published between December 2007 and August 2011. The GZ scientists also regularly updated the GZ volunteers about papers that have been published using GZ data. An example of an update is presented below:

6 April 2010

"Hi Everyone! This is to let you all know that the Galaxy Zoo machine learning paper has now been accepted for publication in the Monthly Notices of the Royal Astronomical Society journal. The final version of the paper is at http://arxiv.org/abs/0908.2033. You can read all about the paper in my previous blog post at http://blogs.zooniverse.org/galaxyzoo/2009/08/05/latest-galaxy-zoo-paper- submitted/. The paper has already attracted a lot of interest from the computer science community demonstrating that your classifications are proving useful and interesting to non-astronomers as well!"[26]

The use of informal language in communicating with the volunteers was common across all GZ Science Team, indicating their perception of the volunteers as collaborators. Every GZ Science Team member who was interviewed made this clear. An explicitly public example of the efforts made by the PI to ensure that the volunteers perceive themselves as collaborators is a blog post[27] which he wrote in response within two days to a New York Times article[28] written by a Harvard University academic which suggested that the volunteers of citizen science "...are not doing the work of scientists...they are doing the work of scientific instruments." In his response in the GZ blog titled: 'Citizen' science and 'real' science, the PI argued that though that may be true for more traditional citizen

---

[26]GZ Blog post titled: Machine Learning Paper Accepted! http://blog.galaxyzoo.org/2010/04/06/machine-learning-paper-accepted/

[27] 'Citizen' science and 'real' science by Chris Lintott, 29 December 2010 http://blog.zooniverse.org/2010/12/

[28] Managing Scientific Inquiry in a Laboratory the Size of the Web by Alex Wright, 27 December 2010 http://www.nytimes.com/2010/12/28/science/28citizen.html?_r=0

science projects which require the volunteers to collect data, GZ operates in true collaboration between the volunteers and the scientists and he went on to argue:

29 December 2010

"If you need to run your own projects, or to acquire a publication record to be' citizen scientists', then consider it an aspirational label. The Zooniverse provides everything you need to do that, although for now, the barriers are still high. Otherwise, if you contribute to our understanding of the Universe is however minor a fashion, then I'll call you a scientist, and I look forward to being able to stop the distinction between professional and citizen." - PI

This blog post elicited a statement through a comment on the GZ Blog post[29] of clarification on the following day, from the academic who had initially written the article:

30 December 2010

"As the utterer of the quote about 'the work of scientific instruments' I can say it wasn't meant as a criticism of the contributors or of the overall effort. I personally am amazed and uplifted by the fact that those without scientific training- and all across the complex continuum- are able to collaborate on project that move science forward. I am a big fan of Zooniverse."

Besides defending how the citizen scientists are perceived by the community outside of GZ, the Science Team was involved in different ways in which outreach and education activities were carried out. An idea for such an effort started on the GZ internal mailing list on the 24th of February, 2011:

19 April 2011

"Hi! Here at Oxford we've had a brilliant idea for an outreach/education activity for our open evenings: Galaxy Top Trumps! :-) It's currently not available as an official Top Trumps pack: http://shop.winningmoves.co.uk/categorylist/top-trumps-every-pack/ But I reckon we and the Zooites are definitely the right people to make one up! We can show a mixture of nearby galaxies with names, and SDSS and HST galaxies as seen in the Zoo, and have characteristics like mass, distance, size, no. of stars, star formation rate, AGN activity, dustiness, rareness, spirality etc etc etc. We're going to make a home-made version of stellar top

[29] GZ Blog post comments: http://blog.zooniverse.org/2010/12/29/citizen-science-and-real-science/#comments

trumps here, but the galaxy pack should really be a Zoo project I think. What do you think? Anyone interested in helping me get started?"

After 27 e-mail responses, one of the highest number of responses to an idea in the team mailing list, an official GZ blog post with a discussion thread was published on the 19th of April, 2011. Such instances of engagement in outreach, science communication and education activities on the part of the GZ scientists illustrate how the public volunteers play a role in the production of knowledge by participating beyond the required primary task, granting them the status of epistemic subjects.

I would argue that the indicators provided by epistemic subjects require some form of translation on the part of the scientists interacting with the volunteers into a new direction of scientific enquiry as afforded by the emergent structure of the epistemic architecture, therefore, necessitating the role of translation 6 (T6). The interactions occurring as a result of translation and mediation in this phase in turn determine how the object is transformed. Although the role of the volunteers as epistemic subjects might be more visible in T6, as a result of Science Team being receptive to the indicating leads provided by the public, that role is not limited to T6 and the following section will demonstrate the ways in which the public can influence practices within GZ.

6.3 Volunteers as Epistemic Subjects

In Knorr-Cetina's (1999) study of high-energy physicists, the notion of epistemic subjects is employed to characterise scientists as the procurers of knowledge, the ones who act upon the knowledge object. In the previous section, it was established that through T6 (translation 6) scientists share their role as epistemic subjects with the volunteers, and this is achieved through sustained communication between the two parties supported by a dialogic space. This section will highlight the role of the volunteers as epistemic subjects in the other stages of translation found in GZ.

Translation Stage	Activities	Public as Epistemic Subjects
T1	Image capture of object	T6 led to request for telescope time application
T2	Image processing and task design	Carried out for a general audience; not for professional scientists
T3	Annotating the object (performed by the public)	Translation by numbers to match expert classification
T4	Data Reduction	Measuring for bias in classification behaviour
T5	Interpretation of data analysis into results	(Evidence not found in data)
T6	Turning public curiosities into scientific inquiry	New astronomical object discoveries, unplanned science questions
T7	Publication of scientific paper	Justification for using citizen science

Table 6.2: Public as Epistemic Subjects

In all stages of translation – except for T5, interpretation of results, as previously described – evidence suggests that there were recurrent instances which established volunteers as epistemic subjects. In T1, some of the GZ scientists decided to apply for observation time to further investigate the new objects discovered by the volunteers, which was facilitated by T6. Therefore, if the objects recognised by the volunteers were not brought to attention of the scientists, then T6 would not have occurred and therefore, T1 would not have been influenced by the volunteers. In T2, the image processing and primary task construction which are inherently linked together in how they are designed, the translation is entirely carried out for a general audience while considering only basic fundamental cognitive capacity (i.e. the natural ability of the human-eye to recognise patterns) needed to complete the tasks. However, when indicators (identification of unusual objects or patterns) arose, which made T6 possible, they had the potential of determining future primary tasks. T3 is entirely carried out by the public, when the characterisation of object was transformed by the epistemic subject, as the task requires annotation. However, a higher number of citizen scientist annotations were needed compared with an expert annotation before the object was considered suitable to undergo data reduction (T4). Once the objects were annotated by the epistemic subjects, any form of bias or error in the dataset had to be corrected by scientists, which allowed for validation of the contribution made by the volunteers. No primary data were found to illustrate how volunteers play a role in the interpretation of the data set. The absence of such data suggests, that this stage of translation is reserved for GZ Science Team members, as it requires expertise gained through professional training in the specific disciplinary field of research. To summarise, no evidence was found in T5 to suggest that public act as epistemic subjects.

Prior to the opening of the discussion forums, the traces of volunteers as epistemic subjects were perhaps most visible and significant in T3, as originally intended in its early design. However, with GZ team members being receptive to the pointers provided by the volunteers, they found themselves engaging in T6, which also influenced other stages of translation. This analysis suggests that a translation activity, which is influenced by the public, can potentially determine the direction of other translation activities.

The role of public epistemic subjects in Translation 7 (T7), which involves the dissemination of knowledge through a scientific paper of or releasing the data set, was a matter of contention that initially had to be defended within the scientific community. Some of the GZ scientists received feedback from reviewers of their research papers in which concerns were expressed about whether or not the public could have carried out the primary task and therefore whether the data were valid. Therefore, the decision to allow non-professionals to carry out T3 had to be explained and defended in the earlier phases of the initiative. The later decision to allow the public to participate in the scientific process by harnessing their cognitive capacities of pattern recognition to meet the pressing need for timely data reduction, revealed the capability for innovation in the practices of the GZ Core Team.

This creative spirit echoes Knorr-Cetina's (1999) assertion of how knowledge practitioners and specialists need to continually reinvent their practices of acquiring knowledge. In the case of GZ, extending the role of the public was conceived as such an endeavour, which gradually formed into a distinctive epistemic environment that was different from what had been originally operationalised as the first GZ project. The combination of the intended structure and emergent structure of the epistemic architecture led to interactions between the primary task (originally GZ's main goal) and the dialogic space. This shift led to an increased awareness among the GZ team of their responsibility towards the volunteers as epistemic subjects. This was later manifested into design principles for future projects, making relevant the notion of 'phronetic design' that will be discussed in chapters 7 and 8.

The role of the public as epistemic subjects becomes of critical importance in citizen science, as they are granted access to the objects of enquiry that previously only professional scientists had access to. Access to those objects allows for the role of the epistemic subjects to be shared between the scientists and the volunteers, the latter having more time to identify the pointers and communicate them to the GZ scientists who, through T6, turn those pointers into science questions. It is the epistemic architecture of GZ that supports the space for the epistemic subject to play a role in the process of scientific enquiry and which requires a distinctive kind of expertise to design such environments.

6.4 Translation 7: Publishing a paper and releasing a data set

The completion of this stage was of utmost importance to the GZ scientists and was recognised as contributing to the body of scientific knowledge. Translation 7 (T7) could occur either after T4 (data reduction) in the form of a data set release and/or after T5 (interpretation of the data set into results). The GZ project has published 24 scientific papers between 2008 and 2011, and two of those are papers on the datasets released by GZ, which has to be cited by any author who wishes to use these datasets. This stage bears similarity to Callon *et al.*'s (2009) translation 3 (t3) whereby the reconfigured macrocosm1 is fed back to macrocosm2. T7, through publication, results in the translation of Object 4 into an aggregated form of newly negotiated knowledge incorporated within a scientific paper, whereby the reduced data are now interpreted and marked with epistemic signatures[30] of the GZ Science Team on the object. The transformation of the object of enquiry, however, does not end at data release or publication. Carrying the interpretation as a contribution to knowledge, it can be used by other scientists outside GZ not only for a better understanding of phenomena but also for further transformation of the object through additional inquiries.

6.5 Concluding Points

The translation practices discussed above transform the knowledge object into a tool to be used in subsequent stages of scientific activity. As the analysis indicates, engagement in understanding and developing public interest and participation has influenced how scientists in the GZ team think about the knowledge they acquired during the process of becoming a scientist. The term epistemic subjects is intended to indicate that the public has a role in providing pointers for further unfolding of knowledge objects (in terms of tools, methods and disciplinary domain knowledge of Astrophysics). The notion of public in

[30] I use the term epistemic signatures to mean the recognisable and identifiable traces of annotations rather than interpretations of the Astronomical image made by the volunteers during the primary task. The use of the term signatures differs from Agamben's (2009:40) "*[m]ultimodal pan-semiotic assemblages, which enact relationships among systems of representation to reference a particular interpretation of an object*." The term 'epistemic signatures' was located in two papers on the most recent search (Last checked on September 8, 2014). While, Dhami and Wallsten (2005) do not provide a clear explanation for 'epistemic signatures' with only a single mention in their paper, Mazawi (2011) refers to epistemic signatures in the context of the political nature of school textbooks.

absentia appears to mediate the relationship between the scientists and the knowledge object pursued in GZ. According to Knorr-Cetina (1999), the subject-object relationship is one of mutuality as objects provide continual indicators and the subjects provide the possibility of the objects to continue and unfold. In GZ, the expertise in the translation practices of the Science Team allows for the public to be a mutual signifier in the subject-object relationship by weaving back those indicators through a dialogic exchange. Therefore, the public does play a role in inspiring new and renewed perspectives to be considered by scientists, the indication of which may be present in the knowledge object but which might have otherwise been overlooked by the scientists.

The role of the public as epistemic subjects lends itself to disruptive practices, which could manifest themselves as what Knorr-Cetina (2001: 83) argues to be "creative and constructive practice - the kind of practice that obtains when we confront non-routine problems - is internally more differentiated than current conception of practice as skill or habitual task performance suggest." Not only did the emergent dialogical support self-directed learning as originally intended, but the discussions on the GZ website led to new discoveries, new science questions and to new enquiries being translated into different kinds and levels of primary task as a result of the feedback from the volunteers. These are some of the ways in which public engagement led the GZ scientists towards pursuing scientific enquiry in an unexpected or previously overlooked direction that was afforded by virtue of public volunteers; and thereby, granting the public a shared role as epistemic subjects in the production of scientific knowledge through practices of translation within the epistemic architecture of GZ.

Science has often been criticised because it has dominated the discourse between science and society taking it towards its own agenda, particularly by ascribing to a public deficit model, which assumes a lack of scientific knowledge among the public that the scientists must counter in engaging with them. Bohm (1996), a prominent physicist and philosopher, argues it is through a dialogic interaction that true sharing of common content occurs, which then leads to collective meaning-making. Bohm (1996), however, contends that this idea of dialogue is therefore alien to the current structure of science, as science aims to seek the truth and not the meaning. GZ offers a challenge to this view. While this chapter was primarily concerned with presenting an analytical account of the translation practices that citizen science demands, the next chapter will examine the practices that emerged through the activities of the Science Team in pursuing the goals of both scientific inquiry and citizen science.

Chapter 7

Objectual Practices Steering Citizen Science in Galaxy Zoo

The aim of this chapter is to identify emerging core practices that steer citizen science, which in turn shape the expertise in managing citizen science initiatives, discussed in greater detail in the next chapter. This chapter provides an analytical account of objectual practices centered on science-related epistemic objects. The role of individual expertise of the Principal Investigator (PI) in moderating the Science Team, is examined in more depth in the following chapter.

The practices that the Science Team dedicated most of their discussions towards are identified and their interrelationships explored through studying the negotiations of their motives to achieve their ultimate goals. While the previous chapter examined translation as a critical scientific practice, this chapter focuses on the collective practices of the Science Team. In terms of the planes within Galaxy Zoo's (GZ) epistemic architecture presented in chapter 5, this chapter examines the practices, mainly focusing on the communication between the Science Team while intersecting the epistemological and ethical planes of analysis.

Citizen science, essentially, is the co-production of knowledge with direct involvement of the lay public in the scientific process. While there are core epistemic practices, such as translation, the inclusion of non-professionals in the knowledge production process requires practices for citizen science. Therefore, it is the concept of practices embedded in the context of co-production of knowledge that provides the framework for analysis presented in this chapter. Although, practice in general can be defined in accordance with Scribner and Cole's definition as - "a recurrent goal-directed sequence of activities using a particular technology and particular systems of knowledge" (1981: 235), as I explained in chapter 3, I draw more specifically on Knorr-Cetina's (2001) notion of 'objectual practice' prevalent in knowledge-based professional work and centered around epistemic objects. These objects are open-ended, unfolding, complex non-human entities through which practitioners engage in epistemic practices that form a specific knowledge culture. The relationship between the epistemic objects and subjects (practitioners) is what affords professional learning. It is those very objects that enable both stability and transformation of practice, as they are conceptualised as artefacts that serve both as a tool and the objects of inquiry. It is the simultaneous engagement with multiple elusive epistemic objects that accounts for the richness in knowledge-based professional practice.

In this chapter, the concept of practice is operationalised as purposeful action towards enabling the co-production of scientific knowledge. It is these practices that reveal the motives of the individual and the group, as they shape and define the epistemic object. For the purpose of analytical clarity, the epistemic objects for the GZ scientists are the emergent practices of co-production as they learn to recognise what is important to their object-driven or objectual practice (Knorr-Cetina, 2001) of citizen science. This is a form of practice that they are not familiar with as professional scientists. Decisions for co-production demand negotiation between purposes of science and citizen science, which leads to newly negotiated purposes and practices. With continuing discussions and negotiations of the simultaneously competing and complementing purposes of the project, new understandings of how to complete the task at hand are extended and expanded, providing evidence for the kind of professional learning that occurs within the GZ Science Team. The joint negotiation of purposes within the team supports building new common knowledge to create the culture forming the foundations for new emerging citizen science practices that characterise GZ. Therefore, the analytical lens employed in this chapter aims to focus on the negotiations between the purposes of science and citizen science that shape the objectual practices of the GZ team.

The unfolding of objectual practices of co-production that characterise GZ can be argued to be akin to what Fleck (1935) conceptualised as a ‘thought collective’. In what has been described as the pioneering work of Sociology of Science preceding Kuhn’s seminal Structures of Scientific Revolutions, Fleck (1935) offers an account of how scientists are socialised into what he calls a ‘thought collective’ through mutual exchange of ideas and intellectual interactions whereby certain ways of perceiving and thinking are not only adopted but go through continuous transformations. He argues that as a result of this interpersonal exchange, members develop ‘thought styles’ that are distinct from another thought collective, forming a particular ‘collective mood’ which persuades them to act and think in certain ways. This resonates significantly with Knorr-Cetina’s (2001) notion of how practitioners become engrossed with their epistemic objects in ways that characterise certain epistemic communities and their objectual practice. However, what makes Fleck’s notion more pertinent to citizen science practices is his recognition of the role of the exoteric or the lay public in scientific practice. He argues that when a thought collective develops a particular thought style, the specialists form a small esoteric circle demarcating the boundary between them and the wider exoteric circle. Although the esoteric circles are the source for knowledge for the exoteric public, they are not independent of them and rely on the exoteric for the ‘public opinion’ that science needs to survive. Within that framing, citizen science qualifies as an attempt to close the gap between the two circles. Therefore, the development of citizen science as a thought style and the GZ team as a collective analysed through Fleck’s (1935) framework contextualises the role of Knorr-Cetina’s (1999) epistemic objects in co-production of scientific knowledge.

7.1 Identifying Science Team Practices

As the beginning of the data collection phase of the research coincided with the preparations for the launch of GZ’s third iteration - GZ Hubble, it provided an excellent opportunity to observe the Science Team’s activities from a pre-launch stage, through its successful launch, to discussions about the succeeding version of the project. Discussions over this two-year period were observed and analysed, resulting in an understanding of the practices they were building on and their motives as they evolved over time, enabling a description of a GZ ‘s thought style in development.

Outlined below are the practices categorised in correspondence to stakeholders with whom the GZ scientists had to interact in order to complete those tasks. The purpose of presenting this table is to distinguish the tasks carried out by professionals that were specific to ‘citizen science’ from those of ‘normal science’ that were carried out by professionals[31]. Therefore, the most important practices of the GZ Science Team are listed and then classified as ‘science’, ‘citizen science’ and both ‘science and CS’ in instances where the task is critical regardless of the mode of scientific inquiry. It is by identifying the practices critical to citizen science that emerge and evolve as a result of GZ’s engagement with the public in the production of scientific knowledge, that the changes in practices within science are revealed.

The table below is organised in the following manner:

h) The stakeholders: Science Team, Funding Bodies, GZ volunteers, Media, General Public and the Scientific Community;

i) The tasks which need to be carried out by all GZ members including Science, Development, Education and Management (the Core Team);

j) The nature of the task, i.e. whether the task is specific to ‘normal’ science or specific to citizen science;

k) The GZ members responsible for the completion of the task; and

l) A summary of the purpose of that particular task highlighting its importance to the functioning of GZ. (P.T.O.)

[31] For the purpose of clarity in this research, by ‘normal science’, I simply mean science done without the direct involvement of non-professional public.

a) Stakeholder	b) Task	c) CS/Sci	d) Responsible unit	e) Purpose
Science Team	Negotiating Collaboration	Sci	CL/ ST	
	Primary Task + Decision Tree	CS	ST	To get annotations from volunteers
	Publish Papers	Sci/CS	ST/CL	Acceptance by scientific community
	Science Communication	CS	CL/ST	To maintain volunteer interest
	Data Release	CS/Sci	ST/CL	Promote open data, Ethical
	Conferences	Sci/CS	ST	Validity by scientific community/ Recruit peers
	Comparing against expert annotations	CS	CL	Proving validity of volunteers classifications
Development	Assess feasibility	CS	TL	
Funding Bodies	Submit grant applications	CS/Sci	CSA	Funds for Development Team; project management
	Prepare reports and advisory meetings	CS/Sci	CL/CSA	Justify expenses to funding bodies;performance review
GZ Volunteers	Attract traffic	CS	CL/ST	To get annotations from volunteers
	Maintain traffic - sci comm/outreach	CS	CL/ST	
	Motivation and Behavioural Research	ET	Outsourced	To allocate expertise-specific tasks
Media	Promote project launch and results	CS/Sci	CL	Maintain traffic and attract new projects
General Public	Attract and retain attention	CS	CL/ST	Maintain brand imag; attract projects and volunteers

Collaborators	Machine Learning		Outsourced	
	Motivation and Behavioural Research	CS		
Scientific Community	Open Data Policy, Open Science, Astroinformatics, Big Data			

Table 7.1: Role and Responsibilities of the GZ Teams

Key: CS: Citizen Science; Sci: Science; ET: Education Team; PI: Principal Investigator; TL: Technical Lead; ST: Science Team; and CSA: Citizen Science Alliance

These tasks have been compiled using the data generated from the conceptual framework co-created with the PI to visualise the structure and practices of GZ (refer to Figure 4B). These key practices have been recognised, refined and established over time and through experience as the GZ team faced new and different challenges and negotiations. In the process of being responsive to novel challenges, their practices have been subjected to reflection by team members at the individual and collective levels, as a result of which their responses to challenges have varied across different instances. In completing the tasks under circumstances with varied constraints, learning in the GZ Science Team has continually evolved. The following sections will present the events and instances that the GZ Science Team faced in their decision making to meet the objective of the task at hand, which is how the expertise in running citizen science projects developed and was learnt over a process of response, reflection and recognition within the environment in which they operated.

7.2 Objectual Practices for Citizen Science

GZ is first and foremost a scientific endeavour. Therefore, the most important outcome for the GZ Science Team was to publish research papers in scientific journals. The first and one of the most critical activity aimed at achieving that goal, is to design a primary task for the GZ volunteers. The process of designing the primary task was continually negotiated with each iteration of the project and was not defined or standardized, as it involved negotiating the possibilities of what the volunteers could be asked to do in relation to the potential tradeoffs that might result from that design decision, which is not only driven by science, but also by the anticipated but as yet unknown behaviour of the Zooites. These are the kind of negotiations that will be illustrated through excerpts from discussions between the Science Team members on the team mailing list. Although the primary task design was understood to be fundamental to GZ from the project's initial stages, the need for communicating with the volunteers arose immediately after the launch. Communication began with e-mail queries from the volunteers and, as I have already indicated, later became an integral element of GZ's model of citizen science. After successfully publishing papers based on GZ data and proving its reliability, decisions pertinent to collaborations and sharing of data also became a recurring focus of the discussion, as non-GZ scientists expressed their interest in using GZ data. The practices that characterised GZ's citizen science were therefore not predefined but emerged as responses to demands from, for example, volunteers and external scientists, and unfolded as the team recognised the demands and created new objectual practices while fulfilling their scientific goal.

In the following sections, excerpts from the Science Team discussions will be presented in order to examine the complexity of the issues and negotiations of purposes in pursuit of the epistemic objects they recognisedand the extracts will allow an examination of the practices they engaged in while aiming at good science.

7.2.1 Setting Up Collaborations and Primary Task Design

As it was established earlier, the GZ scientists required images to be classified by the volunteers and obtaining those images from various possible sources became one of the first collaboration-related negotiations for the Science Team. For GZ1 and GZ 2, images from a publicly open database (SDSS) were used and therefore moving on to the images from the Hubble telescope was an entirely different undertaking that required discussions between the GZ Science Team and the potential collaborators and called for a re-negotiation of purposes within the team. The discussion on moving to GZ Hubble on the mailing list was initiated by the PI. He discussed the change with the Science Team and identified potential hurdles and the decisions to be made prior to the launch of the new site. He reassured the team that they had improved on the technical aspect since their previous version of GZ, and focused on the challenges related to scientific outcome in obtaining images and the questions they might want to ask the volunteers. He therefore, defined the focus of the discussion around the tasks that need to be recognised in negotiating a collaborative relationship with non-GZ partners. All extracts presented are from exchanges on the e-mail list unless stated otherwise.

February 2, 2010

“I wanted to bring you up to date with where I think we are in progress towards 'Galaxy Zoo 3' aka Hubble Zoo […]. However, to avoid the long wait between Zoo 1 and Zoo 2 I think we should get a plan in place as soon as possible.”- PI

The PI indicated the need for a plan for the third iteration of GZ, GZ Hubble, based on their experience with the previous launch, suggesting that he wanted the team to start thinking collectively about the tasks that needed to be completed for a successful launch. As there were no established protocols for how the tasks were to be carried out, the practice of negotiating collaboration was open for joint discussion, beginning from identifying the potential problems that required the team's attention and decision, to determining a new course of action. Since the nature of the problem was not predictable, each problem called for a new set of conceptual resources and new ways of thinking about the problem. This practice of constantly providing pointers towards the epistemic object (understanding the practice of negotiation collaborations) opens a path for the object to evolve resulting in the activity of creating an objectual practice that the Science Team collectively embarked upon while, negotiating and reconfiguring their own purposes and understanding.

25 February 2010

"In my own opinion, we might end up needing to change the list of questions rather radically, but in the absence of any consensus on what to change them to [it] seems sensible to start with what we have got."

The major technical hurdle is in obtaining the images. However - and correct me if I'm wrong - it seems that neither set is quite right for our needs." - PI

As the PI began by outlining the possible science-related difficulties they might face in obtaining the images of galaxies and the questions those images might allow, he went on to lead a discussion on citizen science-related concerns regarding the positive publicity they might receive as a result of this collaboration. He also brought the Science Team's attention to their favourable bargaining position in this collaboration, encouraging the team to engage positively with the negotiations.

25 February 2010

“While musing about all of this, I was contacted by [name] at Space Telescope Science Institute[32]. She reported that a cluster of people there had been discussing Citizen Science and - inspired by us - they would like to get involved. Luckily, we inspired them so much that they want to help us rather than run their own project.” - PI

“I didn't commit Galaxy Zoo to anything without talking to the team, and there are some long term ideas which we didn't get into (ideas for other Zoos, or for extra features and tools) but in the short term the good news is that they would like to make the launch of HST Zoo or something like it the centrepiece of their online celebration of Hubble's 20th anniversary, on April 24th. I think that if we can meet that timescale this would be an excellent way to ensure that the new version of the site gets started with the requisite publicity, and would like to see if we can make it happen. I've sent a list of requirements for 'good' images, which they're going to see if they can produce.

Obviously any changes will have to be discussed amongst this group, but we should see whether they can produce acceptable images rapidly first.

P.P.S. In an ideal world, I think we could run a limited beta and make decisions about questions before the big launch, but again as we haven't seen images timescales are very unclear.

P.P.P.S. We could of course use the other image sets above and thus hit the deadline, but I am still very worried about the disappointing experience of viewing them.” - PI

The PI assumed the responsibility for laying out all the potential issues, steering the conversation around those that he perceived to be of importance in negotiating the collaboration, thereby shaping the space for objectual practices to unfold. He emphasised the advantage of a significant publicity event offered by the Space Telescope Science Institute (STSI), critical for the purpose of attracting traffic and encouraging volunteers to engage in the primary task. In

[32] Referring to the chair of the advisory panel on one of the Zooniverse National Science Foundation grants

addition, although the decisions around image quality and selection was primarily based on scientific standards, he positioned the problem of image quality in relation to acceptability by the volunteers. Meeting the expectations of the volunteers in how they experience the objects they classify becomes a concern, highlighting some of the tensions within the science and citizen science goals of the project. This excerpt is also an illustration of the PI's expertise in communicating the inclusive nature of decision making within the Science Team and the ability to perceive tensions in relation to both science and citizen science. Consequently, he displayed his 'relational expertise' (Edwards A, 2010), a concept further explored in chapter 8. Throughout the chapter, more instances of relational expertise in practice will be highlighted in how the Science Team functions, building towards evidence to support the criticality of such kind of expertise in the working of GZ.

Responses to the issues brought forth by the PI included some tradeoffs from a project management perspective, as a senior scientist, one of the very few in the GZ Science Team, brought the purposes of science more prominently into the discussion of the criteria for collaboration. He directed the team's attention to the potential advantage of the external expertise brought by the potential collaborators; but strongly expressed concern regarding control over the science output and warned of the tradeoff between scientific rigour and publicity benefits when designing the next project.

February 2, 2012

"I think having the expertise, know-how and personpower from STScI will help a lot. This is clearly a big task and to get it right will take many people working hard together. So I'm all in favor of joining forces.

I have two cautionary comments. First, we (GZTEAM) should remain involved in the science and be leading that with our STScI colleagues. So I would not want to lose that from your negotiations. Second, I think we need to be very cautious about rushing this just to satisfy their PR event. Doing this wrong would be worse than missing April 24th in my opinion. As you have demonstrated numerous times, the users keep coming back when we ask them, so I don't think we (GZ) need the STScI event to raise the GZ profile and get clicks. We will get the clicks regardless of them (although it may take a little longer). We also know that doing excellent science with GZ has gained us both professional and public kudos - we should not sacrifice that principle just to satisfy their deadlines." - GZ Science Team Member 2

February 2, 2010

"Just to reassure, there's no sense in any of my conversation with STScI that they want any of the scientific return - they're happy to leave that to us.

I think that we do need a big press event. Lots of the Galaxy Zoo 2 classifications came during our initial big peak, and so the progress will be sped up if we can get a similar boost here. I agree that we need to be careful to have everything in place before we commit to any date, though. The good news is that from a technical point of view we are already where we need to be, with the exception of the images. This is a nice contrast with Zoo 2 where we reached this state a few days before launch!" - PI

After a few other team members agreed with the PI on going ahead with the launch date coinciding with Hubble's 20^{th} anniversary to take advantage of the press event, the discussion was then directed on working with images to make them suitable for GZ's use. In particular the team considered the science questions involved in designing the path of the primary task by altering the existing decision tree as the excerpts below illustrate:

February 9, 2010

"The next stage will be to decide which data and when, on which more shortly when I've worked out what our options are. I promise not to make any decisions without consulting. One thing to think about in the meantime is whether we want to make changes to the GZ questions. I'd say we want - as I think S2 said in the last thread - to keep them the same, but I wonder if we should drop some of them. Thoughts?." - PI

Determining the design of the decision tree, another key objectual practice, for GZ Hubble brought forth questions not only relating to the quality of classifications, but also volunteer retention and comparability with the dataset of GZ2. As some of the members began discussing the details of possible changes in the decision tree, one of the GZ Science Team members specialising in Astronomy education research highlighted his concern over how certain changes would influence the behaviour of the volunteers, and advocated keeping the decision tree closer to the original version. This negotiation illustrated another tension that was perceived between detailed primary tasks with a science-driven purpose and consequences of reliability and accuracy of volunteer classifications as one of the multiple epistemic objects in co-production practices, and for collective emerging expertise in citizen science design.

February 11, 2010

"The closer we keep the classification tree to the original GZ tree, the cleaner our experiment is from a social science perspective; we should put some thought into this given how we saw Zooite behaviour changing after the bias images were added."- GZ Scientist 3 (Also member of the Education Team)

More details of the discussion on designing the decision tree for GZ Hubble are not presented here due to constraints in word limit and technical details that require advanced domain knowledge in Astronomy. However, the extracts selected for this chapter illustrate the complexity of simultaneous engagement with multiple epistemic objects by the GZ team and how resulting decisions shaped the objectual practices that characterised how the 'thought collective' functioned. These were professional scientists collectively attempting to speculate and design for non-professional engagement with their knowledge objects (images of galaxies) and recognise what was important for their science in this new practice of citizen science. The excerpt from the Science Lead's (SL) email below shows one of the many attempts at making design changes to the new decision tree for GZ Hubble, while attending to the agreed aim of maintaining consistency with the previous decision tree, by sustaining complementary purposes for both science and citizen science. That consistency would facilitate comparative analysis of classification behaviour by the scientists and there was an assumption that familiarity in primary task path would not be discouraging for the existing volunteers with experience of GZ2.

11 March 2010

"I agree we want to maintain the decision tree as close to GZ2 as possible. However, I think the way to do this is not to create a separate tree, but to add the extra branch as low as possible, so the questions above are less affected by the change. I'd actually consider pushing the "new Q2" down further."
- Science Lead

26 February 2010

"The aim is to move focus completely over to GZ:HST. If GZ2 were to keep running the rate at which it collects classifications would be very low. As far as I am aware, the only question added since launch was dust lanes, and that has now been running for some time. What are the statistics for numbers of objects classified with the dust lane button active (I think there should already be a reasonable number of classifications)? If you can argue quantitatively that it is very important to the science to continue GZ2, then we may need to rethink things, but I am reticent to continue collecting classifications for the sake of that one button." - Science Lead

The suggestions that were brought into the discussion also presented the motives that differed across the team members. The debate on the transition from GZ 2 to GZ Hubble also instigated a discussion of an emergent principle for the scientific justification for project transition. The SL was heavily involved in data reduction for GZ and demanded a strong scientific justification for holding back the transition onto a new database. His concern was that the delay was to serve a small part of a single paper being written by one of the team members who had earlier argued in favour of a delay to the proposed transition. This is one of the examples where the principles on which GZ operated was made explicit to the Science Team. As already noted, GZ began with little organisational structure and no explicit formal rules regarding GZ activities. This kind of openness was deliberately sustained to maintain flexibility in team practices. However, the need for some rules and principles was eventually recognised by Core Team members when they perceived a need for them to be made explicit to the rest of the Science Team. The rules and principles then took the form of epistemic objects as they were collectively explored and negotiated, lending themselves to strengthening the elements that constituted their own 'thought style', in this case characterising the aspect of their objectual practice concerned with ethics of citizen science practice.

As Hubble's 20th anniversary approached, the PI summarised the discussions about the important issues and emphasised the publicity offer made by their collaborators. However, as the discussion on expanding the decision tree seemed to extend into the launch month of GZ Hubble, he questioned the need for such design changes and proposed a solution based on GZ's previous experience. In doing so he reminded the team that they would be operating on a more technically powerful platform, putting them in a better position to launch. Apart from the need to look at the problem from a perspective of their increased capability to derive solutions to new problems, he focused on the need for consistency between GZ 2 and Hubble samples and anticipated which design decisions could potentially have a negative impact on volunteer motivation. What stands out in this instance is the PI's ability to understand the bigger picture and how decisions that may seem insignificant can affect the sustainability and success of the entire project. His constant emphasis on understanding the changes from the volunteers' perspective indicated development in his relational expertise in the process of understanding what objectual practices were significant for GZ's success. In establishing the practices that mattered to GZ, he drew on his understanding of success factors from previous experiences in the practice of citizen science and established knowledge in the current collective understanding and pursuit of their objectual practice.

25 March 2010

"Given that we're nearly done with Zoo 2 (see previous email) my strong recommendation is that we take advantage of the added publicity of April 24th's anniversary. STScI are offering to feature us prominently both in the press release and via the front page of hubblesite.org. The latter is important as we will thus benefit from anyone who decides to write about the Hubble anniversary, which is bound to be a huge media splash. We can also do our own press release as normal.

We have strongly benefited from press attention for the launches of Zoo 1 and Zoo 2, but the trick won't repeat for ever without a strong extra hook. The Hubble anniversary is our best opportunity (barring another voorwerp sized story) to do that for ages, and as we can continue to collect SDSS queries at the same time I say we should take advantage of it." - PI

2 April 2010
"I don't mean to be awkward (and I agree that we want to gather this information one way or another) but I'm not convinced that we should be expanding the tree like this. Our experience from Zoo 1 is that - despite the admittedly slightly misleading button design - it was easy to distinguish true mergers, with SD finding that anything with more than 40% of the vote in this category being a true merger and most of the things with more than 20% also fitting. This creates an easy way to find a subsample which could then be interrogated for these more detailed questions. For Zoo 1, SD had to do this himself but with the now much more powerful Zooniverse platform (and more programmer time) it would be easy to create a standalone merger Zoo.

This would be preferable to making the decision tree (and accompanying tutorial) longer. Avoiding making this change will also make it easier for us to compare completeness between Zoo 2 and HST samples (I'm worried that by essentially committing users to answering more questions if they click merger we create a disincentive to do so. I know that the same is true of elliptical or not, but the effect will be larger at the end of a normal tree).

Finally, I'm worried about asking questions about a companion when the users can't tell us which galaxy they are considering to be the companion [….] I suggest that the 'Merger Zoo' interface allows people to select an object before answering its morphology.

I'm happy to make constructing such an interface a priority and even to have some threshold at which galaxies are automatically passed across to it from the main Zoo. But I think it's an error to tack it on to the end of the main merger tree, and would be strongly against making any such move in switching to HST Zoo."
- PI

Apart from carefully designing the primary task and running tests among the Science Team, the Development Team led by the Technical Lead (TL), implemented the changes and pointed out any corrections and technical bugs on the GZ site before it was to be launched. To help the volunteers classify objects as accurately as possible, the tutorials that had been prepared had to be checked thoroughly by the Science Team. That check required anticipating the most common questions that needed answering, to provide adequate guidance to the volunteers when making the transition between GZ 2 and GZ Hubble. Therefore, constructing the FAQs became an objectual practice of speculating questions that the Science Team expected the volunteers to raise and providing answers. The process indicated a common understanding that had been established between the Science Team members that the volunteers were capable and curious enough to ask questions about the source of the images based on the change in quality of the 'graininess' of the images. Creating the FAQ was an exercise in enabling the engagement of Fleck's (1935) exoteric circle with the knowledge objects that were determined within the esoteric circle of scientific practice. Therefore, jointly configuring the language and level of domain-knowledge in Astronomy to communicate with the Zooites became a key practice which could be categorised as 'science communication', and which is discussed later in this chapter.

17 April 2010

"Maybe the graininess should be an FAQ entry. There is also the issue that the HST images are only 2 colour, rather than 3, which may also be worth an FAQ. Shall I add them?" - Science Lead

19 April 2010

"In the FAQs page, we need something like:
Some of the images look like they are from SDSS, not Hubble A. We have left in some SDSS images to see how your classifications compare between the two telescopes." - GZ Scientist

"We currently have Q. Where did these images come from? The galaxies in Galaxy Zoo: Hubble are drawn from several large surveys completed with the Advanced Camera for Surveys. So that we can compare the results from Galaxy Zoo 2 and Galaxy Zoo: Hubble we've included a selection of images from Sloan. You can find out more about these surveys on their individual webpages." - PI

The launch of GZ Hubble was successful on the 20th anniversary of the Hubble Telescope with a press release from NASA encouraging the public to join GZ as volunteers, as promised by the collaborators, establishing the importance of the role of publicity in launching GZ projects among the GZ Science Team. In summary, the practices that the GZ Science Team engaged in setting up with their collaborators, involved science-centered purposes that manifested themselves in issues of image quality, control over science output, design of primary tasks and altering the decision trees. All of these were entangled within the joint purposes of citizen science, such as the appearance of the image, consistency between the old and new decision trees' details, and anticipating essentials FAQs to guide volunteers into the new system. Therefore, within the practice of setting up a collaboration before a launch of a different version of the project, the details of the tasks involved were jointly negotiated and discussed, thus framing the nature of learning that occurred within the unfamiliar practices of running a citizen science initiative. The advantage of media coverage played a critical role in encouraging the Science Team towards this particular collaboration, and establishing the criticality of citizen science-driven purposes within the GZ's thought collective. The examination of the team discussions demonstrates how epistemic objects are identified in the process of negotiating the objectual practice of making decisions regarding collaborations.

Therefore, in setting up a new collaboration, the epistemic object in focus in this section, in particular the terms of negotiation of purposes among the Science Team results in externalisation of new understanding, which re-shaped the practices and tasks. The understanding of such externalisations, taking in this instance the form of written text through the team mailing list, led to building common knowledge through a joint negotiation of the purposes of science and citizen science; the common knowledge that characterises the thought style of GZ.

7.2.2 Data Reduction for Citizen Science and Data Release: Towards Open Science

The most significant advantage of having nearly half a million volunteers carry out the tasks that previously had to be completed by scientists was the high number of classifications that were given within a significantly shorter period of time. However, as we have seen, the processes also required stringent measures to identify and rectify errors in the classifications obtained (i.e. annotations) compared to that of the professional Astronomers. This stage of translation was identified as T4 in the previous chapter. Therefore, the annotations made by volunteers were required to undergo various checks and verifications by the Science Team before the data set was considered to be reliable for analyses that could lead to scientific papers. As the PI pointed out in an interview, GZ was designed as a scientific experiment from the very beginning:

13th February 2010

"We said that it was designed as a scientific experiment. [...] So we did proper data reduction, the statistics were impeccable. I shouldn't say that. We put a lot of effort into the statistics. So we were actually, we weren't saying this is an elliptical, we were saying this is an elliptical with, this appears to be an elliptical because 8 out of 9 people have said this." - PI

After completing the original version of GZ- GZ1, the Science Team had realised that the annotations obtained from the GZ website needed to be processed to eliminate 'bad' classifications, including intentional and technical errors and biases. They then needed to be ordered into a data catalogue that could be shared within the GZ Science Team and eventually publicly released for other scientists. Releasing the GZ data served an important purpose for the GZ team as it confirmed and reinforced the validity and usefulness of their efforts and normalised citizen science as a legitimate way of doing science within their research community. Therefore, the data release was aimed primarily at two categories of audiences: a) The volunteers/ public and b) the scientific community in the field of Astronomy.

Soon after the data were made public, a post was published on the GZ blog informing the volunteers that their work would be beneficial for the professional

Astronomy research community and encouraged them to ‘play’ with the data. This practice to make data ‘open’ was critical to the values of GZ’s citizen science, and was a result of being inspired not only by open data policies of others, such as the SDSS database they used for GZ1, but also resulted from the demands of the volunteers. Making the data available exposed them to scrutiny by a wider audience, both professional and amateur, enforcing the need for reliability that met not only professional standards, but encouraged legitimacy through reuse.

Therefore, the reliability of the dataset became the primary epistemic object that drove the objectual practice of data reduction to render the GZ data suitable for public release. In their own words, the purpose of the data reduction process was to “turn clicks into science” which sums up GZ’s primary goal. The following sections lay out the practices that are central the process of data release.

The motives for writing a data release paper for the PI were more than just one. In an early interview, he explained how working on such a paper was necessary for a) proving that the project would contribute to science and b) allowing other non-GZ scientists to refer to the work.

13 February 2010

“[t]he first thing I wanted to do was a data release paper. A paper describing the project and data so that other people could cite it. That is useful because you would not have all that information at one place. [...] So we did that. That needed to have some proof that we were doing something sensible. So I did a lot of work comparing [...] to existing professional classifications.” - PI

This data release paper[33]outlined the steps taken to ensure reliability of the classifications made by volunteers. These included ensuring strength in numbers of classifications for each galaxy observed, employing data reduction methods to account for all possible errors and biases, and comparing the results to expert classifications for accuracy measures. The following are extracts from the journal article:

"Each galaxy in our sample was thus viewed and classified multiple times, with a mean of ~38 classifications per galaxy. A variety of strategies are available to convert from these raw classifications to a final catalogue.

The first step in data reduction involves removing obviously bogus classifications. A small number of users seem to have recorded a number of these classifications, either using some sort of automated mechanism or due to some unknown problem with their browser. [Text omitted] There are 36 such potentially malicious users, amounting to less than 0.05 per cent of the total number of participants. Furthermore, in order to account for accidental double clicks, if a user has classified the same galaxy more than once, we take into account only the first classification from each user.

By combining the classifications of more than 100,000 participants in the largest astronomical collaboration in history, we are able to produce catalogues of simple galaxy morphology which agree with those compiled by professional astronomers to an accuracy of better than 10 per cent. Our results thus suggest that the general public can reliably classify large sets of galaxies with a similar accuracy to professional astronomers." (Lintott *et al*. 2008:1182)

Almost 14 papers and two years since the first GZ paper, the data set from GZ 1 was made available for members of the public and non-GZ scientists alike. The data were uploaded to the website and came with a clause (stated below) for its open data policy. The clause included measures taken to a) ensure that the project was understood along with its limitations and strengths before it was

[33] Paper titled Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey [Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R. C., Szalay, A., Andreescu, D., Murray, P. and Vandenberg, J. (2008)]. As of April 2013, it had 166 citations on ADSABS Harvard Service

interpreted by a non-GZ member; b) encourage the use of the data; and c) allow for citation tracking of others who might use the data.

Galaxy Zoo 1 data release

The original Galaxy Zoo project ran from July 2007 until February 2009. It was replaced by Galaxy Zoo 2,Galaxy Zoo: Hubble, and Galaxy Zoo: CANDELS. In the original Galaxy Zoo project, volunteers classified images of Sloan Digital Sky Survey galaxies as belonging to one of six categories - elliptical, clockwise spiral, anticlockwise spiral, edge-on , star/don't know, or merger.

Full catalog

This webpage allows anyone to download the resulting GZ classifications of nearly 900,000 galaxies in the project.

Galaxy Zoo is described in Lintott *et al.* 2008, MNRAS, 389, 1179 and the data release is described in Lintott *et al.* 2011, 410, 166. Anyone making use of the data should cite at least one of these papers in any resulting publications.

Figure 7 A: Data Release Post on Galaxy Zoo Website

This formal data release was accompanied by another paper titled: Galaxy Zoo 1:data release of morphological classifications for nearly 900 000 galaxies by Lintott *et al*. (2010).

The abstract from the data release article provided a brief summary of the process of data reduction for that particular data set for professional scientists. The purpose of writing a paper on GZ 1 data release was accomplished, as the article was highly cited. The recognition of their efforts were reflected not only in the number of citations in the journal they published in, but the article was also selected as one of the most popular papers in the NASA

Harvard Abstract service. Therefore, with positive reinforcement measured by citations and popularity indexes, publishing a paper along with the data release became an established objectual practice in GZ.

The decision to release GZ 1 data turned out to be a favourable one for the GZ team as one of the reputable catalogues (SDSS) decided to make GZ dataset available through their platform. This indicates the acceptance and appreciation of GZ's efforts to do scientific research with the help of public participation. The team worked towards the goal they had of legitimising citizen science as a way of doing sensible science, and an acknowledgement from their professional community was an encouragement for the GZ team, as pointed out by a senior scientist, who congratulated their efforts.

21 December 2010

"Dear GZ team, just in case no-one knows, the SDSS-III team have included galaxyzoo DR1 [GZ1 Data Release] classifications in their forthcoming Data Release 8, including integrating the data into the SDSS CAS. This will allow everyone to access the GZ1 data via SQL queries alongside all the other excellent SDSS parameters etc. This speaks volumes to me as it demonstrates the importance of GZ1 and shows our colleagues respect the job you guys have done. Bravo." - GZ Science Team Member 2 (Senior Scientist)

It was perhaps no surprise that the data release was not received with the level of apprehension from the scientific community that the PI anticipated. The interview excerpt below gives some indication of his previous concerns. From the response that the Science Team received and the absence of significant objections or doubts, the GZ team successfully established credibility of their 'sensible science' done with the help of GZ volunteers. As the PI explains in an early interview, what was once identified as a critical objectual practice of justifying the accuracy and reliability of public involvement in the scientific process of GZ gradually evolved to a more ancillary role as a result of having proven the ability to understand the epistemic object, i.e. appropriate data reduction for science-ready data.

13 February 2010

"It was surprising right? I had spent a lot of time comparing to existing professional classifications, almost getting ready for the battle because an obvious response is they are the public, they don't know what they are doing. And actually we got really very little of that, which is bizarre. Out of about 20 papers submitted, we had one referee who said - this is insane, how can the public do that. And it's the same in the seminars, I have gradually reduced the amount of time I give to claiming that public could do this. Just because no one ever asks." - PI

Soon after the GZ 1 dataset was released, the SL informed the rest of the Science Team that the data set from GZ2 was ready for use, while offering some cautionary advice on how to use it. He noted that with amateurs remotely classifying galaxies through the web-based platform, types of accidental and technical errors could be expected. The SL was therefore indicating that the data set was not ready for immediate analysis owing to the complexity involved in removing all biases, warning the Science Team to take the necessary measures prior to interpreting the data. In an email to the Science Team, he offered a helpful analogy for Observational Astronomers on the peculiarities of the data set. It appeared to be a trade-off between the speed at which the scientific results could be interpreted and the checks that needed to be carried out before the data could be used with confidence. The status of the data therefore called for a different kind of attention and skill in understanding the data. The development of this ability arose as a consequence of amateur involvement in the process. As a result, the epistemological routine of scientists using this data set demanded a change in their objectual practice of analysis when pursuing the knowledge object of their scientific inquiry.

5 August 2010

“It is my pleasure to announce the availability of the science-ready data from the complete Galaxy Zoo 2. This table supersedes all previous GZ2 data releases. It is based on the complete set of GZ2 clicks, correctly retains information about the different samples and images, and removes the influence of 'bad' users.

A fairly complete description of the GZ2 sample selection, data collection, reduction and resulting tables is at [private link] This is essential reading for anyone using GZ2 data.

A friendly warning: Although the GZ2 data is neatly presented in a table, it is not at all well understood. It needs to be tested, calibrated and tested some more before it can be used for science. As many of you are observational astronomers, an appropriate analogy is that GZ2 is a new instrument on a telescope. This dataset represents the first raw images from that instrument (in our case we have already done an entire survey!). The next steps are quality control and calibration: mapping the bad pixels, removing distortions, flatfielding, measuring the gain, read noise, non-linearities, determining the zeropoint. All of these procedures are reasonable analogies for the various steps that must now be taken with the GZ2 data. Obviously the goal is science, and we should get to that as quickly as we can, but if we haven't done the necessary checks and calibrations along the way then we could end up being very embarrassed.” - Science Lead

18 September 2010

“Just a note that the GZ2 results table has been updated. The reduction process now includes the cleaning of multiple classifications of the same object by a given user (apparently usually a case of impatient multiple clicking because of a slow network connection or possibly a browser bug). The statistics now only use the last classification a user makes for each object. This affects at most ~1% of the sample at all, and only ~0.01% are affected enough for it to result in a significant change in their overall classification. Nevertheless, please use the latest version in your work.” - Science Lead

The challenges of ensuring reliability and accuracy that come with involving non-professionals in the scientific process could be distilled at the stage of data reduction, where the data needs to be ‘cleaned’ up as the probabilities of errors can be of a different nature. For example, they could be the outcome of the lack of patience on the behalf of the volunteer or a technical error on the Internet browser; elements that were completely out of the control of the GZ Science Team. Therefore, seeking out and identifying these errors and sources of errors became a part of the reduction process, and eventually an established practice as a result of operating in unforeseeable environments and unfamiliar classification behaviour of hundreds and thousands of volunteers.

The decisions made by the Science Team in terms of measures taken to make a strong scientific case for GZ data, and publishing the method used along with the explanation of the project proved beneficial for GZ's image in their professional community. The success that followed in the later versions of the project builds on the perception of this very ability of the GZ team. Understanding the epistemic object in the early stages of the project was one of the challenges perceived by the team, and having spent considerable amount of effort directed to the understanding and anticipation of the necessary measures, proved to be of great advantaged for the team's reputation in their own professional community.

To summarise, the volunteer annotations call for additional and novel data reduction practices to detect and test for biases and remove various kinds of errors in the classifications. In addition, the amateur classifications had to be compared to expert classifications for accuracy of classifications. Therefore, the epistemic object of 'sensible science' demanded the objectual practice of data reduction and public release, which can be characterised as open science practices. Hence, Open Science was the emerging thought style of GZ as a result of such objectual practices.

There is also a markedly different shift in the quality of perceptual attention given to the data as a result of the data reduction process. For instance, there was an understanding of how to use the data with annotations of public as epistemic subjects, as the SL cautioned. The shift in the quality of attention required to use the data sets to achieve legitimate scientific results, indicated the signifiers for the unfolding of the epistemic object as the scientists pursued it. As they did so they transformed the practices of their epistemological enquiry. The ability of the GZ Science Team to perceive the epistemic object which provided a direction for engagement and demanded a certain quality of individual and joint attention. This quality was instrumental in the evolutionary development of the expertise necessary for running GZ's citizen science, to be discussed further in chapter 8.

7.2.3 Dissemination

This section illuminates some of the research dissemination practices that call attention to measures directly related to citizen science.

A) "Clicks into Papers!"

As discussed in the previous section, using GZ data sets for science required additional measures in analysis and interpretation (stage t2 of translation in Callon *et al*. (2009) or T5 as shown in chapter 6). Publishing scientific papers is the standard with which the GZ team determined its own success. The PI consistently argued that ability to produce scientific papers was what distinguished GZ from previous citizen science projects such as SETI@Home. As with every stage in the translation process, cautionary measures for ensuring rigour in citizen science methods had to be taken. An e-mail exchange below illustrates some of the questions GZ scientists needed to construct in order to ensure their analysis was rigorous, as the PI pointed out possible problem areas created by a GZ Science Team member. He also highlighted the importance of consistency and coherence in the way the data were interpreted between GZ papers. There was a concern that individual team members may not necessarily pay attention to overall consistency in the scientific output, but it was of importance for GZ's reputation as a serious science initiative. Therefore, ensuring that the papers published by GZ scientists using GZ data went through robust checks in how the data analysis was carried out was a critical practice, also supported by the SL who emphasised the need by characterising the dataset as 'unusual'.

3 August 2010

"Thanks for sending this round; I've got some detailed comments but before I get to them I have two high level questions:

If I'm reading the results at the end of section 3 correctly, out of the 5373 galaxies identified in the early Zoo 2 results as having a bar only 3195 are classified by Google are being barred. This seems a little lower than I would expect; I wonder if this is to do with the lower resolution in the Google interface? Have you visually inspected any of the missing bars to try and see what's going on? It would be great to tie this down as it will obviously be important in tying this paper together with others.

You do a good job of making a virtue out of the multiple classifications by the most active users. However, I'm a bit worried that you use only galaxies with more than 3 bar identifications when more than one of these are likely to come from the same user in many cases. Aren't we horribly biasing the data toward the opinions of those two superusers?" - PI

3 August 2010

"I would like to urge all those writing GZ2 papers to circulate drafts early so that we can benefit from each others' experiences in using this unusual dataset before finalising any analyses." - Science Lead

The excerpt above illustrates what the PI labelled as ‘high level questions’ relating to the use of a citizen science dataset, and drew the Science Team’s attention to issues regarding the possible discrepancy in the results linked to technical features of the web browser used. The emphasis was on the need for coherence between the GZ papers and he questioned a possible biased user classification behaviour linked to the data presented in the paper. Therefore, developing and working on the quality control measures framed by the PI were the objectual practices that led the GZ Science Team members to engage with the object of reliability of results. Here was a display of a combination of expertise in the domain of the scientific discipline and citizen science, which was in constant development throughout the process of decision-making regarding the scientific output from GZ.

In addition to the practice of taking measures for reliability, publication decisions for GZ also included decisions, such as which journal to publish in, which reflected the tension the GZ scientists perceived as advocates of 'Open Science' and 'open data’ practices. A request from an open access journal was refused as most of the GZ team members were early career scientists, and the impact factor of journals was critical for working towards establishing credibility and reputation in the scientific community. Even though an open access publication would appear to be congruent with their Open Science values, they preferred to uphold the decision against it. In so doing they reinforced an objectual practice that centered on building reputation, in this case based on impact factors critical for early career scientists.

28 July 2011

“The point is that they are open access journal (a good thing if done well) with low volume (9 articles this year!) and low impact (a bad thing) trying to solicit some decent papers. It’s sad, but for the sake of our careers, I suggest we stick to MNRAS and ApJ.” - Science Lead

However, decisions of publication practice revealed not just tensions and tradeoffs between motives and objects. The GZ Science Team also found themselves practising a previously unfamiliar objectual practice of formally acknowledging a citizen science method or 'a forum project' in a scientific publication. An excerpt from a senior scientist's e-mail to the team indicated the enthusiasm which revealed his emotional engagement with the epistemic object of citizen science practices and, as Knorr-Cetina (1999) argues, further engrossment with the knowledge object.

21 January 2011

“[a]nything to add in Zoo acknowledgements, since a forum project with so many participants is sort of new publication ground? [...] Forum science rules! (Or at least make a significant contribution)." - GZ Science Team Member 2 (Senior Scientist)

B) Representing Galaxy Zoo in the Scientific Community

The role of academic conferences for the GZ team was not only limited to the purpose of communicating their science results to the rest of their professional community, but also to providing a space for promoting citizen science as an acceptable and reliable method of producing scientific knowledge, to establish it as normal scientific practice. Therefore, in the initial stages of GZ, the object was identified as the acceptance of citizen science as a reliable method in Astrophysics research, inspiring the following objectual practices.

One of the first conferences discussed between the team during the period of data collection was the National Astronomy Meeting (NAM) in April 2010. GZ's SL identified some options for sessions that might be of interest and importance to GZ. The selection indicated three science sessions that focused on the results, methods and surveys in Astronomy and the fourth one on public engagement. Even though the team had significant experience in the latter area to qualify for the session, the PI urged the Science Team to focus on the science sessions, indicating a reorientation of the priorities of the team to the epistemic object that was most important to the project, to establish credibility within the scientific community as a science project and not an 'outreach' project. The PI had expressed concerns about being seen as such an ‘outreach’ project during his interviews as discussed in the following chapter.

3 March 2010

“Has anyone submitted an abstract to talk about Galaxy Zoo? Galaxy Zoo could fit in several sessions:
P01[34]: Galaxy Formation and Evolution in the Low-Redshift Universe With Multi-Wavelength Spectroscopic Surveys
P10: Software Astronomy: How to Do Astronomy by Looking Through a Computer
P26: The Sloan Digital Sky Survey: the Legacy and Future
P33: Public Engagement in Astronomy, Solar and Solar System Physics” - Science Lead

3 March 2010

“I’m going, and should probably give a talk. We should try and cover at least the three science sessions.” - PI

[34] P01 and the following numbers indicate the sessions from the conference.

More than a year after the decision to have a visible presence in the science sessions, the strength of GZ results qualified it for an entire session at the American Astronomical Society (AAS) Meeting in May 2011, with six talks scheduled and chaired by the PI. An entire conference session on GZ's science results marked the recognition of a distinct identity of GZ within the professional community of Astronomy research, indicating that the objectual practices aimed at the object of establishing GZ as a credible science endeavour were successful. This event marked the moment in GZ's history where it positioned itself as pursuing the epistemic object of galaxy evolution in its own disciplinary field and established its identity as a coherent research group, as the extract from the GZ blog post on June 3, 2011 post reveals:

"Cosmic Evolution from Galaxy Zoo: Galaxy Zoo (www.galaxyzoo.org) is familiar to many as a hugely successful public engagement project. Hundreds of thousands of members of the public have contributed to Galaxy Zoo which collects visual classifications of galaxies in Sloan Digital Sky Survey images (and most recently Hubble Space Telescope) using an internet tool. Classifications from phase one of Galaxy Zoo (the basic morphology of SDSS galaxies) have recently been made public.

Galaxy Zoo has also shown itself, in a series of peer reviewed papers, to be a fantastic database for the study of galaxy evolution. In this session Galaxy Zoo team members will hi-light some of the most recent scientific results using Galaxy Zoo data, including the first results from phase two of the project (which collected more detailed morphologies)."

The conference was not only unique for the presentation of GZ's scientific pursuits as a collaborative effort to the scientific community, but also because one of its volunteers and active forum moderator was invited to attend the conference as a representative of the volunteers. One of the Science Team members summarised the session for the volunteers on the GZ blog (see Figure 7B) and also shared her presentation slides, links to abstracts of the other GZ presentations and a link to a post from another Science Team on a sister project's blog. The efforts to communicate their activities to the volunteer community was a practice that the Science Team began engaging in more regularly and will be explored in further detail in section 7.2.5 of this chapter.

Our Galaxy Zoo Session at the Boston AAS

As many of you may know, several Galaxy Zoo scientists were at the recent meeting of the American Astronomical Society meeting in Boston, USA. This included Chris, Kevin, myself, Carie Cardamone and Brooke Simmons; Lucy Fortson (who recently did her first blog post about a review article we wrote), Alfredo Carpinati (from UCL) and Ivy Wong (who recently moved from Yale back to her native Australia).

Galaxy Zoo volunteer and forum moderator, Alice was also there – and has written about some of it on the forum (under "why I'm going to be a bit quiet for 3 weeks"). Kevin has written some of his AAS highlights for the Planethunters blog.

But nothing has been written yet about our wonderful session on the science from Galaxy Zoo (except from the @galaxyzoo Tweets during the session), so I thought I'd take a bit of time to tell you about it.

It's been a busy few weeks for me in the lead up to and following the session last Wednesday, so I hope you'll forgive me for not doing this sooner.

Figure 7 B: Excerpt from GZ Blog Post on the AAS Conference, June 3 2011[35]

Although the session at AAS in Boston was well-received, a similar bid for a GZ session at an Astronomy meeting they had been to previously was not accepted. However, the exchange among the GZ Science Team indicates a shift towards an interest in presenting not only the science of GZ, but other Astronomy related sister projects, expanding the sense of citizen science community even if the members themselves are not directly involved in the science. Conferences appear to provide such a space to expand. The statement by the SL reflects the sense of identity as Zooniverse scientists, expanding beyond their domains of expertise into collective citizen-science led expertise.

[35] GZ Blog Post titled: Our Galaxy Zoo Session at the Boston AAS http://blog.galaxyzoo.org/2011/06/03/our-galaxy-zoo-session-at-the-boston-aas/

28 July 2011

"See the below call for session for NAM 2012. Should we think about some sort of Galaxy Zoo/Zooniverse session? I really enjoyed the one in Boston, and I think it could go down well with the NAM audience. :) Perhaps there are not enough new Galaxy Zoo results to justify doing another session on the science from Galaxy Zoo (?), but how about a session on "Doing Astronomy with Citizen Scientists" in which any of the astronomy related Zooniverse projects could report[...]?" - GZ Science Team Member 4

28 July 2011

"I think that's a nice suggestion. It would be great to get people from all the Astronomy-related Zooniverse projects together to show off to the community, if there is enough interest. We should see what everyone on the other projects thinks of the idea." - Science Lead

7.2.4 DotAstronomy 3: Creating Spaces for New Epistemic Communities

The conference that captured the spirit of GZ-led citizen science was held at Oxford in April 2011 and was one of the two major events I had the opportunity to observe as a participant. This conference was not designed for discussing the scientific results in the field of Astronomy, as its objective was to share and explore web-based tools and innovative ways of conducting research. The participants varied from academic institutions to Wikipedia, Amazon Web Services, Microsoft Research and SETI Institute (Search for Extra Terrestrial Intelligence) - one of the first pioneers of distributed computing projects that inspired a generation of crowdsourcing. A brief description of the event provided by the organisers was as follows:

“The Internet provides an incredible platform for astronomers and astrophysical research. .Astronomy (pronounced ‘dot-astronomy’) aims to bring together an international community of astronomy researchers, developers, educators and communicators to showcase and build upon these many web-based projects, from outreach and education to research tools and data analysis.

.Astronomy events are held (almost) once a year […] These meetings bring people together to talk, make and do cool stuff for just a few days. The events are focused, but informal, and encourage collaboration on new ideas that can benefit astronomy in a variety of ways."

The conference deliberately lacked a rigid structure with only a few plenary speakers every morning, followed by 'unconference' sessions and 'hack days' in the afternoons. Only the speaker sessions were pre-planned and the unconference and hack sessions were decided on the day as participants brainstormed and volunteered ideas for the issues and skills they wanted to discuss and learn from each other. The objectual practice of design decisions of structuring the conference with open and flexible sessions was inspired by the object of spontaneous and serendipitous sharing of issues, interests, skills and solutions that would not have emerged within a predetermined rigid structure of a conventional academic conference.

The talks varied in domain topics and included the Director of SETI who discussed the need to tap into human recognition capabilities to train the next generation of computer algorithms, different technologies that support a large scale data intensive project, ending with an expression of interest in working with citizen scientists. SETI was the pioneering project in distributing computing and citizen science and is known to work on a distributed computing model whereby human effort is minimal. For the founding director of an established project to approach the Zooniverse team, based on the success of GZ for their expertise in citizen science, was a significant turn of events.

The other organised talks were from representatives from Wikipedia, Amazon Web Services - a platform which hosts the Zooniverse projects, a biologist and a prominent advocate of Open Science, an engineer leading the grant in Europe to build robotic telescope which will be partially remotely controlled by public volunteers, and an expert in Python (a programming language). The conference was also open to non-Astronomers who were interested in similar issues of new directions and methods in scientific research. Only three out of the 15 talks planned were delivered by members involved in the Zooniverse projects, which was a considered decision.

As the PI later explained in an interview, the dotAstronomy projects were not intended to focus on solely Zooniverse projects, but were intended as a space for the members of the scientific community who shared similar values and interest, to learn and share new ideas and methods. The decision to keep the number of Zooniverse participants relatively low was also deliberate. He nonetheless reported that they had found collaborators during the conference who later got involved in Zooniverse projects, the founder of the conference being a good example. Therefore, dotAstronomy served as a distinctive epistemic space where professionals voluntarily participated, identifying themselves as potential collaborators and contributors to citizen science.

This event also marked the first real time observation in the present study of social media use, mainly Twitter, as a parallel channel used by scientists to communicate with peers in their community and for communication with the public. Nearly all of the participants had an active Twitter account during the conference. The organisers encouraged the use of this social media channel and as the PI described in a follow-up interview:

11 April 2008

“I opened my Twitter account at dotAstro 1 and really started using it at dotAstro 2. It's a backchannel for the conferences, it's wonderful to sit in a lecture and see all these incredibly smart people discussing it at the same time and posting links. I always lose followers in dotAstronomy, no one else cares. But that's fine.” - PI

The significance of Twitter usage lay not only for communication among the participants, but also the public, as it is publicly accessible and delivered to individuals subscribing to the tweets which include their professional network. This was an act of identifying themselves as participants at the dotAstronomy as professional scientists engaging in creating and sharing new methods and new web-based technologies, but also new approaches in public engagement in science such as citizen science. In a way, dotAstronomy became a space to voluntarily gather potential contributors and collaborators to the cause of Open and Innovative Science.

This epistemic space served as a critical resource not only to identify new collaborators, but for ideas and expertise from fields other than Astronomy to focus their attention towards solutions that could potentially shape the direction of citizen science as a movement. There was the potential for them to create an epistemic network for GZ and its sister projects - an indispensable objectual practice for the emerging style of GZ’s citizen science.

Below are some of the statements made publicly on Twitter by the participants of dotAstronomy 3, which allows a sense of the issues of interest, lessons for reflection and learning that the participants were engaging in during the event:

“Citizen science is most effective when it creates a change of perspective among citizens”

“Interesting point by (PI) about browsing habits tending to limit serendipity to within our interests. Need bigger jumps”

"All the interesting science is on the edges where things are broken"

“Learning about the cool things coming in @the_zooniverse”

“Big data unconference turns into astronomy software debate turns into career conversation.”

Following dotAstronomy, the Zooniverse team continued to organise events to encourage scientists in the research community from various fields to participate in innovative thinking and promote engagement with the public.

7.2.5 Communicating with the Zooites

The first instance of communicating with the volunteers occurred right after the launch of their first project in 2007 as the GZ team (As there were only a few founding members at that point) began receiving hundreds of emails after the initial launch of the project. As I have already indicated, this unprecedented demand on communication led to the building of a forum on the GZ website, which was the first established platform for communication between the scientists and the volunteers. Initially, the primary epistemic object was the understanding of how to sustain the incoming traffic of volunteers. During the course of the project’s inception, success and launch of various versions (for e.g. GZ 2, GZ Hubble), the GZ team responded with the objectual practice of communicating with the volunteers through various channels such as Twitter, the GZ official Blog, the mailing list and live Google Hangouts. Their contributions provided explanations for the various decisions they made as GZ scientists. In this section, communication refers to the exchanges between the GZ Science and Development Teams and the volunteers, in addition to the texts and tasks on the GZ website. Therefore, it does not include FAQs or tasks tutorials, which could also be argued to be a form of communication.

Essentially, the purpose of communication was to build a foundation of trust with the volunteers that not only ensured traffic on the GZ website, which translated into completed annotations through the GZ website, but it also created a sense of collaboration between the GZ scientists and the volunteers. Therefore, the epistemic object at the core of communication was trust, which included the idea of ethical participation on the behalf of both parties, that enabled public engagement in scientific inquiry. The practice of science communication with the volunteers allowed for negotiating the social contract (Nowtony, 2001) of mutual ethical practice of carrying out science in the open. It was the active engagement of the GZ Science and Development Teams in the communication practices that emerged and evolved at times in response to the demand and at times as an experiment, inspired by engrossment in the epistemic objects.

The GZ Science and Core Team operated on multiple communication channels, and snapshots from different platforms have been presented as data throughout the study. Therefore, the examples presented in this section are selected to demonstrate the variation in communication practices that reflect the perceptions and understanding of science communication as objectual practice and the volunteers as epistemic subjects.

One of the practices that the GZ Science Team was particular about was the need for acknowledgement of the volunteers. These responses were from the Science Team at the forum on the event of the third anniversary of GZ:

11 July 2010

“Hello Zookeepers, Just thought you might all like to see the Galaxy Zoo 3rd birthday card http://www.galaxyzooforum.org/index.php?topic=277992.0 and wondered if any of you would like to blog it?” – Science Team

In response to the mail from the Zooite forum moderator, the PI and two most active members of the Science Team made the following posts:

“Thank you so much! This feels like a surprise birthday party...It's tempting to spend paragraphs looking back at the past, but it honestly feels like the blink of an eye (Our three years is approximately one-onehundredmillionth of a percent of the age of the Universe). It seems only yesterday that I opened my inbox to find tens of thousands of messages about the Zoo. [Text omitted] Instead, perhaps we should look forward - the next year is going to be the most exciting yet for the Zoo, as results from Hubble begin to flow and a new project or three spring up. Thank you again.” - PI

“Wow! Thanks to everyone for making Galaxy Zoo a huge success, and showing everyone the potential of citizen science. Along with the whole Zooniverse team, I am looking forward to working with you all (and all your lovely data) for many more years to come.”

“Wow, thanks all! You made it possible.”

Some of the members of the Science Team at GZ were active in science communication prior to the GZ’s launch, and worked towards promoting GZ’s brand and work by engaging in outreach podcasts in Astronomy, and writing science articles for a general audience. These practices within GZ were not only aimed at promoting GZ as a science project, but also to recruit new volunteers and sustain interest in the project within the scientific and amateur communities of Astronomy. The involvement of Science Team members in science communication activities gradually increased, not only in frequency but also in the communication platforms and technologies used. An in depth analysis of the technology-enabled communication was not carried out for this study, as that would require an entirely new thesis, however, the social media communication through blogs, internal mailing lists, and Twitter were periodically observed and examined to determine issues of importance and patterns relevant to the object of this study.

The email excerpt from the team mailing list below reflects the encouraging attitude towards science communication. At this point, the purpose and criticality of science communication was understood to be a form of common knowledge indispensable to the success of GZ as a citizen project and the lack of explicit justification is an indication of a common understanding established among the team.

11 July, 2010

“Thought the rest of you might want to see this. Calvin and I also contributed today's podcast for the 365 days of astronomy series [...] this was actually incredibly easy to do, and so I'd like to do a series of these. If anyone would like to volunteer to be interviewed about anything Zooy, let me know.”- PI

Similarly, acknowledging and promoting the science communication efforts of the members was an important mechanism to direct the collective attention of the group to promote and encourage such practices. Among the GZ Science Team members, the use of Twitter as a social media platform was a common way of communication with other GZ members in addition to peers in the scientific community. It positioned the discussions around GZ in a public sphere, which included non-GZ scientists, institutions and media who might be following the work of the scientists affiliated with GZ. Therefore, publicly mentioning the science communication efforts of team members was a practice that was observed frequently.

29 September 2011

“Well done to Calvin for his articles about Science from Galaxy Zoo and Citizen Science in general in Sky and Telescope[36] this month. :) Just had them pointed out to me on Twitter, and will try to get a copy when I can.” - GZ Science Team Member 4

GZ’s science communication practice also latterly included experimentation with live interaction on Google Hangouts with members of the Science Team discussing GZ-related science and taking questions from Twitter. The structural flexibility of GZ’s organisation can be attributed to such spontaneous experimentation practices with social media for communicating with volunteers. These practices were indicative of the culture of innovative practices within the GZ team. The e-mail from one of the second-generation GZ Science Team member (active members who joined GZ after the launch of the first GZ) reflected that ethos of innovation-driven practice in science communication.

One of the emerging practices supported by Internet technologies such as live online chat hosted by Google Hangout that the Science Team decided to try as new medium of communicating with the GZ volunteers. Below are extracts from GZ’s Twitter feed inviting interested parties to ask questions for the live chat (Figure 7C) and an official blog post summarising the live chat session for the GZ volunteers.

[36] Sky and Telescope is a popular Astronomy magazine aimed towards an amateur audience http://www.skyandtelescope.com/

Figure 7 C: Screenshot of GZ's Official Twitter Feed before the First Google Hangout

Spiral Galaxies and the Future of Citizen Science: a Live Chat

Last week Karen Masters suggested that we start doing Galaxy Zoo live chats a little more often. I thought that sounded like a great idea, and we figured we'd just have an informal chat about whatever galaxy/Zooniverse topic we felt like discussing that day.

We were joined by Kyle Willett and Kevin Schawinski, and the four of us started talking about this paper, which presents an automated system for classifying and measuring spiral arms. It compares to Galaxy Zoo 2 data within the text, and we talked about what the fact that the computers did pretty well means for the future of Galaxy Zoo. We didn't prepare anything in advance, and I didn't even start reading the paper until about 20 minutes before we got going. So my favorite part of the chat is where I put forward a few definitions of pitch angle and get them all wrong. Science in action!

We also introduced the jargon gong, which we used on each other whenever one of us said something in insider-speak. I think this is a feature worth keeping, and we also plan to invite viewers to gong us themselves via Google+ or Twitter for the next chat.

Figure 7 D: Excerpt from the GZ's Blog Post Summarising the First Session on Google Hangout [37]

17 January 2013

"After the success of the last GZ live chat, Carey and I are planning another Google Hangout. We're thinking of having slightly more frequent chats, just very informal, and shorter than what we're used to (more like 15 minutes instead of

[37] Source Link: http://blog.galaxyzoo.org/2013/01/25/spiral-galaxies-and-the-future-of-citizen-science-a-live-chat/

30). We're planning to try it out next Friday [Text omitted]. If you'd like to join us please feel free!”

One of the Google Hangouts was observed in real time on the February 8, 2013, which was the second ever hosted. Much like the structure of GZ itself, the structure of the video chat was observed to be informal and open to serendipitous occurrences. In this instance, it was an unplanned appearance by a prominent scientist who willingly participated. The live video stream was initiated by four members of the Science Team who answered questions taken from Twitter. In order to ensure that the language used was for a general audience, a ‘jargon gong’ was introduced. It was employed every time a Science Team member used technical jargon, which the other members would alert them through a feature of Google Hangouts or members of the public by typing ‘jargon gong’ on Twitter. This was also revealed to be an unexpected exercise for the Science Team members in terms of assuming certain level of knowledge on the part of the audience. For example, there was a gong for ‘infrared’ but ‘redshift’ was taken as a given. It is also a representative case of an improvised solution in response to a perceived problem in communicating with the Zooites, which was a recurring characteristic demonstrated by the GZ team in practices relating to citizen science.

Unexpected events appear to have regularly occurred in GZ activities, partly owing to their open and flexible practices that permitted such occurrences. During the observed live video stream, one of the leading Astronomers who was also introduced to the audience as a ‘morphologist extraordinaire’ walked into the office of one of the GZ’s senior scientists. Unaware of the public broadcast at first, he began engaging in a conversation about galaxy morphologies and his positive impressions of GZ's work. When he was made aware of the live public broadcast he had accidentally joined, he continued sharing his anecdotes. It was also the first meeting he had ever had with the rest of the GZ team members. The live video chat lasted longer than planned and took a completely unanticipated turn with the arrival of an unexpected guest. This flexibility was not uncharacteristic of GZ where such serendipitous events were not only welcomed but actively supported. The technology available to the scientists allowed them to conduct real time conversations with their volunteers supported by Twitter, but it was also the ease with which the team handled the unexpected that was characteristic of their science communication practice. What was evident in the live chat was how the spirit of experimentation in action among the GZ team members in shaping their objectual practice in using new communication technology for science communication.

Although the GZ team has been proactive in communicating with their volunteers from the very beginning of the project, and have been successful in maintaining trust and open communication, their efforts have encountered tensions. Prior to one of the Zooniverse conferences where issues brought up by Zooites were scheduled to be discussed, one of the members noted a post in the GZ forum which lends an insight into the problematic tensions among some of the Zooites and the communication efforts of the team.

23 Mar 2013

"Some longtime Zooites are deeply incensed at the SN [Supernovae] Zoo analysis of voting patterns, feeling somehow betrayed at any analysis beyond the stated use of finding SN. Also, there are issues of telling users when they've had a simulated image, and once more requests to know how psychologists are affecting Zoo design. There is also an intriguing mention of mandated open-access publication and whether any Zoo results should ever be behind paywalls (arxiv notwithstanding). Just sharing the heartburn." - GZ Science Team Member 2 (Senior Scientist)

Some of the issues brought up by a few Zooites therefore revolved around a possible expression of breach of trust, which was the core epistemic object driving GZ's science communication practices, and a demand for what was perceived as an ethical practice by the Zooites, of making all GZ publications open access. The discussion on the forum reflected a sense of expression of rights as perceived by a few Zooites. However, dissent by a few was sufficient to warrant a discussion among the GZ team. The importance lies in the approach that was revealed in managing the tension. In response, the PI elaborated on the issue that there had been a history to the conversations and previous efforts to reconcile differences had not been successful. He therefore recommended suspending the issue. He also acknowledged the efforts of the team in communicating with the Zooites and making clear that the issue was open for discussion if any of the team members had serious concerns. His response reflected a recognition that such problems may arise despite their best efforts and in that light, disengaging with the disruptive few was acceptable. Here non-action as a response was seen as a legitimate objectual practice in certain situations.

24 March 2013

"Just a note to say that we've tried again and again to have a sensible discussion with these users, but to no avail. If anyone thinks it's creating enough fuss that we should intervene somehow please say, but otherwise I suggest leaving alone.

The irony is I think we're doing better than we have in a long while about keeping people informed about what's going on (thanks in particular to the GZ team members doing that!).” - PI

One of the founding members of GZ suggested addressing the issue of open access in the next live chat session, to which another member replied with skepticism. The issue of Open Science was discussed briefly in one of the email exchanges presented in the section on research dissemination, and perhaps the team was not ready to have a debate about open access at this point.

24 March 2013

“I'd be wary of doing anything that might stir them up further; and nearly everything stirs them up further. The TL's post in the forum thread (for example) is excellent, but I have little hope that it will actually help.

Just because the vocal minority often communicate with hostility, irrationality and entitlement doesn't mean they're necessarily always wrong. And, particularly if you read the posts of the others on the same thread, there is signal in the noise. I think we haven't always been great about communication, but we have been doing better and will keep trying to do so in the future.”- - GZ Science Team Member 5

The response of the second-generation Science Team member quoted above indicated that they were actively monitoring and paying attention to the discussions on the forum not just as general noise. However, the Science Team was selective about what they needed to act on. The epistemic object of trust between the volunteers and the GZ team members remained a critical one, continuing to drive innovative and experiment objectual practices.

What was evident in the discussion between the GZ Science Team about their public effort was that science communication was critical for GZ’s citizen science practices. It defined their thought style, with trust building as an epistemic object that they continually pursued by exploring new tools, technological and semantics and designing the epistemic architecture to carry out those practices. In recognising the public as epistemic subjects, communicating for the GZ Science Team became indispensable. Therefore, it can be argued that the communication practices of GZ team evolved in response to their perception of the volunteers, supported by the experimental ethos within the GZ team to embrace and explore new technological tools for communication.

7.3 Conclusion

The objective of this chapter was to examine the practices that are driven by the constant pursuit of the epistemic objects recognised by the GZ Science Team, while taking citizen science forward as a reliable method of conducting scientific inquiry. From negotiating tensions involved in setting up new collaborations, to the rigorous pursuit of reliability in data reduction techniques, the use of citizen science data sets, openly releasing data sets for use by non-GZ scientists and dissemination practices of results, the team demonstrated continuous engrossment in multiple epistemic objects simultaneously by virtue of being a collaboration, resulting in the refinement of multiple objectual practices. Although the practices are presented in categorised sections, they often occurred simultaneously, therefore the GZ Science Team was presented with multiple epistemic objects. In that pursuit, the Team demonstrated how they learnt to negotiate and understand those emergent epistemic objects that were recognised by the PI and the Science Team to be of importance and which were relevant to citizen science as an unfamiliar and emerging practice. Embracing innovative approaches in refining existing practices and an experimental attitude to learning that practice was reflected in the Team's ethos as the examples in the section have demonstrated.

What was also evident was the emerging role of active second-generation scientists in GZ, the members who were recruited after the launch of GZ. In applying Fleck's (1935) notion of a thought style to GZ, one might argue that a thought style was recognised with a successive generation of its followers who preserved and refined some of the established practices exhibited in the initial thought style. GZ's emergent thought style was demonstrated in a gradual increase in the activities of the second generation scientists, especially in their role in science communication, which was established as one of the most critical practices driving GZ's science and its knowledge production process, hence shaping its epistemic architecture.

Following the identification and examination of the roles and responsibilities of the Science Team, the focus of the next chapter will be the development expertise of the Core Team, through the practices and motives of its two main participants, the PI and the TL. The relationship between the Core and the Science Teams was crucial as the Science Team negotiated the process of scientific output by dictating where the public was positioned as epistemic subjects, and located in the process of knowledge production. In essence, they were designing the volunteers' interaction with the data, creating a space for them within the epistemic architecture.

However, over a period of time, although the Core Team shared the Science Team's primary objective, their attention also branched to the wider implications of their practices. In brief, in addition to the Science Team's primary interest in - how to make citizen science work for science, the Core Team also became interested in - what can citizen science do for science or what citizen science and its methods mean for science? Some of the discussions between the Science and the Core Teams that were presented in this chapter reveal the tensions between the two parallel but equally important set of motives and the decisions that were taken in the interest of the sustainability of the project. All this was also taking place as Zooniverse initiative grew bigger in volunteer strength, media popularity and its affiliated projects.

Chapter 8

Expertise for Citizen Science

The aim of this chapter is to reveal the practices and expertise of the Core Team developed through their work in Galaxy Zoo (GZ), shaping a distinctive epistemic architecture, while engaging in citizen science. I shall argue that the evolving nature of the expertise in managing citizen science projects is emergent through iterative actions which comprise of practices being created in response to the pursuit of citizen science as an epistemic object.

By the end of the fieldwork, the Principal Investigator (PI) and Technical Lead (TL) had been recognised as experts in building citizen science projects and as by the end of 2011, they had launched more than 15 citizen science projects (By September 2014, the Zooniverse team had had built 30 projects). In chapter 5, GZ was described as a knowledge production system comprising of various forms of distributed intelligence, expertise and knowledge, and the relationships between the various elements that comprise the epistemic architecture were identified. In chapters 6 and 7, it was revealed that a set of practices which involved operating across expert and non-expert boundaries was critical in successfully running GZ. These are practices which have developed through various iterative processes in the project. In this chapter, the types of expertise that are evident in managing GZ as a citizen science project will be examined. The initial epistemic architecture was built on expertise which could be attributed to the scientific background of the founding members. However, the expertise for citizen science and the epistemic architecture had co-evolved, shaping each other over the course of three different iterations of GZ by the time the data gathering was concluded. What guided the evolution of the epistemic architecture were identified to be the design principles that began to emerge and undergo a process of refinement over the course of various iterations of the project.

8.1 Identifying Relevant Types of Expertise for Citizen Science

Before embarking on an examination of expertise for citizen science, it is necessary to identify the different kinds of expertise to be found at play in the practices observed and rehearse briefly how the definitions of different forms of expertise are employed in the present study. The constant pursuit of understanding and managing the relationship between experts and non-experts at the boundaries of science provided the grounds for the use of different sets of expertise that emerged from GZ's innovative citizen science practices. This chapter primarily addresses the third sub-question of this research study, which is 'what kinds of expertise are evident in these citizen science practices and how are they developed?' While the two previous findings chapters have presented analytical accounts of scientific practices and practices for citizen science, this chapter addresses the remaining part of the overarching question through an analysis of the interplay between existing and emerging expertise in GZ's citizen science.

At the inception of the first GZ project, the problem at hand of time-consuming classification of thousands of galaxies, was met with a potential solution of recruiting volunteers from the non-scientific community. Finding the solution required a combination of practices and expertise that were not normally a part of the standard expertise required for postdoctoral researchers within an academic research institution. I have chosen to describe the set of expertise among the Core Team members as an ‘expertise portfolio’ that was identified in the work of the PI and TL. The term portfolio has been employed to reflect evidence that suggests how different kinds of expertise are called into play in various activities that are part of the practices which have emerged in the development of GZ.

The relevant categories of expertise are as follows:

i) Domain Expertise in Astronomy D(A) refers to the expertise characterised by the use of domain-specific knowledge in the academic discipline of Astronomy/Astrophysics, which qualifies an individual as a professional to be able to contribute to scientific inquiry within that field of research. Professional recognition of domain-expertise is what provides the legitimacy required to be recognised as a ‘scientist’. The participation of an individual in a systematic scientific inquiry within an academic institution is limited or restricted by the level of domain expertise one demonstrates. One of the unfaltering features of GZ is the utmost priority given to achieving scientific results, therefore demonstrating significant levels of domain expertise at the core, which appears to be critical to the expertise evident in the work of the PI and TL.

ii) Interactional Expertise (IE) refers to the ability of individuals to communicate with experts outside of their domain of expertise. The term was proposed by Collins and Evans (2002) in their analysis of the work of physicists and relates to the ability of social scientists to talk about Physics without being able to act as Physicists. Interactional expertise in this study refers primarily to the ability of the Core Team members to communicate their goals and purposes of GZ when solving a common problem, in situations where a common knowledge may not be shared with other team members or collaborators. In the present study, this form of expertise requires the expert in one domain to have sufficient knowledge and skills to communicate with an expert in another domain facilitates joint problem-solving without having to call for in depth domainexpertise. Interactional expertise was exhibited in discussion between the Core Team and those outside the team including other experts such as developers, media agencies and education researchers. Working with these groups is not typical of academic researchers. However, the ability to effectively and successfully interact outside their professional boundaries is attributed to the next category of expertise, relational expertise.

iii) Relational Expertise (RE) comes into play when individuals increasingly interact outside of their professional boundaries, in ways which requires them to recognise, acknowledge and understand the standpoint of the other and take actions by aligning motives that matter to both parties. “Exercising this additional type of expertise is not simply a question of collaboration, but is itself a complex phenomenon. Rather, this additional capacity involves recognising how others interpret and react to problems and aligning one's own interpretations and responses to theirs to produce enriched understandings and practices.” (Edwards A, 2010:2).

First proposed by Edwards A (2005), the need to identify relational expertise arose with increasing professionals collaborating with other specialist practitioners on complex problems. Edwards A (2011) suggests that relational expertise arises through a dynamic process of i) working collaboratively to expand "[t]ask being worked on by recognising the motives and the resources that others bring to bear as they, too, interpret it; and (ii) aligning one's own responses to the newly enhanced interpretations with the responses being made by the other professionals while acting on the expanded object" (Edwards A, 2011:34). She argues that while working with others, one's existing expertise does not suffer dilution as a consequence of weaving in the conceptual resources and motives of others into a joint activity. In fact, she suggests that relational expertise enhances professional practice, by demanding that practitioners become more aware of their own expertise and the contribution it makes to collaborations. She distinguishes relational expertise from 'interactional expertise' proposed by Collins (2004), which has been introduced in this section. While interactional expertise requires minimal expert knowledge for communication and assumes a level of immersion into the domain of other expertise, relational expertise requires recognition of what others can contribute to the shared activity, while explicating what matters to you (Edwards A, 2012).

This type of expertise plays a key role in allowing the GZ team to engage in collaborative work, not only beyond their professional boundaries defined by their domain-expertise in Astronomy and Astrophysics, but it also enables them to work collaboratively with various teams within GZ, namely the Science Team, the

Development Team and the Education Team. Relational expertise was demonstrated to be a dominant category of expertise exhibited by the PI and the TL, some instances of which were presented in the previous chapter.

iv) Adaptive Expertise (AE): Involving volunteers from the public directly into the scientific process requires opening up of the boundaries of science. The process of identifying the primary task required deconstructing the steps of scientific inquiry into activities that could be completed by individuals with minimal prior training. Those steps placed the boundaries of scientific practice into the public domain, and the feedback format and mechanism was designed to meet the requirements for data analysis by the Science Team. Such practices are disruptive in the sense that they do not conform to the current established and accepted methods of scientific enquiry and may be considered controversial. Through such practices, GZ members pushed their own boundaries into other domains and challenged the boundaries within and outside of their own professional/expert sphere. For instance, by inviting the public to participate in ‘real science’, not only did GZ benefit from the multiplied cognitive capacities of its volunteers, but unprecedented discoveries occurred as a result of that entrepreneurial undertaking. This ability of the GZ project to successfully implement citizen science methods for scientific inquiry and find innovative solutions for sustainable practice of citizen science is characterised as adaptive expertise.

As Hatano (1982) argues, the development of adaptive expertise calls for a varied nature of tasks and playful situations and environments where the values lie more in understanding, than efficiency in performance. By challenging and disrupting the boundaries of professional practice of scientific research, GZ scientists were able to permeate other kinds of boundaries to acquire resources from educational funding agencies, and form collaborative alliances within their disciplinary field and outside, for example, machine learning and psychology. They were greatly enabled by the technological tools available to them. Engaging with the public had unprecedented outcomes for GZ, which shaped how the epistemic architecture was constantly reconfigured. GZ created a new boundary space that was open to professionals from other domains of expertise and from the public sphere. In doing so, the different GZ Teams engaged in disruptive practices operating at the margins/boundaries of these different spheres. They repeatedly demonstrated capability to work with such non-routine practices as innovative solutions to achieving their science goals, is a key feature of adaptive expertise in the context of citizen science.

v) Technical Expertise D(T) in the context of this study, is a type of domain expertise which is characterised by informed practice in relation to the technological elements involved in building, running and sustaining an online citizen science initiative. Scientists in the field of Astronomy may have the technical expertise to build such projects, but it is clearly not a required element of their expertise portfolios as scientists. The technical challenges for GZ were of a varied and incremental nature from building a website in order upload and host the primary task of images to be classified by volunteers at its initial stages, to designing and building multiple scalable projects that ranged outside of their disciplinary domain. Therefore, the technical expertise of building citizen science projects was transferable across different research disciplines. This capacity was demonstrated by the various projects that the Zooniverse Team (including GZ's Core Team and the Development Team) built including: transcription of weather data, whale song classification and identifying extra terrestrial signals. The technical expertise of the Zooniverse Team also directed them towards collaborating with machine learning experts near the end of the data gathering phase, which was not anticipated in the initial stages of GZ's inception. This collaboration resulted in their working towards improving their approach to citizen science, taking it from a crowd-sourcing approach to a more refined targeted approach; in particular, matching individual classifiers with the tasks they might be most proficient in carrying out. The adaptive nature of the technical core of Zooniverse allowed for professionals from other technical and research domains to contribute to solutions by offering ways of moving citizen science forward that had not previously been achieved.

The interplay of these different types of expertise will be examined in the following sections where critical practices of the GZ Core Team will be presented through individual and joint accounts of their practices, and responses to the challenges perceived by the PI and the TL. It will be argued that the adaptive dimension of expertise arising in managing citizen science projects, manifests itself in both domain expertise and technical expertise and might allow for a greater degree of relational expertise. Although their domain and relational expertise are strongly at play in decision making as the Core Team, it is adaptive expertise that lends the dimension of innovative thinking to problem-solving as a response to the emerging unprecedented challenges.

It must be understood that the GZ Core Team not only began to be recognised as experts in building and running citizen science projects, they were pioneering what it meant to be experts in building citizen science projects and therefore, defining the domain-expertise of running such projects. Working in a field that lacked existing set of best practices and expertise for reference, the development of the Core Team's expertise and understanding did not follow established stages of expertise acquisition (Dreyfus and Dreyfus, 2005). The development of the Core Team's expertise was not an apprenticeship in running citizen science. They were, navigating unchartered territories of potential practices in pursuit of multiple epistemic objects, some of which were examined in the previous chapter.

They established themselves as experts in building and running citizen science projects by constantly seeking out challenges and problems, and further expanding the potential of citizen science. It is that trajectory of intentional expertise development that is discussed, by following the evolution of the expertise of the two key Core Team members. The notion of intention in relation to the expertise of the PI and TL is particularly relevant, as it allows recognition of the intention to construct a responsive mechanism for citizen science and involves aligning the motives of both citizen science and a robust scientific practice, therefore engaging in collective meaning-making.

8.2 The Collective Expertise Profile of the GZ Team

One of the key factors for the success of GZ in directly involving the public as participants in the process of scientific knowledge production, has been the numerous practices of calibrating the reliability of citizen science methods to that of the standard of scientific practice of professional scientists. The success that matters to the entire GZ team was defined in terms of acceptance of the science results by the scientific community, in the form of peer-reviewed publication and citations by members of that scientific community. As chapter 7 explained, from the very beginning, measures were taken to prove how the results published were as reliable as any other academic scientific study. This was constantly one of the major concerns for the PI who viewed achieving the highest scientific standards in research output to be of utmost priority.

In order to examine the development of the expertise of Core Team members in running GZ and its impact on GZ, I have drawn on a framing which allows examination of both the individual intentions of the two key informants, and also how they were reflected within the broader 'thought collective' that was GZ. To do this, I draw on the theoretical framework proposed by Fleck (1935).[1] He argues that almost every individual belongs to several thought collectives, for example, a professional scientist might be member of a political group, a religious sect, a musical group and a outdoor enthusiasts' club. He also contends that it is most likely that individuals belong to thought collectives that are more distant to each other in order to avoid conflicts between different thought styles. He draws a distinction between the esoteric circle, which comprises of the specialists in the thought style and the wider group around them known as an exoteric circle. Furthermore, he argues that most individuals belong to exoteric circles of the collectives and only few occupy a position in the esoteric circle, which perhaps corresponds to the level of expertise one is perceived to have by the members of the collective.

The following sections will present the analysis of data from interviews, social media publications, team mailing list and grant proposals gathered over a period of two years, focusing on the formation and evolution of expertise in citizen science practice.

In order to place the Core Team's expertise in the context of the expertise of the entire GZ team including the Science and the Development Teams, Table 8.1 below presents the expertise exhibited in the practices that were compiled from the team roles presented in chapter 5, the epistemic practices in chapter 6 and the objectual practices of the Science Team in chapter 7.

Stakeholder	Practice	CS/Sci	Responsible individual/team	Expertise
Science	Identifying Problem/Idea	Sci	ST	D(A), D(CS)
	Primary Task (Decision Tree)	CS	ST	D(A), D(CS)
	Publishing Papers	Sci/CS	ST/PI	D(A)
	Science Communication/Outreach	CS	PI/ST	D(A), D(CS), RE
	Data Release	CS/Sci	ST/PI	D(A), D(CS)
	Conferences	Sci/CS	ST	D(A)
	Comparing against expert annotations	CS	PI	D(A), D(CS)
Development	Assess feasibility	CS	TL	D(T), D(CS)
	Hosting and developing	CS	TL	D(T)
	Launch	CS	TL	D(T)
	Maintenance	CS	TL	D(T)
	Scalability	CS	TL	D(T), D(CS)

	Ethics	CS	TL/CS	RE, D(CS)
	Submit grant applications	CS/Sci	CSA	D(A), D(CS)
	Prepare reports and advisory meetings	CS/Sci	PI/CSA	D(A), D(CS)
GZ Volunteers	Attract traffic	CS	PI/ST	D(CS), AE
	Maintain traffic - science communication	CS	PI/ST	
	Motivation and Behavioural Research	ET	Outsourced	
Media	Promote project launch and results	CS/Sci	PI	D(A), D(CS)
Collaborators	Machine Learning		Outsourced	AE, IE
	Motivation	CS		AE
Scientific Community	Open Data Policy, Open Science, Astroinformatics, Big Data		ST, SL, TL	D(A), D(CS),AE

Table 8.1: Practices and Expertise in Running Galaxy Zoo

Key: CS: Citizen Science; Sci: Science; PI: Principal Investigator; TL: Technical Lead; ST: Science Team: CSA: Citizen Science Alliance; D(A): Domain Expertise in Astronomy; D(CS) Domain Expertise in Citizen Science; RE: Relational Expertise; D(T) Domain Expertise in Technical Development; AE: Adaptive Expertise; IE: Interactional Expertise

Table 8.1 reveals the collective portfolio of expertise that was evident in running GZ as a successful citizen science project, and provides a summarised presentation illustrating how the roles of the Core Team was differentiated from the rest of the GZ teams, and what kinds of expertise was at play in the established practices of GZ. The following section focuses on the expertise profiles of the two Core Team members, the PI and the TL.

8.3 The Expertise Profiles of the Galaxy Zoo Core Team

With the experience in building multiple citizen science projects based on their initial model of GZ, what began emerging was the role of non-domain expertise which allowed professional scientists to extend their expertise to the development of citizen science. Initially the non-domain expertise, namely interactional, relational and adaptive, allowed for the expertise extension of one domain specific expertise (scientific) into another (citizen science), thereby changing practices of professional scientists.

This shift appeared to allow for a dialectical and recursive exchange between the two domains of expertise (science and citizen science), which will be examined through analysis of decisions made by the Core Team in steering the development of GZ. In undertaking the analysis, the personal motives of the Core Team members and their visions for citizen science were examined, as they influenced the direction of practices that define GZ's work. It is through the engagement with citizen science practices that professional scientists were presented with circumstances that enabled them to recognise and refine their domain expertise in building citizen science projects, a reflexive practice that is critical to repeated successful performance. It is hypothesised that inter-domain expertise (relational, adaptive and interactional) are aspects of these closely connected kinds of expertise that facilitate a particular thought style over time; constantly shaping the design of the epistemic architecture through acting on new practices, refining previous ones, and constructing innovative solutions in the technological and communications plane.

The purpose of presenting the profiles and unfolding the emergent design principles of the epistemic architecture within the Core Team is to characterise a particular thought style that is specific to GZ. Initially, the individual participants had their own visions and purposes for GZ; but their interactions influenced each other and other colleagues as the project evolved and decisions had to be made to meet unprecedented challenges. By the end of the data gathering period, the participants appeared to have developed design principles that governed and guided the practices they engaged in to further citizen science, and developed their own expertise in the domain of delivering on citizen science projects. The outcome did, I suggest, characterise a thought style that was shaped by their interactions with not only each other, but with the internal teams focusing on science, education and development; various stakeholders, such as media partners, and funding agencies; and experts from within and outside their domain, such as machine learning experts, programmers and developers. Fleck (1935) proposed that interactions with various thought styles and thought collectives shape and inform one other. With that in mind, it is essential to examine how the GZ Core Team might have been influenced, as their model of citizen science was under continuous development.

I use the term 'areas of attention' to signify how the challenges for GZ citizen science were perceived and prioritised by the Core Team. It was necessary to do so, as in the initial stages the GZ team lacked any structure of management and instead responded to the situations and challenges as they arose, mainly through improvised responses. Over time, the GZ scientists began to recognise patterns of challenges and problems, which led to a more structured manner of engaging with them. They began to take a long-term view as the response to citizen science grew and GZ as a stand-alone project and a thought collective gained stability with its own established practices and growing expertise among the Core Team. The term 'areas of attention' refers to the categories of challenges that were perceived by the Core Team, gaining the status of epistemic objects, the pursuit of which shaped GZ's thought style in relation to citizen science.

The next section outlines the functions within an 'area of attention', and then discusses the expertise that the PI and TL exercised in those functions. Relational and adaptive expertise will be identified as the primary types of expertise that mediated different kinds of domain expertise. The presentation entails teasing out what relational expertise may appear in different combinations within the different functions performed by various members. Ultimately, GZ's citizen science thought style encompasses creating conditions for pursuing good science. That was the motive that gave direction to the work of the Core Team of GZ.

8.4 Areas of Attention: Epistemic Objects for the Core Team

In this section, the principal areas of attention are identified, and were the epistemic objects that the Core Team spent significant amounts of time discussing and negotiating among one another, in order to understand the practices those areas required. Hence, the 'areas of attention' denotes the elements related to the epistemic architecture that required the attention of the Core Team. They were as follows: design, the Science Team, development, volunteers, media, and funding agencies. The different kinds of expertise that are employed in attending to these functions or areas of attention will be discussed, beginning with a brief unpacking of the terms used.

By design, I mean the deliberate positioning, structuring, and organising of all the different elements of GZ's epistemic architecture. This is the most complex of all the other functions as it involves considering and incorporating all the other areas of attention, functions and science. It also involves anticipating and identifying the expertise and the skills needed by the volunteers; the expertise and the science questions from the Science Team, the skills and the time and resources of the Development Team, and other constraints from the external and internal environment. Design is, therefore, not simply about what is visible on the website. The invisible functions and processes behind the interface are structured to produce the kind of data that scientists need, and hence are extremely important. I therefore argue that the design decisions of GZ are most critical, and have led to its success.

The Science Team has been introduced in the previous chapters and their practices have been outlined. It was primarily the Science Team who made decisions regarding the science questions for GZ and designed the decision tree that outlined the logical path of of activities to be completed for the volunteers. Not only did the Science Team provide the primary task, but it was their responsibility to produce scientific results and publish articles and scientific journals from the data obtained from the volunteers as annotations. Therefore, managing the Science Team was critical, as it occupied an indispensable role in designing how the volunteers interacted with the task and produced the scientific knowledge.

The term Development concerns the technological elements and mostly the programming-based practices of GZ. The Development Team was responsible for coding, programming and running the online website and database for GZ. It was their capabilities that allowed for and imposed constraints on what design could be translated into the web-site and how the data could be stored, altered, and retrieved by the Science Team. The development function, which was outsourced at its inception, was moved in-house for GZ2 and a chief programmer was hired as the projects expanded with further education funding being secured. The ways in which the Development Team functioned and defined its vision was critical to how the projects were designed in terms of its scope and scale of volunteer involvement.

The volunteers in GZ, who had been identified as epistemic subjects in creating new scientific knowledge were the group that sustained citizen science. Over the last few years they shaped not only the design of the experiment, but also the possibilities of what GZ's citizen science could be. In particular, they played an important part in shaping GZ's thought style over the years. Volunteers in GZ were not passive users of the online interface; rather, they had in many ways inspired the design and development in helping the team imagine a new way of doing science.

The external stakeholders, mainly media and the funding agencies, were critical for the development of GZ. Media had an important role to play in providing a communication and public relations channel that connected the project to the public. The funding models of GZ demonstrated innovative ways of securing grants that were not explicitly for scientific research, but were meant for education and outreach. It is with those grants that they have been able to build more projects and support the Development Team of the Zooniverse. The grant proposals that were written by the GZ members not only influenced the design decisions of the project, but also the reporting requirements of the funding agencies demanded the team members to reflect on the kind of citizen science they were pursuing.

Table 8.2 is a summary of the roles of the two primary participants presented in relation to these broad categories of epistemic objects. The second column from the left (A), outlines the defining roles of the participant; column B highlights that type of expertise that provides legitimacy for the corresponding roles presented in column A; while column C presents the additional types of expertise that are at play in these roles; column D places those roles into three distinctive categories that form the core functions of GZ and finally, column E on the extreme left links the roles (column A) within the knowledge production system of GZ. Each of the roles of the participants will be presented and examined to reveal the types of expertise at play in achieving the primary goal of the project, the production of scientific knowledge.

Elements of Epistemic Architecture (E)	PI(A)	Legitimising Expertise (B)	Interplaying Expertise (C)	Role Category (D)
Epistemic Architecture	Managing Design	D(A)	AE + IE	Project Management
	Managing Science Team	D(A)	RE	Project Management
	Managing Media	D(A)	RE	Science Communication
Epistemic Subject	Managing Volunteers	D(A)	RE + AE	Science Communication
	Managing Funds (CSA)	D(A)		Project Management
	TL (A)			
Technical Architecture	Managing Development (Primary Role)	D(T)	D(A) + AE	Development
Epistemic Architecture	Managing Design	D(T)	AE + D(A)	Development
	Managing Media	D(T)	D(A) + RE	Science Communication
	Managing Funds (CSA)	D(T)	AE	Project Management
	Managing Science Team	D(T)	D(A)	Development

Table 8.2: Roles and Expertise of the Core Galaxy Zoo Team

Key: D(A): Domain Expertise in Astronomy; AE: Adaptive Expertise; IE: Interactional Expertise; RE: Relational Expertise; and D(T): Domain Expertise in Technical Development

8.5 Background Profiles of the Core Team

By the time the expertise profiles were analysed and compiled, the Core Team were recognised as experts in citizen science and had demonstrated it through successful launches of several projects and publication of papers in scientific journals. However, to understand the process behind the development of the expertise of the Core Team, backgrounds of the principal participants prior to the inception of GZ and at the initial stages of their involvement with the project will be presented.

8.5.1 Co-Founder and Principal Investigator: A background

Although the inception of GZ can be attributed to a serendipitous collision of ideas, the professional trajectory and interests of the PI reveal earlier innovative practices in science communication that would shape his ability to make connections, to provide solutions to problems that may not necessarily be within the domain of expertise in his field of research. It was not surprising that he directed his attention to a public experiment, since the idea of engaging the public with science had been brewing for a while:

8 February 2010

"So we realised we need to get through the rest [of the galaxies] and I knew about a project called Stardust@Home. I have been thinking for ages about [...] for years about finding a good way to get more amateurs involved in Modern Astronomy." - PI

The ability to find intersections between various ideas that arrive at a solution to solve a seemingly unrelated problem is a mark of the kind of creativity that is attributable to adaptive expertise. However, it was not just adaptive expertise at play in the idea of the inception of GZ. The drive to engage amateurs in Astronomy had historical foundations in the domain-expertise of Astronomy. But what was added was the ability to understand the nature of the problem when faced with the challenge of classifying thousands of galaxies. The relational dimension was new, and called for new expertise, in particular understanding the standpoint and what matters for the other, in this case the amateurs.

Being resourceful, which is one of the characteristics of adaptive expertise drove the inception of the first GZ website. The PI's recognition of the different kinds of domain-expertise needed to launch GZ was followed by drawing upon the professional networks of expertise to acquire physical resources and intangible services, which included website design, database programming and website hosting. These domains were outside the expertise portfolio of the PI and were not standard for scientists holding a post-doctoral position in an academic research institution such as Oxford University.

8 February 2010

"So we ended up with me calling in favours. So I rang a guy called Bob, who is a software developer in Northern Ireland, a web developer. I knew him because he had done a website for the book I had written, for a popular book I had written. I called him and said do you want to do something for free? He said sure. Lewis [PI's colleague] called a friend of his called Drew who I think Lewis had done his undergrad or masters with. And said can you do something for free. He said yes. So filled it with design and down with the database programming. Then we needed somewhere to then you have to host this thing on the web. Computing here is interesting. Let's say computing in this department strongly favours security over utility. So, Rob I think suggested let's contact Sal from John Hopkins who was the person who led the department. Rob suggested I should email him. But Sal was the person who build Sloan or at least the database for Sloan. And he is one of the most brilliant people I have ever met. So Sal got involved." - PI

Providing the avenue for public participation in scientific inquiry in terms of a technological solution was only one of the challenges, another was the uncertainty of the response by the public. The measure taken against the risk of low participation was to design a task that would be independent of domain-expertise.

8 February 2010

"I guess there was a belief that the public could do this task. Then the task was quite easy. We choose the simplest, possible, useful task." - PI

The task design that balanced simplicity with usefulness for the GZ scientists did prove to be popular within a very short time, and the PI believed that he made an instinctive judgement, which is indicative of the inclination towards innovative practices over a conventional academic role.

13 February 2010

"Then I opened my Galaxy Zoo email and we had tens and thousands of emails. And then we looked on the BBC site and we were at the top of the most popular articles. So it was clear that we had created a bit of a monster. The turning point was that week. That opened up all these other possibilities. It was almost as if certainly for me it was a snap decision." - PI

The decision to choose a career path which positioned the PI in the intersections of science and the public, to support research outside of his own domain-expertise as an Astrochemist, translated into a rather unique position of becoming an expert in building and launching citizen science projects that deliver measurable scientific results. The 'snap decision' made by the PI to take an uncertain professional path reflected a propensity towards a risk-taking behaviour that is characteristic of adaptive expertise. Positioning himself in the middle demanded relational expertise, to be able to understand both the amateurs and the scientists, which the PI had demonstrated throughout his training as a scientist.

13 February 2010

"This is July 2007. So I had money to take me to Sept 2008, so I should have been applying for jobs. And I had not done very much well when I was in Oxford for first nine months. I got distracted by various things. I was supposed to be working very hard on Astrochemistry to save my academic career. And I remember very clearly when you are In Oxford – you have to decide whether you are going to run with this or you are going to just let it to fizzle it out. So I decided that I will work full time and see what happens." - PI

As I have already shown, the GZ Science and Core Teams were meticulous about meeting the standards of scientific rigour from the very beginning. This remained the driving principle for GZ and the Zooniverse projects that followed, and ensured that the team-building process was aligned with the objective of making citizen science a standard method in conducting scientific inquiry.

13 February 2010

"Why can't we have a good team? And we were very careful to meet those standards. Rather than just to do something simple and take an average. I am doing the data release paper. The amount of thought that is coming into is exactly what data we present to be, and how it gets used is huge because if you are going to do serious science with this then you have to do that. I think the reaction changed later when we began to get more science results out." - PI

Throughout his training as a professional scientist, he had positioned himself in a space where relational and adaptive expertise was required in addition to the domain expertise of a professional Astronomer. In fact, the first GZ project had almost nothing to do with his own area of research. It was his drive to solve problems with an innovative approach that led to his engaging with amateur Astronomers. He did this by building a rigorous scientific experiment online, that would later be accepted as a reliable method of reducing data.

8.5.2 The Technical Lead: Expertise Profile and Background

The TL was brought into the GZ after the initial success of GZ 1, when he was referred to the vacant position by one of the lead scientists in the GZ Science Team. His PhD was also in the field of Astrochemistry. His interest in programming began during his PhD after which he decided not to pursue academia, and worked for a new media agency for a brief period before joining "a hybrid academic environment" at the Sanger Institute known for its Genome sequencing research, where he worked in a collaborative environment with researchers. He recalled his initial motives for joining the initiative:

17 March, 2010

"But it became clear that the project was much bigger than Galaxy Zoo. It was really something, really interesting to get into. So I was only really interested because I thought that it was not just Galaxy Zoo. I did not just wanted to run a website. It is a much grander thing in my head.

It pushes my buttons in the, I think that scientists do not write very good software. Scientists typically are not very good programmers and I have seen that in lots of different fields.

There is a whole expertise in programming, and the whole kind, the whole field of software development is a massive massive area, that scientists never get exposed to that, they never get interested in." - TL

His enthusiasm for the possibility of developing his expertise though GZ was one of the factors that drew him to the project.

18 March 2010

"I am really interested in building expertise in building stuffs." - TL

He recollected his initial ambition as a challenge which he describes below:

18 March 2010

"So now I have got the sort of challenge of like instilling what I think is the best approach into people who are actually just like me at the end of my PhD. We have got two new guys who have started, literally finished their PhDs, and they are kind of interested in programming in the web. That is exactly where I was, but that was me four years ago. So can I get them interested in what I am interested in. That is going to be cool. That is going to be very interesting. Because their skills sets are very similar to where I was." - TL

One of the first challenges the TL set for himself was to initiate and recruit the next generation into GZ's thought collective, who would contribute to his thought style. It should be noted that there was no prior Development Team or software development expert in the GZ team before the TL's arrival, therefore, it was his responsibility to define and give direction to the development in GZ.

Although the TL did not have the extensive experience in science communication that PI had, he was not a novice at using social media in communicating his domain expertise, in this case technical expertise, to his existing network of expertise, mostly through Twitter and blogging. Two of his earliest tweets presented below illustrate an open quality in a very public form of communication to solve the new, sometimes trivial, problems that he faced within an academic environment:

1 January 2009

"Looking forward to getting started on Galaxy Zoo."

16 January 2009
"Does anyone know if there is another way to pay for #aws
services other than by credit card (fussy university department)" - TL

Soon after the launch of GZ2, the TL shared his experience through his first blog post, where he gave the reasons for making some significant technical changes in GZ's operations. The blog demonstrated his commitment and motivation to encourage scientists to develop their programming skills, and develop his expertise in building projects.

19 March 2009

"It's been nearly a month now since the launch of Galaxy Zoo 2 and with close to 15 million classifications the volume of traffic has exceeded even our wildest expectations. I joined the Galaxy Zoo team in January this year and in the six weeks before launch worked pretty much non-stop to re-implement the Galaxy Zoo 2 beta site in Rails as well as write a web service to capture back the results. Now I know it's almost always best to avoid the big rewrite but we had many good reasons for moving away from the old infrastructure and codebase.

The original Galaxy Zoo project was really an accidental success - the team had no idea that what they had created would become so popular so quickly and the story of the melting web server is Zoo folklore these days. With this in mind we were keen for a smooth launch of Zoo 2.

I think one of the most significant moves we made for the launch was to host the new Galaxy Zoo website and API on Amazon Web Services (AWS). AWS has a pay by the hour pricing model which was perfect for our very public launch. Below is a diagram of the production web stack we were running on for launch day."

The communication of ongoing work through social media channels and invited talks and lectures were not mandatory activities, and could be attributed to the principles of openness that later became clearer to the GZ team as one of their core principles. This stage of expertise development can be seen as the beginning stages of a thought collective, where the distinction between the esoteric and exoteric cannot be clearly made. However, the following quote illustrates that the communication function not only served a purpose to promote GZ's efforts to an external audience, but also helped TL internalise GZ's aims. PI had championed the art of communicating with a wide range of audiences, and it was an activity that became part of TL's responsibilities not only as the leader of the Development Team, but also member of the Core GZ Team.

18 March 2010

"And over probably the next three months, I started to actually get really passionate about the idea and really believe that we were doing something different. And so that took months. There was no point where I was like yes, there was a moment […] that was the moment when I started to believe. Partly from brainwashing […] I probably started to give couple of talks, I started talking about it […] and you have to really…when you are going to say a word you have to really think about what you are going to say." - TL

GZ was first made possible by reaching out to experts beyond the field of the founders themselves, and they continued the pattern offline and online, especially through social media. Trained as a researcher in Astrochemistry and after having worked in new media and bioinformatics as a programmer, the TL's trajectory led him back to an academic in a research initiative built on new media platform, not as a domain-based scientist, but as a developer who would work with the Science Team. Therefore, the role was to design and build new technologies for scientific knowledge production. He joined after the initial success of the first version of GZ and therefore in Fleck's (1935) framework, he moved into the newly formed esoteric circle of GZ's thought style, who would later go on to heavily influence the course of GZ's vision citizen science.

8.6 The Design Principles: GZ's Design Thought Style

The design principles for both participants emerged over time as a result of their sustained interaction between the public, the Science Team and their personal visions for the project. It was in the unfolding of the design principles as an epistemic object that the complexity of interplay of all kinds of expertise mentioned earlier in this chapter are illustrated. The different kinds of expertise that the PI and TL brought to the design decisions of GZ will be illustrated through excerpts from their interviews over two years, throughout which they occupied the roles of PI and the TL for GZ, and were board members of the Zooniverse suite of citizen science projects. Although most of the excerpts presented here relate to GZ, there will be references made to other Zooniverse projects as the expertise and design decisions of the projects influenced and shaped each other.

Therefore, to understand the design of GZ, the decisions behind other projects will be discussed when appropriate. The Core Team that manages GZ and Zooniverse are the same, and therefore, in the instances chosen for this purposes, the distinction made even for analytical purposes would be artificial; although attempt has been made to keep the discussions relevant mostly to GZ to maintain a defined scope of this thesis.

Design is perhaps the single most important area of attention for GZ in shaping how the elements of epistemic architecture are positioned, and interact with each other to eventually construct the model of citizen science, jointly envisioned by the PI and the TL. The definition of design I follow is "The arrangement of the features of an artefact, as produced from following a plan or drawing[38]". As issues around design became a recurring and important challenge, the following themes that influenced the design decisions emerged from the data and were broadly categorised into the following principles of doing good science, breaking the mould and communicating science with the volunteers.

Doing good science perhaps includes all the other themes, as it is not only the primary goal, but one that encompassed the ethos of the entire endeavour. It was the means and the end of the initiative. As PI elaborated, breaking the established model was essential to the spirit of the scientific process, constantly questioning and challenging what the ways of knowing are today and what is known. The existing model can be conceptualised as the current epistemic

[38] Source: http://www.oxforddictionaries.com/definition/english/design

architecture, with its design subject to continuous changes and refinement to meet the purposes of the project.

As a part of science education of the public, one aim of GZ was to build a community of volunteers who could understand and contribute to scientific research. One of the most important steps in achieving that was breaking the misperception of science among the volunteers. As the project progressed into successful iterations of the original version, the possibility of the amateur public and professional scientists interacting on the same platform became a vision that the PI wanted to follow. As the GZ scientists began to recognise that the volunteers could complete more complex tasks, and take the initiative of undertaking their own research, it became possible to imagine designing a system whereby the volunteers would have moderated access to GZ scientists as potential collaborators. This may appear to be a rather ambitious goal, given the epistemic authority professional scientists are afforded, but for the PI and TL it was seen as a design goal they could strive for.

8.6.1 Doing Good Science: Making Science the Primary goal

This section presents the different ways in which the Core Team understood what good science meant to them, and attempted to examine how good science related to design in the context of citizen science, and how good design facilitated good science.

On GZ being labelled an outreach project, the PI asserted:

13 February 2010

"It does outreach an education for free, but the purpose is to produce science that we could not do any other way. And we have been very careful trying to make sure we deliver on that."

"[t]he best way to do this is if we are looking at a new project and it does not have a science case so I do not care how good the education is, I do not care what it is going to teach me, I do not care how many pages you want to use. So, it is necessary to and sufficient for it to have a science case."

"Because I have never claimed and said [sic] based on one person, I am usually making claim based on 20 people. And so we have that problem solved by having statistics about how good the classifications are. And I can do a comparison with experts classifications […] so that is quite a sound way of working, it makes our life easier. And so it is this data reduction Citizen Science that I think it is easier to produce stuff that fits into standard academic environment." - PI

The excerpt above clarifies science to be the utmost priority in GZ, therefore the citizen science efforts are determined by whether or not the science case is sound. The design element that is most important to good science is the primary task as TL explained, it also served as a standard by which they measured their own success.

28 May 2010

"Primary Task is the bit that we really worry about for achieving the research goals of the given project. Zooniverse [...] all of the projects, if the primary task does not work, by definition the project is not successful." -TL

The design of the primary task was discussed in chapter 7 as an objectual practice of the Science Team, and therefore will not be examined here. It was one of the most critical practices of the Science Team, and it was the responsibility of the Development Team to ensure that the task ran exactly along the decision tree that was agreed by the Science Tam, and which determined the quality of the results that could be used for further data reduction. Therefore, the design of the primary task required domain-specific expertise in the scientific discipline of GZ.

Elaborating on his understanding of good citizen science, the PI drew a distinction between citizen science that produced good science, and what may appear like citizen science but does not really contribute to the body of scientific knowledge.

4 June 2010

"Goes back to the definition of citizen science. If you call it something like contribution of non-professional in the scientific process, then that is actually quite a high bar. Because you are [saying] it is implicit that it is a meaningful contribution.

It is very easy to design projects that look like citizen science, but would not even produce anything that would be used by scientists, so a good example of this might be getting primary school kids to count the number of species of plant in their school playing field [...] you might do that in a few thousands schools [...] I am making this up. But the odds that there is not going to be anyone writing a scientific paper." - PI

Among the many tasks, the design process in citizen science involved ensuring the details of the process were explicit. The TL gave an example of how the specifics of the image processing decisions were a scientific issue:

8 March 2010

"[y]ou have to be sure that you are doing it in a robust way so that you do not affect the scientific outcome. So all of the discussion, it was about two weeks of discussion, with S going to one of the guys there, you have to demonstrate to me that you know exactly how you made these images [...] so we have to goback and try and understand why the bias, [...] for these galaxies we apply slightly different colour and that appears to have affecting the result [...] Steph was not completely comfortable in the way the images were being made because he made these ones, he prepared these images. So he knows exactly how they were made. Whereas when you send somebody else, you make the images, you cannot [...] how exactly do you make that, why exactly does it look like that? So it is kind of quite an extended discussion." - TL

The above excerpt is also an example of how TL's domain-expertise in Astronomy privileged him in understanding concerns of a Science Team member, which a professional from a purely software development background would not have had. It was precisely this sensibility demonstrating relational expertise which formed the core expertise in for both the PI and TL.

However, the PI and the TL also demonstrated differences in the ways they understood what good science constituted. For instance, in an interview almost a year into the data gathering process and TL's arrival at GZ, a decision regarding a primary task to be repeated had to be made.

13 June 2011

"Kasey has written two papers from GZ 2. But she said, well I think we should do GZ 2 again because I have got some other questions I would like to answer. And I thought that was a terrible idea. We cannot ask everybody to go through the database again just because you thought one extra question you would like to ask. And she was like: Why not, I am going to get good science out of this. Right? Not the point. If we did not think to do it the first time, we cannot just say 'Hey it is kind of like GZ2 but with another question.' No, we are not doing that." - TL

This is an illustration of how TL considered that the volunteers' time and effort related to good citizen science, even if it meant not answering a scientific question. The term proposed for the intersection of good science with good citizen science is 'phronetic design,' which I suggest involves a demonstration of relational expertise on behalf of the scientists responsible. The TL was also a strong advocate of ethical considerations in citizen science and in the interview excerpt below, he elaborated:

13 June 2011

"Does that warrant a year long effort by tens and thousands of people for that one paper? No! It does not. But it was interesting that from Kasey's perspective, she was like, but it is fine, we will just get people... There is a very clear idea that some people in the Science Team have not.

So sometimes we surprise our collaborators, especially the GZ Team, when we are like [...] no no [...] you cannot do that. Why on Earth not? I want to do science.

If I am doing science, I am doing the right thing. But, you are wasting people's time. Because they are not even thinking about ethics of it all, why should they. They have kind of been isolated from the Zooniverse, they have got the biggest project but they are not thinking about the domain as whole.

They are not thinking about the Zooniverse and how you would vet new projects, they are just thinking about Galaxy Zoo science, which is absolutely fine [...] But there is certainly [...] I do not think there are many people in the world who are thinking the thoughts we are thinking."

They are not rules [...] they are founding principles. They are not the founding principles of GZ, they are the founding principles of the Zooniverse." - TL

There is perhaps a relation between having a systemic view of the larger picture as a function of scalability issues and the practice of phronetic design that is facilitated by the presence of relational expertise as the capacity to take the standpoint of the other and work with them on their problems. The TL explained his stance over wasting volunteer time with the argument that was connected to the larger picture and the future of GZ. He also continued to defend the design of the Zooniverse projects publicly (see Appendix G). He saw the problem in relation to the need for an advancement of methods that could match human capabilities of classification and improve the quality of data received from GZ:

13 June 2011

"Actually what we are interested in is people providing classifications where methods are not good enough and then when they provide enough data and the methods should improve and we should stop asking them to do those tasks." – TL

GZ's efforts in science education set it apart from many of its citizen science counterparts. PI explained how he understood what makes citizen science good:

20 December 2011

"Citizen science necessarily at its best includes [...] at least our stuff [...] we try and do science education as well but it's secondary. So you can never do citizen science without science education. I would argue that FoldIt does it [...] But you do not learn anything about protein folding. They do not teach you that by moving two red blobs together, you are bringing hydrogens together and because they are repulsive, then that raises the energy of the molecule. You just learn by playing the game, that if you keep the red things far apart, then your score goes up. And you get meaningful results without science education. I am not saying that is ideal either. But I think that is better than to say somebody [...] for fun, you could do this for fun." - PI

Science education remained a core function of GZ, which shaped GZ's thought style vis-à-vis citizen science. Whether it was due to the PI's long held passion for science communication, or the unprecedented nature of communication demanded by the public at the inception of GZ, science education was an essential element in GZ's design. For this purpose, the team secured a large grant from the US specifically to design and implement the web-based tools know as User Based Research and Education Tools and (UBRETs), that enabled volunteers to navigate and explore the objects they had to classify.

While large data sets provided for Astronomers have transformed Astronomy into a statistical science, as the PI maintained, he also cautioned against making a positive correlation between large surveys and good science. He gave the example of definitional differences between the two databases as a source of unchallenged assumptions that could lead to compromised scientific results. Therefore, design decisions for citizen science would include using the catalogs that had been examined and verified for the needs of GZ.

20 December 2011

"One thing that the surveys do, it is incredibly powerful, makes it easier to do good science, but it also makes it incredibly easy to do bad science. Anyone can download two datasets and compare them [...] a good example of this is [...] like just defining what a galaxy is. Two surveys might have a very different [...] radio or optical might be very different definitions of what a galaxy is, but naively, I can come along and say 'I want to look at radio galaxy, so I will just download the two datasets and where they cross, go alright that's the radio galaxy. But I have implicitly accepted a whole host of assumptions on both sides. And in fact, what I should have done is, there are people who have spent years doing a proper cross match between these catalogs and understand these issues, and I should just use their catalog. But it is very easy to do the wrong thing. But there is always bad science, it is easy to do it now." - PI

In addition to being recognised as contributing scientists in their own research community, the PI explained in an interview how being within academia inspired an experimental approach to building GZ. GZ appeared to be a project that aimed to break the existing structure of the previous version with every new iteration.

4 June 2010

"Because what's different is we are not building website for the LHC...the website is the experiment. One of the joys of working in an academic environment is that I can try and deliberately break the model." -PI

Following the first meeting with the advisory committee for the large National Science Foundation (NSF) grant in the US, the PI elaborated on the tensions that prevailed for an academic project that was required to adhere to project protocols in the non-profit sector. This had been the advice given to GZ at the meeting. The idea of formalising, he explained would hinder the very approach to design that led to GZ's success and would lead to its future successes.

4 May 2010

"More philosophically, because we are an academic project, we are deliberately trying to break the mould with every project. Every project we have chosen to do, has been chosen because it's significantly different in some way. So, Mergers is a whole different level of interaction' Supernovae is live data [....] is also video and....MZ[39] is probably closest to GZ [...] it's an intermediate in difficulty between GZ and Mergers. Papyri is ships, logs, three kinda of interactions rather than just one - tell us the stuff we know is there, what might be there and free form tidying.

[39] Referring in Moon Zoo, another citizen science project built by the Galaxy Zoo Team

It's a deliberate policy to do lots of different things and see what works and what doesn't and I've said to Arfon and I mean it that if all projects succeed, then we've done something wrong. We've stayed too close to... we haven't gained any information about what works and what doesn't and in a few years' time, I want to look at applications and say, that'll work and that won't work because we've done that. We'll still be doing the ones that break the mould and the moment you start documenting...the moment you formalise what it is to be a Zoo, the moment I write down...anything more than a philosophical level[....]so we do have rules right, we have documents that say things like the Zoo must produce science which is defined by one publishable paper." - PI

The PI's understanding was that the ability to innovate and break the mould of citizen science projects, would be weakened by the formalisation of the process of building those projects. The lack of formalisation itself was an element of the design principle, which allowed for adaptive expertise to be exercised. In essence, one of the most defining characteristics of GZ's epistemic architecture was its constant reconfiguration, which could be traced to this design principle of innovating with every iteration, perhaps by virtue of GZ being located within an academic culture. It did appear that his inspiration for a scientific experimental approach to citizen science drew parallels with his personal approach to scientific inquiry.

4 May 2010

"I guess I don't have any faith[...]in very formalised ways of documenting things. This is a very personal style of mine, everyone disagrees. Once you write down the rules, instead of thinking about whether something is a good idea or not, you do and look and see whether it's in the rule book. I want us to do projects because they are good novel ideas, not because they agree with the seven fundamental rules of[....]" - PI

However, as the project progressed the need to define a loose structure arose. This change was particularly needed after some of its operations were based at the Adler planetarium in Chicago.

The PI's own vision for citizen science for good science meant that he created his own demands. Each previous success appeared to be the starting point for identifying and addressing the next demands.

21 December 2011

"From my point of view, we are always trying to do the next hard thing. And so, there's some expertise there, there's a track record there that we will outperform anyone else" - PI

Therefore, seeking design challenges including ensuring good science required the specific expertise that the GZ Core Team repeatedly exhibited. Indeed, as their citizen science efforts expanded, its success continued to be measured by the scientific output, as well as recognition by their peers within the community of citizen science project designers.

The expertise in designing good citizen science projects for science included the adaptive and relational expertise exhibited by the PI and TL. It was exemplified by the practice of translation (as discussed in chapter 6), allowing the public to participate in scientific knowledge production. However, in order to ensure good science, domain expertise in the scientific field of astrophysics appeared to be the most critical kind of expertise required, serving as a foundation for adaptive expertise to be practised. Good science as a design principle was a function of domain expertise, which was integral to an understanding of citizen science as an epistemic object for the PI and the TL.

8.6.2 New Methods for Scientific Inquiry

Astronomy as a field has been at the forefront of facing the data deluge discussed in chapters 1 and 3. This section presents the implications for expertise in advancing new methods of scientific inquiry, that is, citizen science as an innovative method. As a design principle, innovation in methods required the foundation of domain expertise in Astrophysics. However, adaptive expertise appeared the most prevalent and was exercised and refined through the pursuit of creating new methods for scientific enquiry. Drawing on interviews with the PI and the TL, this section reveals their commitment to developing and refining the existing methods of scientific enquiry through citizen science.

The PI summarised the current trend in Astronomy of moving towards becoming a statistical science, while still engaging in the more traditional form of observation.

4 May 2010

"Astronomy has become a statistical science. We still do both. Whereas, people before did a detailed study of a small number of objects. It's now normal that a paper includes a million. It's a shift to survey science. It's very hard to say it's a hard and fast line. Lots of people do both. When you find an interesting object, you switch to the other mode. And there is no, 'it's 30 galaxy a survey'. There's a shift. And that has interesting implications. It's almost as if we were professionalising data collection.

In some sense it's inevitable because of the size of the datasets. If you have someone who has never encountered raw data, then it's a problem...I think that respect for data is quite healthy." - PI

The initial undertaking of applying the techniques from one field to another is evidence of the practice of adaptive expertise, which defined how GZ had been operating since its inception, and expanding into several projects, some even outside the discipline of Astronomy and physical sciences.

8 February 2010

"I wanted to use the tools learnt from Astrochemistry to try investigate that [....]. So we realised we needed to get through the rest and I knew about a project called Stadurst @Home. I have been thinking for ages about finding a good way to get more amateurs involved in Modern Astronomy. And so that they came together [...]. so just talking to Melvin and we came up with the idea and that it sort of was born at that point." - PI

Anticipating the use of the new astronomical survey, the PI reflected on the developments in the field whereby most scientists would be focusing on statistics-based astronomy research, and wondered if there were any qualified groups to carry out the observation-based research.

20 December 2011

"We were talking last night to various people, the next surveys [...] LSST and Co are going to produce so many interesting objects. I don't think there are enough astronomers to follow up on serendipitous discoveries. And so I think that...one response is that 'ok we're doing LSST because we care about the bulk statistics about the galaxies, which is perfectly fair and that science will get done' [...] So that leads you to do the idea that you do need the high levels of citizen scientists involved. But I don't think we're quite there yet." - PI

One of the unprecedented developments in GZ was the application of machine learning in order to improve galaxy classifications and the continued collaboration with the field of machine learning. This development was unanticipated by the Core Team, as it was an area very different from their own domain-expertise of Astrophysics, belonging instead to the field of Computer Science and Artificial Intelligence. Attempts to understand and integrate machine learning in GZ's knowledge production system was a demonstration of adaptive expertise, and interactional expertise, since the team did not have the domain expertise in machine learning. However, they had the ability to communicate with machine learning experts and understand their needs, and consider a solution that could be beneficial to both domains, allowing machine learning to occupy a critical role in GZ's citizen science.

The PI outlined the role of machine learning in tackling not only the increasing quantity of data influx, but also emphasising the qualitative difference it can contribute to dealing with large datasets. GZ classification by the volunteers had been used to improve machine learning classification algorithms, and machine learning came to form an integral part of GZ's principle of phronetic design.

30 June 2010

"It will be an interest to machine learning people [...]The general trend in citizen science projects we will see [....]The data sets get bigger and bigger, computers get better and better, the proportion of data you have to send to humans get smaller and smaller. But because the datasets get bigger, if we get it right, the absolute no. of galaxies that you need to show humans probably stays about the same. But the key to that is to get the machines to tell you when they need human help. Which is doable[...]but no one's done it yet." - PI

In suggesting how to solve a problem of machine learning, the PI resorted to his problem solving method, of getting the attention of established experts. This was already demonstrated during the first launch of GZ, where Sal was approached to fix their database. It is a mark of his adaptive expertise, to be able to draw on his existing network to solve a problem in which one has no domain knowledge.

30 June 2010

"This is a machine learning problem[...]we need to put a lot of effort. We can get a lot of publicity in the academic community for it, then we might get some high-powered people thinking about the problem. That's what we need." - PI

The TL later made the following observation about the machine learning grant put forward by one of the Science Team members:

28 May 2010

"Actually, it's not clear how we will [...] Machine learning, I rarely think about. It's quite an exploratory grant. I think a few wins would be successful. We should see this honestly. I don't know what would be successful for this grant." - TL

The TL similarly exhibited adaptive expertise when he became involved in a collaboration with artificial intelligence (AI) researcher groups at Oxford in order to test classifier expertise, as he elaborates:

13 June 2011

"[b]uilding some of the intelligence in some of the methods that they use for improving classification of data into the Zoo. So, it's the lab stuff. So I think, quite possibly, I hope this is the case, the introduction of the Zooniverse labs is essentially where innovation is going to happen."- TL

Over time, GZ's interest in machine learning and its potential increased despite the lack of domain expertise in the area. The GZ Science Team published two papers from a machine learning perspective, and the use of machine learning techniques in GZ papers began to be referenced more frequently after the first paper was published in 2010[40]. The publication of an entire chapter by the GZ Science Team, the first of its kind in the field, in a book titled 'Advances in machine learning and data mining for astronomy' (Way *et al*., 2012[41]), was evidence that GZ was part of the movement in encouraging collaboration between the domain of machine learning and astronomy through citizen science. Among the nine authors of the chapters on GZ, only two were machine learning experts, with one previously a high-energy physicist, one an astronomy

[40] Papers on machine learning include,
a) Banerji, M., Lahav, O., Lintott, C. J., Abdalla, F. B., Schawinski, K., Bamford, S. P., and Vandenberg, J. (2010). Galaxy Zoo: reproducing galaxy morphologies via machine learning. Monthly Notices of the Royal Astronomical Society, 406(1), 342-353.

b) Fortson, L., Masters, K., Nichol, R., Borne, K., Edmondson, E., Lintott, C., and Wallin, J. (2011). Galaxy Zoo: morphological classification and citizen science. arXiv preprint arXiv:1104.5513.

c) Smith, A. M., Lynn, S., Sullivan, M., Lintott, C. J., Nugent, P. E., Botyanszki, J., and Walters, R. (2011). Galaxy zoo supernovae. Monthly Notices of the Royal Astronomical Society, 412(2), 1309-1319.

[41] Way, M. J., Scargle, J. D., Ali, K. M., and Srivastava, A. N. (Eds.). (2012). Advances in machine learning and data mining for astronomy. CRC Press.

education expert along with the GZ Science Team members trained in Astrochemistry and Astrophysics. GZ was evidently forming new epistemic communities and networks directed towards common goals and shaping a particular thought style of citizen science, (Fleck, 1935).

In the excerpt below, the TL elaborated in an interview on this understanding of the possible ways in which machine learning can help make design decisions in GZ:

13 June 2011

"I think what we are going to learn there...by not having to sweat the details about that doesn't look very good, can we ask people to...tell people who clever it is 'coz the people who get involved in the labs projects will be interested and would have understood the very very special and unique abilities of people's vs machines. And we can say this is the space we are helping to further the actual research discipline of citizen science or man and machine or whatever. So I think innovation is going to happen in that space. The examples of that are where you are testing the ability of the classifier as you go along.

I disagree with sometimes on what constitutes people's waste of time. I would say that, I don't think we should do any more supernovae stuff, I don't think we should take more candidates for the project, because...I think that it really is a solvable problem and I think it is our duty to innovate. We're in the space where, we've got large datasets and large numbers of individual classifiers, there are understood ways of how to improve algorithms by using large good quality training sets. Or if we are to do more Supernovae science, I would like to give each project a year. OK 12 months and then you'll be done. Like we're gonna close after 12 months. Unless [....] unless we see the efforts of the community fed back into improved algorithms and then you start sending us slightly different candidates."

[....] you shouldn't just be building these projects because you can't be bothered to put in the effort into the algorithms." -TL

The TL conveyed a very strong sense of responsibility towards the volunteers, which he translated into their "duty to innovate", which I would argue is evidence of relational expertise that called upon adaptive expertise to be exercised, building on the foundations of technical expertise. Therefore, the design principle of 'breaking the mould' or innovative practice calls for the interplay of different kinds of expertise that are also tied with ethical practice (phronetic design) in citizen science.

One of the ways in which GZ approached citizen science resonated with their understanding of the scientific spirit of trial and error, which extended to their approach in experimenting with citizen science activities that did not necessarily yield a scientific outcome. Known as the Labs, they designed a space where short-term, small-scale experiments could run for a week or few months, as the TL elaborated in an interview below:

13 June 2011

"[t]his is not our poster child....this is just a small scale project, it might be experimental, the results might not be guaranteed to produce science, but it's a part of an ongoing experiment. It's kind of labs stuff. So that's easy, you can say it's labs, Zooniverse Labs. And at the door when you step into the labs, there is not guarantee that you're doing things of scientific value anymore because it really might be trial and error. And we might go those results were junk, that was a waste of time. But that's ok, that's the buy in." - TL

8.6.3 'Equal Footing' Between Scientists and the Public

The design principle, creating a space for interaction between science and the public, called for relational expertise that built on the domain expertise of the GZ scientists and the interactional expertise of the public to intersect. From the moment the GZ team was flooded with emails from volunteers after the first launch, communicating science to the public was an ongoing effort. Communicating with the volunteers was an established objectual practice in the

GZ team as discussed in chapter 7. Here, the objectives of the Core Team to develop GZ's relationship with its volunteers are presented, and how those aims were translated into design decisions are explored.

In the initial phases of GZ, the volunteers were perceived as one homogenous group of individuals who would perform a few classifications. The GZ scientists' perception of volunteers as it evolved was discussed in chapter 6, exploring ways in which they participated in the stages of translation proposed by Callon et al. (2009). The present chapter specifically addresses the design issues that the GZ Core Team had to consider in keeping the volunteers as an audience. GZ has evolved from being a simple classification project, to one that provides platforms for amateur research, and engaged science communication.

In their attempts to understand the volunteers, the members of GZ Education Team carried out research to interpret their motivations for participating in GZ. The findings then influenced their approach to communicating science with their volunteers. As the excerpt below suggests, the willingness of the volunteers to contribute to research instilled a sense of volunteer responsibility in the project, and influenced the design of GZ to emphasise a pedagogic purpose in relation to widening access to science:

20 December 2011

"You see lots of resources to get kids involved in science, but it doesn't go anywhere. I think we have to be very careful because we know that our volunteers, the large part of why they do it because they want to contribute to science. A large part of what we're trying to teach is that science is open for contributions from everybody and the attitude of science I would like to change is that its reserved for scientists." - PI

The Education Team's research also revealed that within the volunteer groups, there were sub-groups with different levels of expertise and commitment, which was a factor that later influenced how features of GZ were designed. What determined the design that allowed for communication and interaction between the volunteers and the scientists was how the designers, primarily the PI and TL, perceived the volunteers, the distinctions they made between different kinds of volunteers, how they related to them, the ambitions they had for developing the relationships, and the constraints that resources and time imposed on their efforts.

20 December 2011

"But we know that people are having higher level...so we are trying to build the communications[...]there are various discussions going on. So we are trying to rebuild the forum so that it's easier for people to get involved in discussions.

On the Interface: I think what we're doing is the stuff that is important to us. The sites have to be slick, they have to work, we have to understand motivation, and I think that's beginning to become clear to other people." - PI

The PI was explicit in his objectives of eroding misperceptions of science in the public, and he explained his vision of creating a space for 'equal footing' between scientists and the public. The efforts of GZ towards that goal through the process of iterative design were apparent in the possibilities they were exploring to accomplish that goal.

12 February 2010

"Even in the volunteer community, it is interesting in the volunteer community there are experts and none-experts. And the question whether the community structures itself to recognise those or whether we need to encourage it to do so." -PI

30 June 2010

“One way of making this easier is, rather than drop them in with the scientists, [is to] find a way to allow them to work with their peers so that they reach a point where it is useful for them and the scientist to talk. It's a hard problem.”
- PI

In the following excerpt, the PI conceptualised the design of a system that would allow for mechanisms to communicate matters of scientific importance to reach the scientists. In particular, he considered the issue of scalability, which was part of the vocabulary from the TL’s domain expertise, suggesting the joint development of a thought style between the PI and the TL, with evidence of merging expertise, in line with Fleck’s discussions of thought collectives (Fleck, 1935). Over the period of the data collection, the talk of ‘design’ was observed to become more frequent.

20 December 2011

“And the minute we have a human system, where we rely on the moderators to do that. I'm not sure that is scalable. A system where good stuff can bubble to the top and be recognised and worked on or with professionals.” - PI

Such a system was required as time was reported to be a scarce resource for the scientists. The PI explained that the scientists were overworked and barely had time to do their own research, therefore the system should be extremely effective if it was to realise his vision of citizen science. He found that, in addition to doing good science, having hundreds of thousands of volunteers participate in the projects for which he is responsible was a motivating factor for him. He explained in an interview as follows:

30 June 2011

“I finally got to give my GZ lecture I'm happy with after about three years of giving them. [.....] It's all about stuff we have discovered in collaboration with people in the audience, which is pretty cool. It is about that widening participation.” - PI

He also made some of his domain-specific values explicit regarding the methods of the discipline. For example, he reflected on the current state of astronomy where astronomers lack the time to observe and study astronomical phenomena. He talked about designing a system for serendipitous discovery. This ambition was a considerable challenge, since the Core Team did not yet quite understand what features and elements of the system supported the two big discoveries the volunteers made.

30 June 2011

“The other thing is this idea of serendipitous discovery. Even in the face of these large data sets, maintaining the ability of looking at this stuff is weird and we should pay attention to it. I think that's really important too.”- PI

The high levels of response and the commitment from volunteers, was something the team had not anticipated. Therefore, to their surprise when events such as a discovery of new objects occurred and volunteers began to undertake their own research, it opened up design challenges for GZ.

30 June 2011

“The Peas[...] the peas were amazing. They have basically written the middle bit when that came to us. By the time it came to us, they had already organised their own search, found lots of candidates, sorted the candidates in categories, looked at the spectra of all of them, classified by spectra, they had literally done sections 2 and 3 of the paper.

There is clicking the buttons, there is arguing on the forum about what we found, but some of them have been already using spectra’s, some of them are writing, not many, data base where reasons are...But we are trying to provide the sort of extra stuff for people so they can take more of the scientific process.” - PI

Within a year of data gathering, the TL's idea of phronetic design appeared in PI 's explanation of not wasting volunteer time as the founding principle for the Zooniverse. Here the thought styles intersect and strengthen in the collective.

20 June 2011

"That's the thought behind Zooniverse. If we can be a stamp of guarantee that we will not waste your time." - PI

The PI believed that GZ's strength lay teaching the process of science as it is currently carried out. He reviewed what kind of science education the project was providing the public in an interview:

8 February 2010

"If we are trying to teach people about process of science. That's a good example of how modern science is done. It's not quite experiment hypotheses test. It's models of fiddling and so on." – PI

Ultimately, the PI had placed himself at the intersections of science and the public in advancing methods of data-intensive science at the same time as promoting science education, and engaging with the public. This engagement was possible through designing a system led by principles that accounted for good science, new technologies and the public.

30 June 2010

"If it is as powerful as we think it is, if we can show that, I might still be sitting here in 20 years' time developing its mode of data reduction. And I was never going to be the person to rewrite the theory of relativity to explain dark matter. I was never going to crack the hard problems of astrochemistry. It's quite interesting to be doing something on a subset of science, those problems that have large visual rich data, that we might actually make a contribution. That's pretty exciting." - PI

8.7 Emerging Ideas on Design

The idea of 'design' talk was initially observed during extensive interviews with the TL, and I began to explore the concept in more depth towards the end of the data collection stage. As the TL explains, it is perhaps the function of scalability and long term vision that the thinking in terms of design became increasingly important to GZ. It was his role to oversee the technical plane of the epistemic architecture, and the concept of design shaped how the epistemic architecture and the related expertise co-evolved.

13 June 2011

“So the design of the system is very deliberate. I would say that as we grow, it's more and more important that we don't lose sight of the design. Because you don't [...] ill-thought design can send you down a path that in two years you can't innovate because of a particular way you've done something. You're tied into a particular technology because you said- let's do it this way because it's easier.” - TL

The TL expressed his understanding of their contribution to science as quite succinct and ‘straight forward.’ He pointed out:

13 June 2011

“We sort of positioned ourselves as the people who can help a Science Team to distill what can be quite a complex classification scheme into something very straightforward.” - TL

Towards the end of the data collection period in June 2011, it was clear that the expertise that the Core Team had developed was applied to projects beyond GZ and its various iterations as the TL recalled:

13 June 2011

“I can never remember what projects we have now. (laughs) It's too many now.” - TL

The background of the TL also influenced GZ’s thought style. It was the kind of thinking he brought into GZ from his software design background that was reflected in how GZ lacked organisational structure. In explaining the lack of rules in his field of expertise, he emphasised the use of the term ‘design’:

13 June 2011

"I don't think that software engineering is a good term. There aren't rules that make good software. There are some rules, there are guiding principles that you can follow and best practice that you can follow, but fundamentally it is a craft. And so that's why design is a better term. I don't mind the term architecture.

Understanding the design and the architecture of the Zooniverse is pretty much what I think mostly about. It's also entirely abstract usually, so you can talk about it without writing any code. It's almost like you get the feeling of the aesthetics of the design even if it's code, that actually that doesn't feel right. And that's a very hard thing to explain to somebody [without] experience [...] no I don't like the way that feels. Why not? They are logical home of things." - TL

It could be suggested that the TL applied his skills of abstract thinking from the domain of software design, skills also identified by Kramer (2007) and Nerland and Jensen (2010), to the management of GZ. But that capacity for abstract conceptualisation in this area also developed as he worked on GZ and the Zooniverse. Towards the end of the data gathering period, it was clear that the TL's responsibility was expanded beyond the technical plane of the epistemic architecture of GZ, extending to thinking about the entire design of Zooniverse suite of projects and management functions. These included media relations, hiring developers, assessing collaborations, and promoting the work in GZ within the software development community.

However, in working closely together towards individual goals and shared collective goals for GZ, the expertise at the Core of GZ was primarily a merger of the expertise portfolios of the PI and the TL. They both extended their relational expertise in order to understand the goals of the other, and imagine situations and problems from their perspective, as they shared an intersection of responsibilities, and sometimes made management decisions on each other's behalf. While their individual strengths and expertise varied in terms of the functions that they were primarily responsible for, it became evident that they were able to recognise and perceive challenges through the standpoint of the other, as illustrated by this interview excerpt on the design and architecture of the Zooniverse:

13 June 2011

"I spend quite a lot of time actually thinking about the architecture of the Zooniverse. For e.g. those events, where we post back data between projects, we've been wanting to do that for ages. And actually I'm really pleased with the design of how we ended up doing it. And that meant that we'd have been doing it for the last year, but I'm sure we'd have done it in a different way [...] whereas the way we're doing it now, I can see it scaling up to lots of projects and it's still working and that is [...] Design and thought went into that. So some people would call it the how we architected that [...] if that's the word [...] how that's architected is important. And actually a lot of [...] and personally the reason why I was brought into the project, I don't think would have described it in those terms, but he would have said something like we want software that we can reuse across lots of projects." - TL

The competencies involved in the expertise allowed for interactions that otherwise wouldn't have occurred, such as of translation that was covered Chapter 6. These interactions not only reinforced the expertise, but also built and shaped the architecture and the architects themselves through a process of iterative learning. The following excerpt from a blog post [42] by the TL is an instance of how the elements of technical architecture were refined through his understanding the patterns of epistemic practices of the Science Team. It is an example of how the practice of the Science Team shaped the technical plane of the epistemic architecture. It demonstrates a design choice led by technical expertise facilitated by the recognition of a situation through a relational effort in

[42] Zooniverse Blog post titled: How the Zooniverse Works: The Domain Model
http://blog.zooniverse.org/2013/06/20/how-the-zooniverse-works-the-domain-model/

understand how the Science Team used the data collected from the volunteer between the completion of translation 3 and prior to translation 4. It is a demonstration of technical and relational expertise at work.

20 June 2013

“Our domain model has also been heavily influenced by the patterns that have emerged working with science teams. In the early years we spent a lot of time abstracting out each step of the User interaction with a Subject into distinct descriptive entities called Annotations. While in theory these were a more ‘complete’ description of what a User did, the science teams rarely used them and almost never in realtime operations. The vast majority of Zooniverse projects to date collect large numbers of Classifications that are write once, read very much later. Realising this has allowed us to worry less about exactly what we’re storing at a given time and focus on storing data structures that are a convenient for the scientists to work with. Overall the Zooniverse domain model has been a big success. When designing for the Zooniverse we really were developing a new system unlike anything else we knew of. It’s terminology is pervasive in the collaboration and makes conversations much more focused and efficient which can only be a good thing.” - TL

8.8 Concluding Remarks

While the previous chapter examined the objectual practices of the Science Team members, this chapter focused on identifying the different kinds of expertise evident in the practices within GZ, and what drove the development of the new expertise in running citizen science initiatives. This chapter has been concerned with citizen science expertise for Science. As the PI reflected on the nature of his own expertise in citizen science,

23 March 2010

“If we are not careful, you just break the boundary between expert and non-experts. How do you distinguish between what’s actual science and what’s Cargo Cult Science[43]? That’s one of the challenges we have got I think.” - PI

[43] The term cargo cult science was first used by physicist Richard Feynman, and refers to negatively characterised research that may have semblance of a scientific character but lacks scientific integrity.

In order to understand how the expertise in managing citizen science initiative developed, the different kinds of expertise involved in GZ were identified, categorised and presented in correspondence to practices that were identified in chapters 5, 6 and 7. From the synthesised profile of GZ's collective expertise, the expertise demonstrated by the Core Team, mainly the PI and the TL, were isolated and examined. A brief professional background was presented for both participants, which provided the summary of the experiences, expertise and personal motivations prior to their involvement in GZ. Finally, to understand what led the development and refinement of the different kinds of expertise they exhibited, the formulation of the design principles of the epistemic architecture were examined. I argue that the Core Team began with the design principle of doing good science in its initial stages, which explains the skeletal epistemic architecture at the inception of GZ. However, the architecture evolved over time with various iterations and with it, the understanding of the design principles that demanded different kinds of expertise. The expertise in citizen science appears to have evolved alongside the understanding of GZ's design principles. Citizen science expertise therefore presents itself as an epistemic object that is continually unfolding through engrossment in the object along the lines suggested by Knorr-Cetina (1999) when observing scientists.

The findings in this chapter have also illustrated the development in expertise in the domain of citizen science as an emerging field. In the pursuit of expertise in citizen science, what is demonstrated here is precisely the skilful incompleteness by the GZ team. The ability to recognise the incompleteness, and embrace the uncertainty is an attribute of the scientific approach.

It was also revealed that in addition to a strong foundation in domain expertise in astronomy and domain expertise in technical development, the role of relational and adaptive expertise was critical in managing GZ, since the work of the PI and TL was mainly characterised by innovation in spaces where domains intersect, pushing the boundaries of their domain. Therefore, while most scientists would mainly exhibit domain expertise in their disciplines, in order to pursue citizen science as a method for scientific research while maintaining a strong commitment to science communication, open science, and innovation in scientific methods, additional types of expertise were called for. The principal participants began with different goals, the PI aiming to produce scientific results, and the TL aspiring to build on his expertise in software development.

These differences resulted in two distinct purposes directed at the same set of practices. The PI excelled in immediate problem-solving and relational expertise while managing all stakeholders. The TL shaped GZ's structure through the adaptive nature of his technical expertise in creating new tools for citizen science, which the PI had not envisioned prior to TL's involvement in GZ. In negotiating their individual visions for GZ and purposes of their role in GZ, the additional layers of relational and adaptive expertise played a critical role in how they continually negotiated their individual understanding of the design principles of GZ, while developing a joint set of expertise in the domain of running citizen science initiatives.

Chapter 9

Discussion

The aim of this chapter is to demonstrate that the research question for this thesis formulated in chapter 1 and have been addressed and to discuss the theoretical and practical implications of this research for science, society and education.

9.1 Answering the Research Question

The questions driving the study were:

How is citizen science shaping the structure, practices and expertise of Galaxy Zoo (GZ) scientists?

1) In what way is citizen science changing scientific practice?
2) What practices can be identified in the work of Galaxy Zoo as a successful citizen science initiative?
3) What kinds of expertise are evident in these citizen science practices and how are they developed?

To proceed with the research enquiry in teasing out the practices and expertise among GZ scientists, in chapter 5, I proposed constructing a conceptual framework of GZ that would enable understanding the structure of the system within which existing practices are located and where new practices emerge. This conceptual framework, what I call the 'epistemic architecture', draws on the theoretical foundations based on the concepts of distributed intelligence (Pea 1993) and distributed cognition (Hutchins 1996; Salomon 1993). It was assumed that not only scientific practices, but practices that were not necessarily scientific in nature would be significant in making citizen science work at the boundaries of science and society. Preliminarily outlining a structure early in the data gathering stage helped locate the practices that were emerging and evolving, and also verified the simultaneous pursuit of multiple knowledge objects by the GZ scientists, revealing the complexities in citizen science practices. It quickly became evident that the design of an epistemic architecture was continually being reconfigured with the GZ's efforts to meet their objectives. The task was made more demanding by the changing external demands of the volunteers and of the funding agencies, and limitations of the tangible and intangible resources at hand. Therefore, the structure and practices of GZ scientists were examined under the framing of an epistemic architecture in order to capture the complexity within GZ's knowledge production system.

9.1.1 In What Way is Citizen Science Changing Scientific Practice?

In order to determine the changes in scientific practice that citizen science demands, the model of translation as a scientific practice proposed by Callon *et al*. (2009) was introduced in chapter 6 as a conceptual model for 'normal science,'[44] and the basis for detecting possible changes in scientific practice in citizen science. Their model has three distinctive stages, t1 - translating the real world into objects or data that is brought into the laboratory, t2 - constituting organisation, analysis and interpretation, and t3 - representing the return of the results back into the real word; finally reconfiguring the macrocosm from one state to another while those stages are repeated. Building on Callon *et al*.'s (2009) conceptual model of scientific practice, two additional stages of translation

[44] * By 'normal science', I simply mean scientific inquiry entirely carried out by professionals without any involvement of non- professional volunteers from the public.

in the work of GZ were identified, as a result of public participation in the process of scientific inquiry. The two additional stages were labelled T3, where the public volunteers annotated the objects during the primary task and T6 which was a stage before the existing t3, whereby GZ scientists transformed the objects and inquiries identified by the volunteers into a scientific study (refer to Table 6.2).

Drawing upon Knorr-Cetina's (1999) notion of epistemic subjects, it was proposed that the additional stages of translation established the GZ volunteers, who are assumed to be non-professional amateurs in the domain of astrophysics, as epistemic subjects by virtue of their contribution to the process of scientific inquiry. Therefore, even though the volunteers might be located outside the boundaries of professional science in the recognition of their abilities, the epistemic architecture extended beyond the boundaries of professional science by assigning them a status of epistemic subject, as one of the constituent distributed intelligence elements.

What remained unchanged for GZ was the focus on the rigour of the existing scientific practices as discussed in chapter 7 and 8. As a design principle, doing good science, was understood to be the utmost priority for the Science and the Core Teams. However, there was evidence found to support the conclusion that scientific practice of data reduction underwent additional stages to account for biases and errors in volunteer classifications. Therefore, in striving for academic rigour, in addition to new translation stages, reliability checks and data reduction techniques qualitatively changed and became essential elements in the practices within the evolving epistemic architecture.

9.1.2 What Practices Can Be Identified In The Work of Galaxy Zoo as a Successful Citizen Science Initiative?

Chapter 7 addressed the collective practices of GZ's Science Team have directly resulted in a scientific outcome, and examined the practices that supported citizen science, as a process. While the concept of epistemic practice in chapter 6 was restricted to scientific practices, the concept of practice in chapter 7 related to the purposeful actions that carry a supporting function in the production of scientific knowledge. Therefore, to make a distinction from epistemic practice in chapter 6, I draw more specifically on Knorr-Cetina's (2001) notion of 'objectual practice' pervasive in knowledge-based work dictated by epistemic objects.

In answering the question, it was also important to examine how objectual practices were negotiated with the aims of pursuing good science, and the GZ members' developing understanding of good citizen science. Drawing primarily on the' internal email exchanges amongst the Science Team, interviews with the PI, and GZ's public social media content, the critical recurring practices identified included the following: negotiating collaborations with non- GZ scientists; designing primary tasks, which determined the outcome of the translation stage T; data reduction for citizen science; data release; publishing papers; representing GZ in the scientific community; creating spaces for epistemic communities; and communicating with the GZ volunteers.

In setting up a new collaboration, the epistemic object in focus and the terms of negotiation of purposes among the Science Team, resulted in the externalisation of the new understanding, which reshaped the practices to follow. The understanding of such externalisations, in the form of written texts through the team mailing list, resulted in building common knowledge through joint negotiation of the purposes of science and citizen science; this common knowledge characterised the thought style (Fleck, 1935) of GZ. As it was found, deciding on a potential collaboration was not only a precondition to T1, but could influence how T2 (image processing and task design) was carried out. Similarly, the negotiations between the objectives of science and citizen science were revealed to be a common occurrence between the GZ Science Team, where the PI was the most prominent figure in ensuring that the decisions taken served the purpose of sustainable citizen science.

While following the science communication practices of the GZ Science Team, what began emerging as an intriguing development of thought style was the evidence of second-generation scientists, who joined after GZ1, entering the discussion and making decisions relating to openness in practices, and increased and innovative efforts in science communication. Science communication went from an ancillary practice for GZ scientists to a significant practice in terms of the frequency, quality, depth and media platforms used. It was suggested that the recurring objectual practices arose in pursuit of multiple epistemic objects, and those practices demanded existing expertise to be exercised towards a joint purpose of science and citizen science.

9.1.3 What Kinds of Expertise are Evident in these Citizen Science Practices and How are they Developed?

The analysis of the data on Core Team members, and particularly the PI and TL, revealed five categories of expertise that were found to be present in varying combinations depending on the practice examined. Chapter 8 outlined the five categories of expertise, which were i) Domain Expertise in Astrophysics D(A), ii) Domain Expertise in Technical Development D(T), iii) Relational Expertise RE, iv) Adaptive Expertise (AE), and v) Interactional Expertise (IE). The rationale for focusing on the Core Team members in this instance, instead of including the Science Team, was that the Core Team were primarily responsible for the citizen science-specific decisions, and the PI and TL were the only two members of the Team who were working full-time on GZ.

It was proposed that expertise in managing citizen science developed through the pursuit of the design principles. These principles were informed by how the PI and TL negotiated the purpose of GZ to meet both the needs of robust scientific practice and sustainable citizen science. The design principle of 'doing good science' was the primary and only design principle at the inception of GZ and it led to a basic epistemic architecture with translation stages T1 to T7, except for T6 (turning public curiosity into scientific inquiry). Over the period of three years since GZ's launch, two design principles had become central to GZ's design: i) breaking the mould of science (i.e. innovative practice in science), and ii) Science communication. In chapter 8, I examined how the understanding and shaping of the design principles unfolded as an epistemic object itself, and provided the direction for expertise in citizen science to develop, recognising that they were pioneers in the field, constantly expanding the knowledge of citizen science as a method in scientific inquiry.

It also began to emerge in the collective expertise of the PI and the TL that, although they shared a certain domain expertise in Astronomy, or Astrochemistry to be precise, their experiences in science communication and software development, and their distinctive personal ambitions of doing good science and building expertise in programming, respectively, had significantly influenced GZ's particular thought style (Fleck,1935) of citizen science. Even though their scientific expertise legitimised their position and authority amongst the community of astronomers, the GZ volunteers, media, and the funding agencies, their personal goals translated into design principles which shaped the epistemic architecture and vice-versa.

The epistemic architecture was designed for translation stages that supported the citizen science model of GZ. Consequently, the role of adaptive expertise was evident in the decision to create technical features to support T6, and it became a critical feature in the epistemic architecture that was unique to GZ. The technical features that supported T6 were built upon the understanding of the volunteers as epistemic subjects capable of aiding serendipitous discoveries, which might have been overlooked by the scientists themselves. Therefore, the design of the epistemic architecture evolved to support T6 as it was recognised as a practice that would be of scientific value to the GZ team. The translation stages of T3 and T6 signified the function of distributed intelligence in the form of those volunteers who had the capacity to identify what may be of scientific importance and interest to the GZ science team. Good science as a design principle, which reflected their primary goal and purpose of highest importance, called for strong domain expertise in Astronomy, which the GZ Science Team demonstrated in terms of publishing articles in scientific journals, and citations by their peers. However, achieving good science through citizen science also required adaptive expertise. The introduction of stages T3 and T6 demonstrated how the epistemic architecture could be determined by decisions to serve the purpose of doing good science. Therefore, good science as a design principle is a function of domain expertise and adaptive expertise in the context of citizen science.

The second design principle that gained importance among the Core Team was made possible mainly after the entry of the TL into the team as their lead developer. His personal goal to elevate the use of programming in science not only inspired a scalable model of GZ, which supported the launch of numerous projects, but also opened avenues for mutually beneficial collaborations with the field of machine learning, which later played a role in shaping the ethical principles of GZ's citizen science. It was argued that although the domain expertise in programming provided the foundations for the design principle to be defined, it was a combination of adaptive, relational and interactional expertise, which resulted in the ability of the GZ team to create and improve innovative methods in citizen science. These methods supported scientific inquiry through reconfiguring the technical architecture of GZ to be adaptive to the needs and goals of the GZ team and its collaborators. In some sense, the epistemic architecture also exhibited traits of expertise of the Core Team, by virtue of its adaptive ability.

The third principle that led the design of the epistemic architecture of GZ gained significance through the influence of the PI's background and commitment to science communication. The exercise of relational expertise by the GZ scientists in recognising the criticality of communicating with the volunteers and perceiving them as collaborators, which resulted in direct communication between the GZ scientists and the volunteers. The third principle led efforts along the communication plane of the epistemic architecture as the team engaged with different online social media platforms, and offline meetings with volunteers with the intention of explaining them to the scientific process as it happened. These developments not only required technical expertise, but also relational and adaptive expertise in understanding and jointly negotiating the needs and intentions of GZ volunteers. Although science communication was recognised as a critical practice in sustaining GZ's continued success in chapter 7, it was evident that a major driving force behind their communication efforts was the PI's goal to provide a space where scientists and the public would be granted 'equal footing' in a dialogue that would challenge misperceptions about science that the public may hold. Hence, the design of the epistemic architecture evolved to support technical features that would both allow GZ volunteers to establish a community to learn together, and also develop skills in scientific literacy and ultimately produce scientific articles that were qualified for peer-reviewed publication.

Therefore, while existing domain-specific expertise in Astronomy and Software Development was indispensable to GZ's success in establishing citizen science as a valid method of data reduction, I argue that the non-domain expertise, or what I would suggest as 'inter-domain' expertise such as adaptive, relational and interactional, were equally critical in building expertise in a new domain of online Big Data citizen science. It was the pursuit of the design principles as unfolding epistemic objects, that created an an understanding of the practices that were required, while demanding certain combinations of expertise that the GZ Core Team gradually were able to recognise and employ.

While the research question for this study can be answered in the context of other citizen science initiatives, the findings so far can also lead to further research possibilities including enquiring into the following areas.

a) In what ways can phronetic design play a role in achieving collective intelligence through a scientific collaboration?
b) How does the organisation of the technical objects in an epistemic architecture shape the practices of scientists?
c) How does the distribution of expertise between the scientists and the volunteers change the way the epistemic architecture of a citizen science project evolves?
d) Do epistemic architectures of citizen science projects vary between different scientific disciplines?
e) What kind of tradeoffs are there in the work of scientists who manage citizen science rather than purely focus on the scientific outcome?
f) How do scientists perceive and understand their role as experts in science in the context of data-intensive science where computational thinking is becoming increasingly important?

In addition to answering the overarching question that guided this study, there were certain observations and findings that began to emerge as characteristics of GZ as a system, which are briefly explored in section 9.2. Section 9.3 presents a brief deliberation on the challenges for citizen science as an emerging mode of scientific inquiry, while section 9.4 explores the implications of this study in relation to the themes of science and society, training for scientists, and science education. The following section 9.5 reviews the shortcomings and strengths of this study leading to a reflection on the research process in section 9.6. Finally, the thesis concludes in section 9.7 with some questions and areas of research for further study.

9.2 On Structural Characteristics of Galaxy Zoo's Epistemic Architecture

In describing the characteristics of GZ as a knowledge producing system I draw on the concepts of collective intelligence and complex adaptive systems, and attempt to draw connections between these concepts in the context of citizen science, and the significance of epistemic practices, such as translation, and expertise; bringing together the main findings of the study that have been presented in chapters 6, 7 and 8. This study was an attempt to understand GZ as an innovative system within science, one which successfully functions as a result of deliberate design and serendipity. In doing so it recognised GZ's backbone structure of continually evolving expertise, technologies and ways in which the public act as epistemic subjects. The idea collective intelligence, was a central theoretical resource which, as I have already indicated, is based on the phenomenon of amplified intelligence as an outcome of humans and technological tools working together, rather than being simply an outcome that is a simple aggregation of their individual efforts combined (Malone *et al*., 2009). The characteristics exhibited by GZ's epistemic architecture and the collective intelligence within it, GZ suggest it to be an example of a complex adaptive system.

“CAS are dynamic systems able to adapt in and evolve with a changing environment. It is important to realise that there is no separation between a system and its environment in the idea that a system always adapts to a changing environment. Rather, the concept to be examined is that of a system closely linked with all other related systems making up an ecosystem. Within such a context, change needs to be seen in terms of co-evolution with all other related systems, rather than as adaptation to a separate and distinct environment.” - Chan (2001:2)

According to Holland (1992), complex adaptive systems exhibit two distinctive characteristics of a) ‘evolving structures,’ by which he means that the system reorganises and changes its constituent parts to adapt to challenges created by their environment, and b) they have a ‘moving target’, which is what makes such a system difficult to understand. GZ meets both criteria proposed by Holland (1992) in having citizen science as a continuously unfolding epistemic object by deliberately changing the model with every iteration. Owing to its reconfiguring of its epistemic architecture to adapt itself to the external environment, GZ can be characterised as a complex adaptive system which achieves collective intelligence through its epistemic architecture. The parallels between the Core Team and the system itself are important, as both the key actors and the system exhibited adaptive qualities. The resourceful practitioner (Edwards A, 2010) is characterised by an ability to identify which parts of the system to approach to find solutions to a problem, if the problem requires expertise that they lack. What is then required from the practitioner is not only to know who knows what, but also to find out how to know who knows what. This quest to find the expertise needed to address a challenge could involve stepping outside the boundaries of the GZ system. Consequently, determining the conceptual boundaries of GZ becomes an interesting question. The lack of clarity over boundaries can be attributed to the deliberately fluid manner in which GZ functioned and was organised.

The GZ Core Team, brought with them networks of expertise, spread beyond the boundary of the GZ system. These networks, in turn, also challenged the notion of what can be described as a system. Nonetheless, I suggest that conceptual boundaries need to be made explicit for the purposes of the present study. First, to limit the analysis to answering the relatively focused research question; and second, to recognise that the idea of a system can incorporate any number of expanding and contracting conceptual planes and boundaries. Hence, examining the role of the practitioners at the core in controlling or managing the boundaries of GZ becomes important in revealing the dynamic processes that underpinned GZ's epistemic architecture.

In an attempt to understand GZ's epistemic architecture as it developed over time, it may be helpful to imagine a dynamic entity, which contracts and expands as the requirements vary with emerging challenges. New practices are developed to meet these challenges, in turn leading to refining the portfolio of expertise drawn upon and developed. Recognising and accessing expertise within and beyond the GZ membership not only changed the expertise at the core, but also the ways in which the elements within the system interacted with each other and with external resources in creating new knowledge. Therefore, through engaging in resourceful practice, a function of adaptive expertise is facilitated by relational expertise over time, through addressing challenges of a varying and unexpected nature in which two phenomena can be observed. The first is seeking out the expertise lacking within the system; and the second is laying a network of expertise outside the system and enfolding that knowledge and expertise back into the core expertise through reflection and learning. In such a system, external expertise can influence the perception of future challenges, and the actions taken to meet those challenges. These processes illustrate the adaptive expertise at the core of GZ which resonated through the entire system, reflecting those attributes through its interactions with the external environment.

Conceptualising GZ as a complex adaptive system also enables an appreciation of the dynamic nature of expertise in and out of the boundaries of GZ's epistemic architecture. The significance of the term epistemic architecture, as I argue, lies in the dynamism of the word 'epistemic' as opposed to 'knowledge' in knowledge production systems. In my understanding of epistemic architecture, the term architecture lends conceptual stability, while epistemic signifies the dynamism inherent in the pursuit of an epistemic object as Knorr-Cetina (1999) contends. It therefore allows a much more dynamic view of expertise that is congruent with the view of intelligence as 'developing expertise' (Sternberg, 1999).

As Pickering (1992: 6-7) points out the need for "new conceptual frameworks built out of concepts that speak directly to practice", a re-examination of expertise within the context of complex adaptive systems would be an exciting direction to undertake for further research.

9.3 Challenges for Citizen Science

Although Citizen Science is an encouraging movement in solving the data deluge problems for science while engaging with science education for the public, there are several areas of contention in the process of opening up the boundaries of science, which will be explored in this section.

9.3.1. Openness and Ethics in Citizen Science

The debate regarding what Open Science is has caught the attention of the research community in the last few years, with foundations such as the Open Knowledge Foundation, Wikimedia, Creative Commons beginning to recognise the debate and participate in it. David (2004) argues that it was with the introduction of the academic journal in the 1600s, that led to the scientists to share resources in order to meet the societal demands to allow access to scientific knowledge. More recently, online open collaborative projects, such as the Polymath project[45] led by a Field Medalist Tim Gowers, have demonstrated that doing research openly in the public can have remarkable benefits (Gowers and Nielsen, 2009). Gowers and Nielsen (2009), two prominent advocates of Open Science, argue that mass open collaborations between scientific fields can expand the problem solving abilities of humans to new limits. The most active forms of discussion take place not through the publishing of systematic research in academic journals, but in clarifying the terms and considering how they could contribute to the community of science.

The concept of Open Science is a broad one, having the potential for considerable transparency in relation to contributing to an understanding of the scientific process and making it more broadly accessible. Open Science has implications for the public, but perhaps will have a more profound impact for science itself. On the one hand, Open Science might mean fast, more efficient and easier movement of data and methods across science, addressing the failures of redundancies and resource wastage, as Samwald *et al.* (2011) argue

[45] The Polymath Blog: http://polymathprojects.org/

in the context of pharmaceutical research. What is certain is that Open Science is likely to have implications for the inner workings of how science is done, being exposed to scrutiny from the inside and the outside. The problem here is that the scrutiny would be carried out not only by other professional scientists, but by individuals outside of that professional community. From the experience of the GZ scientists, engaging directly with the public has led to greater benefits for the Science Team, but there have been instances where volunteers have created and pursued issues that take up time and resources away from the Science Team, which could have been directed towards a more constructive undetaking. It would not be realistic to expect every individual in society to have the domain specific expertise to engage in a constructive and critical dialogue about science, however what is required is a relational approach. The role of spaces for science communication lies in the potential for development of relational expertise required from all parties. The privilege of openness must come with the responsibility towards understanding what science stands for, such as curiosity, rigour, standards, complexity and uncertainty.

Citizen science presents a rather lucrative proposition for science in terms of the potential it offers for large-scale data analysis and data collection which may be used by a number of projects. Since citizen science is still in the developmental stages as a reliable method, practices are being tested and standards for ethical citizen science have not been established. The Core Team, through their experience of engaging with volunteers formulated a few ethical standards that they decided to strive for, which were inspired by the discussion among themselves and the Science Team. One of the instances where the ethics of citizen science had to be established was discussed in chapter 6, the idea of having volunteers go through a completed data set was rejected by the TL on the grounds that it would be a waste of their time, even though it would not be a difficult task to arrange for the data set to be reclassified. The concept of ‘digital sweatshops’ has been discussed by Zittrain (2009) and poses an ethical concern in citizen science. The Core Team only explicitly detailed three ethical principles for GZ (see Appendix F), which were a) viewing volunteers as collaborators, not just users, b) contribution to real research, c) not wasting volunteers’ time. In addition to the publicly stated ‘rules of citizen science’, their commitment to science education and communication can be discussed under ethical practice of citizen science, and relates to the first rule of seeing volunteers as collaborators. As shown in chapter 8, the TL’s beliefs of the ‘duty to innovate’ was revealed, which resonates with the ‘not wasting people’s time’ rule, but also to strive for advancing the way citizen science is done.

I would argue that the consideration of ethical engagement with volunteers could be a function of relational expertise, as the Core Team demonstrated. Although the Science Team as a whole recognised the contributions of volunteers in their research, it is the Core Team which most consistently exhibited a relational approach towards the volunteers. Their decisions relating to design and communication have reflected the deliberation on what constitutes ethics in citizen science as scientists, which leads to the next challenge of how the public is perceived in the scientific community.

9.3.2 Perception of Public and Public Engagement Among Scientists

The lack of research on scientists’ attitudes towards public engagement poses a challenge to the development of citizen science. In the Survey of Factors

Affecting Science Communication by Scientists and Engineers (Royal Society, 2006) with a sample of 1485 scientists, 20 per cent thought that their colleagues would view peers who engaged in science communication unfavourably. In a questionnaire study conducted in a single university in the UK with approximately 170 participants, it was revealed that the factors that prevented public engagement efforts by scientists were neither time and money constraints, nor lack of recognition, but primarily the lack of experience or perceived skills, and a negative attitude or belief that their peers did not engage in such activities themselves. These findings suggest that the attitudes of scientists towards the public are not helpful, and the value that might be placed on public engagement lacks encouragement.

The attitude towards the volunteers at the inception of GZ was that they were amateurs, as the first primary task on GZ was designed to require minimal training for the volunteers, caliing for only the very basic cognitive capabilities. That attitude changed over time as the volunteers demanded tasks that were more advanced, and the team discovered that certain groups of volunteers had the capability of performing some tasks better than others. These revelations changed the view that scientists had of the volunteers and opened up avenues for complex and useful data reduction. From the public being viewed as a single homogeneous group, the scientists recognised the capabilities and expertise of different groups of volunteers, which had implications for task design. The project also undertook a study to understand the motivations for volunteer participation, demonstrating a willingness to understand the volunteers in order to improve ways in which they could retain their participation and enhance their experience of engaging with GZ. The lesson learnt from GZ’s experience is that what citizen science offers science, depends on how well scientists understand the motivations and capabilities of volunteers.

However, it is within reason to assert that most scientists might not hold that view or have the inclination to perceive the public as collaborators. Nonetheless, there is an encouraging trend of increasing efforts to train scientists for science communication, as institutions such as the Royal Society have begun to present awards for science communication[46], recognising that it needs to be part of the training for scientists who want to engage in citizen science. One of the few qualitative studies on scientists engaging in citizen science by Riesch and Potter (2013) examined the attitudes towards public engagement among the scientists. It revealed that while participants found citizen science worth engaging in, their notions of public engagement seem to remain unchanged. Therefore, it is plausible that some versions of citizen science do not change science communication practices or attitudes among the scientists who engage in it.

9.3.3 Training for Scientists working with Citizen Science

As the findings of this study suggest, working with citizen science requires additional set of practices and expertise that is not a part of the current training that scientists receive. Nonetheless, there is encouragement in forms of: training; prizes for public engagement (Poliakoff and Webb, 2007); public events offered by universities, foundations and organisations such as the Royal Society; and science-events increasingly becoming a part of popular culture. Prominent

[46] The Michael Faraday Prize "The Royal Society Michael Faraday Prize is awarded annually to the scientist or engineer whose expertise in communicating scientific ideas in lay terms is exemplary. Normally, preference will be given to a practising scientist or engineer, but other individuals whose primary expertise is in writing, broadcasting or other relevant forms of communication may also be considered." https://royalsociety.org/awards/michael-faraday-prize/

examples include the annually held World Science Festival in New York, Science Museum 'Lates' for adults only in London, and popular television programmes such as the Big Bang Theory, which reaches an audience of approximately 5 million in the US alone. The latter revolves around the life of a group of professional scientists struggling with their encounters with non-scientists, and their frustrations with the general lack of public interest in science. Although these efforts may be an encouraging one in popularising science among the public, it still falls under the paradigm of 'public understanding of science' where the public plays a largely passive role in understanding the process of science following the deficit model (Bauer, 2009).

As chapters 6 and 7 revealed, scientists need to understand the practice of translation 3 and 6 (Refer to Table 6.1) and perhaps other forms of translation that may emerge. In addition, ensuring the reliability of annotations received from volunteers led the GZ scientists to collaborate in experiments in psychology with researchers outside GZ in order to reveal any biases in classification behaviour. This exercise demonstrated the ability of GZ members to extend their problem to another domain of expertise, evincing adaptive behaviour. Working with citizen science data therefore requires a different set of practices, and a general ability for innovative solutions. Therefore, there are implications for how scientists are trained for citizen science. In addition to relational expertise in communicating with the public, there is the challenge of addressing the fundamental perception of the role of non-professionals as potential collaborators. Not only does Big Data citizen science require skills for working with data-intensive science (Edwards P *et al*. 2013; Jirotka *et al*. 2013), but adaptive and relational expertise could become as important in running successful citizen science projects.

9.3.5 Structural Challenges in Academia

In following the participants through their attempts to sustain GZ within the boundaries of academia, the study identified some struggles that scientists should be aware of. For the first year of the project, the work to support GZ was done on a voluntary basis by the developers and scientists involved, which could not be sustained. Since citizen science was not a proven success at the time of GZ's inception, grants from various sources were secured, a large proportion through science education-related grants. The current environment is a little different as the success of GZ and a few other large-scale citizen science initiatives have now led the way for new projects. However, citizen science still has to maintain its growing reputation. Although GZ might have successfully managed to establish their methods as reliable ways to conduct scientific research in their disciplinary field, skepticism within other research disciplines in relation to publications, conferences and collaborations can be expected.

In addition, support in terms of technical infrastructure needs to be considered. In the case of GZ, the university had initially refused to run the servers from within the university premises, requiring them to be located in another institution in the US. The perceived trade off between security of information technologies and innovation poses a structural challenge that institutions must address if citizen science projects like GZ are to be supported.

The GZ scientists' perception of the public changed only with experience of working with them. It was a fundamental shift that required a supportive structure of incentives and opportunities. Such a structure is currently lacking in academic organisations. One lesson from the present study is that the domain-expertise of project managers legitimises their actions a citizen science team, but also calls for relational and adaptive expertise. Therefore, it should be expected that the individual who undertakes the responsibility will face a trade-off between advancing their disciplinary knowledge and publishing scientific output and making the project work. Therefore, organisations that are willing to support and fund such endeavours need to recognise what the role entails and align their expectations in their hiring and evaluating practices.

9.3.6 The Design Challenge and the Role of Social Science

The talk about 'design' in the GZ project emerged over a few years, and began to be understood as a deliberate process of decision making with an ongoing understanding of how the volunteers would respond and behave to the changes made with every iteration of GZ. The PI emphasised that making responsive modifications was one of the hardest challenges in designing for serendipitous discovery. One of the strong features of GZ's citizen science was the discovery of new astronomical objects that was a result of volunteers scrutinising an unfamiliar object, and pursuing that inquiry which led to a new discovery. The question of how GZ could enable such discoveries and build a system that allows that potential to be realised, is the next difficult question that the team faces. Designing such a system involves understanding how motivation among the volunteers works, and which features of the system facilitate inquiry and which ones do not. In the search of these answers, the GZ project has collaborated on social science research in the hope that those studies will be able to influence their design decisions. However, there are questions relating to the task difficulty and variation that has to be managed. The GZ scientists recognised that even though some individuals were more skilled than others in certain tasks, the lack of variation in tasks might affect performance and motivation of those volunteers. Therefore, understanding human behaviour and motivation, and designing a system that corresponds to that behaviour for the outcome, remains an ongoing challenge, where social science research could contribute. In addition, the notion of phronetic social science (Flyvberg, 2001), when applied to citizen science, could help scientists address some of the ethical concerns in designing volunteer participation for citizen science. Although, GZ's design was not game-based, the notion of epistemic games (Shaffer, 2006) that are designed for the purpose of providing rich learning experiences could also provide lessons to assess the volunteers' developing expertise (Rupp *et al*., 2010).

Although far from an exhaustive discussion, this section presented some challenges in pursuing citizen science in the context of present academic science, based on the findings of this study. Issues around the meaning and practice of open science were explored, followed by the hurdles facing the current state of training for scientists, structural problems in academia and perception of the public held by scientists which were addressed with an emphasis on ongoing efforts to design a system that corresponds to the skills, motivation and behaviour of volunteers. There are, however, a few smaller concerns that can be raised in relation to the success of a citizen science project. The effect of a reputation of an organisation such as a university attached to a project is not clearly understood or known, and even though the role of media was seen to be a significant one in reaching out to the public, the extent of the dependency has not been studied and needs to be examined. Exploring these challenges would certainly provide further questions for research in understanding the practice of citizen science and its implications on society, since this study marked only the beginning of such research, the shortcomings and strengths of which are presented in the following section.

9.4 Implications for Science and Society, Science and Education

The implications of this research will be discussed under four different themes; a) Big Data science, b) the relationship between science and society (past and current paradigm), c) expertise for science, and d) science education.

9.4.1. Data-Intensive Science in Society

As Bell (2009.) argues, the current state of scientific inquiry is large amounts of data for which computing is indispensable. Computational methods in science have increasingly permeated across various disciplines, with genomics and astronomy leading the way in the amount of data generated and gathered and with algorithms and coding becoming staple terminologies. Citizen Science is one of the ways in which scientists have chosen to tackle the challenge of data deluge or Big Data. There are several categories of citizen science initiatives, with differences depending on how and at what level the public is allowed to engage in the process of scientific inquiry. This study explored and examined a model that became based on a symbiotic epistemic relationship between scientists and their volunteers. In addition to tackling the challenge of Big Data science, GZ engaged its volunteers in the kinds of tasks that computer algorithms did not perform as satisfactorily as humans with even minimal training. With collaborators from the field of Artificial Intelligence, the GZ team attempted to advance machine learning using the behaviour of its volunteers to ensure that the human volunteers were only performing the tasks that computers could not yet reliably complete.

Other similar computationally-challenging citizen science initiatives such as EyeWire and FoldIt asks volunteers to perform tasks in which the computers are not as good as humans. However, there is a fundamental difference between the GZ approach and the driving principles behind some of the other citizen science projects. In addition to contributing to scientific research by carrying out simple tasks, the GZ project has encouraged volunteers to learn more about the topic by providing them with spaces where they can further investigate the field of astronomy. The team also engages in communicating the progress of their research publicly through social media and mailing lists, and some of the science team members actively interact with volunteers on GZ's online discussion boards, inspiring ideas for papers and further research.

Citizen science as a data reduction method in the current paradigm of data intensive science is not only significant to science as a discipline, but is worthy of recognition in how we understand the relationship between science and society and public understanding of science. As recent research suggests, there is a significant move towards making science more accountable to the public (Nowotny, 2005), whether that is through more 'public understanding of science' events or the 'impact' measurement of research. It has been contended that a lack of transparency in scientific investigations and lack of consensus among scientists themselves leads to public distrust of science (Wynne, 1995).

GZ's model of citizen science is a symbiotic in nature where in addition to solving the data deluge problem, science education became a significant drive for the

GZ team. Allowing non-professionals to be positioned as epistemic subjects in the epistemic architecture of GZ is an indication of reconfiguration of how science and society relate to each other.

9.4.2 Science Education

The implications of the work of the Science Team of GZ for science education is not just a matter of improved communication with the public. One way of seeing it would be that GZ's approach to citizen science in raising public awareness and understanding of the scientific process can be one of the ways forward in increasing scientific literacy. An individual with Internet access, and curiosity can access scientific processes and a community of volunteers eager and willing to carry out research on their own, with the help of professional scientists. Secondly, GZ's model of citizen science includes discussion about scientific method and how the scientific community works, which provides helpful insights for the volunteers into the social world of science, which so often is a black box for the public. The damage arising from education systems which have taught students in schools to understand science in terms of facts and not process, has a chance to be corrected through citizen science. The 'public misunderstanding of science' (Wynne, 2006: 215) stands to be corrected through the kind of participatory model that GZ has managed to sustain and strengthen over a period of five years from 2007-2013. Thirdly, as a result of learning about science from a citizen science model, there is a huge potential for the individuals to not only be informed about how science is done, but to understand the need for science communication directly from professional scientists and the importance of having a space for that interaction. The complexity inherent in scientific investigations deserve a dialogue between those who carry them out professionally and are called scientists, and the whose lives are dictated by products of scientific and technological endeavours.

In summary, GZ's model of citizen science is an effort to build a platform for that dialogue, greatly enabled by web-technologies and social media. It is driven by principles that place understanding at the heart of knowing, and has implications not for only reconceptualising the relationship between science and society, but calls for reconfiguring the education and training of scientists of the future.

9.5 Reflections On Researcher Role

Before we proceed to the concluding remarks of this thesis, this section presents a reflection on my role as a researcher. There are three key points I will discuss in this section, a) the participants' perception of my role as a researcher, b) the uncertainty and complexity of this research and finally, c) the parallels between my research and my participants' epistemic journeys.

While it took several weeks to negotiate access into the GZ team, I strongly believe that the open-ended nature of the interviews I conducted along with my personal interest in the field of astronomy, influenced the perception of my key participants held of me and perhaps the role I played in the kind of reflexive thinking they engaged in. During an early interview in March 2010, the PI remarked:

"[y]eah we are trying to do it regularly. It is partly your fault, you kept asking about how we communicate [...] so whole lot of that. I thought about that, about how we changed our behaviour." - PI

On multiple occasions, colleagues of the participants would refer to our interview meetings as ‘therapy sessions,’ as my participants confessed that they found that verbally expressing their concerns and challenges, which they would not have the space or the time to talk about otherwise, was enjoyable and ‘not work’. As a researcher, I was initially concerned about being able to engage them in regular conversations about the project over an extended period of time. However, the openness of the Core Team and their curiosity about my curiosity to study their work, were two of the key elements that sustained my access.

One of the methodological advantages of employing Knorr-Cetina’s (1999) epistemic object-subject relations as a way of understanding GZ’s practices and expertise was the recursive analytical note-taking that continually informed my research design, which was discussed in chapter 4. It not only allowed for a continuous dynamic understanding of GZ’s work, but as my questions reflected an evolving understanding, my participants began noticing that I was not merely collecting data, but trying to make sense of their work along with them. My skepticism about being unable to keep up with references to jargon and technical terms in Astronomy was met with enthusiasm that the Core Team had for communicating their work to others. One of the confirmations I received was a statement made by the PI while I was trying to clarify my understanding of their criteria for selecting new projects, implying that I had internalised their philosophy on citizen science. “You’ve drunk the Kool-Aid”

Towards the end of the fieldwork, the PI mentioned that he would be interested in sending me to proposal meetings as he believed that I had understood their selection process and was in a position to make a judgement they would approve of. “I would love to send you to one of the proposal meetings.” Although their comments of confidence allowed me to conduct the unstructured interviews in a way that granted me a privileged position which was no longer of an outsider, but someone who understood them, I had to consciously maintained my focus on my subjective views and remain as objective as possible in raising critical questions, while sharing the enthusiasm for the possibilities ahead of GZ as pioneers in the field of online Big Data citizen science.

I believe it was an advantage that the GZ’s work was in the field of Astronomy, in which I have always had tremendous personal interest. My existing social network of doctoral students in Astrophysics proved to be extremely helpful in understanding the context, and led to resources for references that would otherwise have been overlooked. My participation in early crowdsourcing project SETI@Home was another common point of interest that I shared with the PI early on, which was also a source of inspiration for the GZ team. Therefore, establishing a shared source of inspiration and enthusiasm may not have been important, but it certainly made the research process more engrossing for me as a researcher, and the PI as a participant. Another parallel that we drew in common was our tendencies to be ‘generalists’ with simultaneous engagement in multiple areas of interest, which was reflected in how the PI approached various disciplinary fields for solutions and inspiration. This coincidence allowed me to keep up with his way of thinking about a challenge during the interviews, and sustain his interest with minimal interruptions. Therefore, having shared interest in the pursuit of multiple common epistemic objects with the participants of the study was believed to very helpful during the research process.

Finally, I would like to address the issue of perceived complexity of GZ in my analysis. I find myself in agreement with Tsoukas and Hatch (2001) in their views that while complexity is a characteristic of systems we examine, the ways of organising our thinking about them is also important. Casti (1986) contends that "Complexity is a contingent property arising out of the I interaction between a system S and the observer "O (Casti, 1986:149), which implies that complexity is dependent on the observer and is not an intrinsic property of the study. Checklands (1981) further elaborates that how we define where an event begins and ends as we shift focus from one level of the system to another are not properties of the system, but are interpretative moves. In presenting the findings of this study, the reader must be aware that reflexivity plays an important part as the narrator's inclusion in the narrative adds an additional layer of context as suggested by Tsoukas and Hatch (2001).

In reflecting on the influences on my interpretation of the analysis of the data for this study, I find that the systemic view of society and nature put forward by Capra (1982) and Bohm (1980) has shaped my own thinking. Trained as a theoretical physicist with a background in systems theory, Capra (1982) argues that the Cartesian reductionist views that have become prevalent not only in scientific thinking, but also in how we understand society and nature has contributed to the current ecological, social and economic crisis. Bohm (1980), a prominent theoretical physicist and the protege of both Albert Einstein and Niels Bohr, takes a critical stance on the fragmented nature of academic disciplines in understanding the world and the nature of reality. He later proposed a new way of dialogue by re-imagining the concept of dialogue with the emphasis on understanding as opposed to gaining knowledge, as he stated:

"What is essential here is the presence of the spirit of dialogue, which is in short, the ability to hold many points of view in suspension, along with a primary interest in the creation of common meaning." (Bohm and Peat, 1987:247)

Both Capra (1982) and Bohm (1980) urge the need for a holistic framework to appreciate and account for the complex problems that human society faces today. Both influences are reflected in my approach to synthesising the findings in conceptualising GZ's epistemic architecture as a complex adaptive system.

9.6 Challenges and strengths of this study

While some of the challenges of undertaking this study were anticipated prior to gathering data, several shortcomings and further challenges were noted throughout the process of pursuing this research study.

9.6.1 Challenges and Shortcomings

Prior to discovering GZ, my aim was to investigate the interactions between public intellectuals and the public, primarily mediated by information-technology and social media. I was particularly interested in examining a combination for science communication and technology, when I serendipitously discovered GZ which provided a very intriguing setting to understand scientists who were engaged with the public. What was clear from the beginning was the lack of established theoretical framework that would allow me to refine my analytical concepts, and although it posed a challenge that revealed itself over the process of data collection, it permitted the opportunity to allow my understanding to unfold through observing GZ. The uncertainty of the direction of the project was a risk that I understood, and neither the GZ team nor I could have predicted the development that followed. After almost a year into data gathering, I began to study the framework of collective intelligence, which is an emerging field of inquiry yet to be populated with empirical and theoretical bodies of work. Although initially, I drew analytical inspiration from the fields of distributed cognition to inform my understanding of GZ as a knowledge production system, the framework for collective intelligence allows the synthesis of findings in this study.

Although this study investigates the practices of the Science Team, the evolution of the expertise of only the Core Team was chosen as a focus. It is understood that the inclusion of the entire Science Team would have been more robust, however not only were there constraints of time and resources for this study, but the entry and departure of Science Team members did not offer a stable group of participants for an extended period of data gathering. In addition, the focus of the study was more on the citizen science aspect of GZ. It was determined that the members who were engaged full-time and responsible for all related decisions would provide a better focus. Also, my lack of fundamental domain knowledge of astrophysics might have led to my overlooking potential findings. Cross comparison between projects may have provided a stronger basis for some of the claims in the study, but the scale of that kind of study was not possible in a DPhil.

This study did not focus on the volunteers. Although this was a consequence of the GZ Education Team carrying out research on the volunteers, enlarging the focus in this way would also have made the scope of this study unmanageable. However, including volunteers in the study would have been a more balanced representation of 'co-production' in citizen science. Also, while it was known that the participation of female scientists was in the GZ Science Team was relatively lower in proportion to that of male scientists, I did not include the examine of gendered exclusion or inclusion in the pursuit of physics research as Hasse (2008) and Hasse and Trentemøller (2011) have carried out.

Although this study aimed to capture the communication between the Science Team members, I was made aware of the fact that there were many back channels of communication such as regular video conferencing, face to face meetings in different locations, and e-mails that were exchanged outside of the formal team mailing list. Also, the communication between the PI and the TL occurred not only through tech-mediated platforms, but they relied heavily on personal communication in the year that they were sharing an office space. Clearly, the data used in this study does not include or account for the relevant exchange that might have occurred within the team and therefore, in addition to the mailing list and additional documents provided to me by the core participants, I relied heavily on the report and recollection of events and conversations through interviews at regular intervals with the participants. Having presented some of the challenges faced during this study, it is not without some advantageous opportunities.

9.6.2 Strengths

I acknowledge the opportunity of gaining access to a group of individuals who at the beginning of the study were experimenting with citizen science and over the period of a few years went on to become the most successful citizen science project understood in terms its scalability and number of scientific publications. Neither my participants nor I had anticipated the success that followed. Not only was I allowed access into the workings of GZ, but my participants were very generous with their time and access. The data gathering period began in February 2010 and ended in April 2013, which allowed me to follow events as they happened. Although it was not feasible to include all the members of GZ in the study, I had constant access to the two key members who were the decision-makers and the only two members who were employed full-time to carry out GZ - related work at my point of entry into the field.

In pursuit of understanding how GZ was making citizen science work, I had the opportunity to co-construct conceptual maps and tools with my participants, which eventually led to filling lexical gaps in understanding knowledge systems by introducing the terms epistemic architecture and phronetic design. The lack of a structured and pre-determined research design allowed for engagement with the participants, that was not only enjoyable but revealing for both the participants and myself. We were in pursuit of our respective epistemic objects, with both parties intrigued by the process of the other.

9.8 Concluding Thoughts

This study provides an example of how conceptualising an epistemic architecture, drawing inspiration from the concept of collective intelligence (Malone *et al.,* 2009), can help reveal the critical constituent elements within a system and how they are related to each other. The salient elements in the present study were the practices and the expertise that operated within them to create GZ's approach to citizen science. GZ's model not only produced scientific output by addressing the challenges posed by Big Data, but also recognised science education and communication as a critical practice, signifying a symbiotic relationship between professional scientists and citizen scientist. These feature not only indicate a movement away from the deficit model of public understanding of science, but suggest an encouraging turn towards a relationship between science and society led by curiosity and participation in the scientific method. One of the key differences between crowdsourcing and citizen science in GZ's case, I would argue, is that while the former aims to achieve collective intelligence, the latter is an example of a symbiotic collective intelligence.

In addition to the new practices of translation identified, this study also revealed the interplay of different kinds of expertise exhibited by the scientists in running GZ. The design principles that shaped the development of the Core Team's expertise were identified, and the interplay of dominant expertise for each area of attention recognised. The development of expertise in GZ's citizen science was theorised as: Domain (citizen science) = Domain (Astronomy) + Domain (Development) + Relational Expertise + Adaptive Expertise + Interactional Expertise in Development (On behalf of the PI)

As academic research moves towards Mode 2 (Gibbons *et al*., 1994), which extends knowledge production outside of the boundaries of academic institutions into partnerships with industry, governments and the public sphere, the role of relational and adaptive expertise in addition to domain expertise among scientists becomes significant. As presented in Chapter 7, since GZ''s activities and practices involve crossing boundaries of academic research practices by having to engage with the volunteers, non-academic research institutions, media agencies and science communication with the public at large, the role of relational expertise and adaptive expertise emerged as being significant to GZ's success. The developing expertise of the GZ Core Team calls for revisiting of the term 'expert'. Etymologically, the words expert, experiment and experience, share the same root, which means to try, to test[47]. Therefore, an expert can be seen as an individual who attempts to understand and act, in pursuit of his or her epistemic object (Knorr-Cetina, 1999) able to work with skillful incompleteness (Lindblom, 1979). As the object is inherently incomplete, and the subject perceives the signals of incompleteness and reflects. Epistemic object and the subject interact in their engagement with one another. While a relational approach (Edwards A, 2010) is needed to understand this complexity, acting on an understanding of the complexity requires adaptive abilities of an expert, or adaptive expertise. In other words, an expert may not be someone who knows, but someone who is repeatedly willing to try and test. Therefore, the examination of the expertise in this study sheds light on how expertise in new domains is developed rather than on the acquisition of established domains of expertise.

---

[47] Source: http://www.etymonline.com/index.php?term=expert

The findings of this study reveal the emerging practices of scientists who engage in citizen science while being committed to science and science education, and shed light on the significance of relational and adaptive expertise in addition to domain expertise in working across various boundaries to make citizen science work. It is my hope that the findings contribute to the empirical literature in citizen science studies and offer epistemic architecture and phronetic design as conceptual tools in further investigating citizen science. This research study has been an engrossing epistemological journey, and I look forward to continuing my pursuit of understanding the conceptual, methodological and empirical findings, as Bohm (1980:ix) so eloquently expresses:

"I would say that in my scientific and philosophical work, my main concern has been with understanding the nature of reality in general and of consciousness in particular as a coherent whole, which is never static or complete but which is an unending process of movement and unfoldment."

10. Reference List

Agamben G and D'Isanto L (2009) The signature of all things: on method, New York: Zone Books.

Anderson C (2008) The End of Theory, Wired Magazine, 16.

Argote L, Beckman S L, and Epple D (1990) The persistence and transfer of learning in industrial settings, Management science 36(2) 140-154.

Banerji M, Lahav O, Lintott C J, Abdalla F B, Schawinski K, Bamford S P and Vandenberg J (2010) Galaxy Zoo: reproducing galaxy morphologies via machine learning, Monthly Notices of the Royal Astronomical Society 406(1) 342-353.

Baraniuk R G (2011) More is less: signal processing and the data deluge, Science 331(6018) 717-719.

Barnes S J (2001) Knowledge management systems: theory and practice, Cengage Learning.

Bartky I R (1989) The Adoption of Standard Time, Technology and Culture 30(1) 25-56.

Bauer M W (2008) Survey research and the public understanding of science (pp. 111-130), Routledge.

Bauer M W (2009) The evolution of public understanding of science—discourse and comparative evidence, Science Technology & Society 14(2) 221-240.

Bauer M W (2014) A word from the Editor on the special issue on 'Public Engagement', Public Understanding of Science 23(1) 3-3.

BeaulieuA (2004) Mediating Ethnography: Objectivity and the Making of Ethnographies of the Internet, Social Epistemology 18(2-3) 139-163.

Beaulieu A (2005) Sociable hyperlinks: An ethnographic approach to connectivity, Virtual methods: Issues in social research on the Internet 183-198.

Beaulieu A (2010) Research Note: From co-location to co-presence: Shifts in the use of ethnography for the study of knowledge, Social Studies of Science 40(3) 453-470.

Beaulieu A and Simakova E (2006) Textured connectivity: An ethnographic approach to understanding the timescape of hyperlinks, Cybermetrics 10(1).

Beaulieu A, Scharnhorst A and Wouters P (2007) Not another case study: A middle-range interrogation of ethnographic case studies in the exploration of e-science, Science, Technology & Human Values 32 672–692.

Bechhofer S, Ainsworth J, Bhagat J, Buchan I, Couch P, Cruickshank D, Roure D D, Delderfield M, Dunlop I, and Gamble M (2010) Why Linked Data Is Not Enough for Scientists, 2010 IEEE Sixth International Conference on e-Science 300-307.

Beecher T and Trowler P R (1989) Academic Tribes and Territories, Society for Research into Higher Education and Open University Press.

Bell (2009) In Hey T, Tansley S, and Tolle K (Eds.). (2009). The Fourth Paradigm: Data-Intensive Scientific Discovery, Redmond WA: Microsoft Research.

British Educational Research Association (BERA) (2004) Revised Ethical Guidelines for Educational Research.

Bhaskar R A (1987) Scientific Realism and Human Emancipation, London: Verso.

Bhaskar R A (1989) Reclaiming Reality: A Critical Introduction to Contemporary Philosophy, London: Verso.

Bhaskar R A (1997) A Realist Theory of Science, London: Verso.

Blanchette J-F (2012) Burdens of Proof: Cryptographic Culture and Evidence Law in the Age of Electronic Documents, Cambridge: MIT Press.

Bohm D (1996) On Dialogue, New York: Routledge.

Bohm D and Peat F D (1987) Science, order, and creativity: A dramatic new look at the creative roots of science and life, New York: Bantam.

David B (1980). Wholeness and the implicate order. Routledge & Kegan, London.

Barron, B. (2000). Problem solving in video-based microworlds: Collaborative and individual outcomes of high achieving sixth grade students. Journal of Educational Psychology, 92(2), 391–398.

Bond C E, Shipton Z K, Jones R R, Butler R W H and Gibbs A D (2007) Knowledge transfer in a digital world: Field data acquisition, uncertainty, visualization, and data management, Geosphere [Online] 3(6) 568-576.

Borgman C L, Wallis J C, Enyedy N, and Mayernik M S (2006) Capturing Habitat Ecology in Reusable Forms: A Case Study with Embedded Networked Sensor Technology. Paper presented at the Annual Meeting of the Society for the Social Studies of Science, Vancouver, BC.

Bredekamp (2012) in Way M J, Scargle J D, Ali K M, and Srivastava A N (Eds.) (2012) Advances in machine learning and data mining for astronomy, CRC Press.

Brooks R (2001) The relationship between matter and life, Nature, 409-412

Buchanan W J, Kwecka Z, and Ekonomou E (2013) A privacy preserving method using privacy enhancing techniques for location based services, Mobile Networks and Applications 18(5) 728-737.

Burns T W, O'Connor D J and Stocklmayer S M (2003) Science commjunication: A contemporary definition, Public Understanding of Science 12 183-202.

Burton M, and Jackson S J (2012) Constancy and Change in Scientific Collaboration: Coherence and Integrity in Long- Term Ecological Data Production. Proceedings of the 2012 Hawaii International Conference on Systems Science, 353-362.

Busher H and Clarke S (1990) The ethics of using video in educational research In Anning A, Broadhead P, Busher H, Clark S, Dodgson H, Taggart L, White S and Wlison R (Eds.) Using video recordings for Teacher Professional Development, Leeds: University of Leeds, School of Education.

Callon M, Lascoumes P, and BartheY (2009). Acting in an uncertain world: An essay on technical democracy, MIT Press: Cambridge MA

Callon M and Latour B (1981) Unscrewing the big Leviathan: how actors macro-structure reality and how sociologists help them to do so. Advances in social theory and methodology: toward an integration of micro and macro-sociologies, 277-303.

Campbell P (2005) Escape from the impact factor, Nature 435 pp. 1003.

Capra F (1982) The Turning Point: Science Society and the Rising Culture, Wildwood House London.

Carley K (1992) Organizational learning and personnel turnover, Organization Science, 3(1) 20-46.

Carlile P R (2002) A Pragmatic View of Knowledge and Boundaries: Boundary Objects in New Product Development, Organization Science 13(4) 442-455.

Carvalho, (2007) Ideological cultures and media discourses on scientific knowledge: re-reading news on climate change, Public understanding of science [Online] 16(2) 223.

Castells M, Fernandez-Ardevol M, Qiu J L, and Sey A (2009) Mobile communication and society: A global perspective, MIT Press.

Casti J (1986) On system complexity: Identification, measurement, and management. In J. Casti & A. Karlqvist (Eds), Complexity, language, and life: Mathematical approaches. Berlin: Springer-Verlag.

Knorr-Cetina K (2008) Objectual Practice. Knowledge as Social Order: Rethinking the Sociology 83.

Chan S (2001) Complex Adaptive Systems. In research seminar in Engineering Systems, Research Seminar in Engineering Systems October 31, 2001/November 6, 2001

Chi M T (2011) Theoretical perspectives, methodological approaches, and trends in the study of expertise. In Expertise in mathematics instruction (p17-39) Springer US.

Chi M T, Glaser R, and Farr M J (Eds.) (2014) The nature of expertise, Psychology Press.

Chompalov I, Genuth J and Shrum W (2002) The organization of scientific collaborations, Research Policy 31(5) 749-767.

Cohen M J (1997) Risk society and ecological modernisation alternative visions for post-industrial nations, Futures 29(2) 105-119.

Collins H M and Evans R (2002) The Third Wave of Science Studies: Studies of Expertise and Experience, Social Studies of Science 32(2) 235-296.

Collins H M(1996) In Praise of Futile Gestures: How Scientific is the Sociology of Scientific Knowledge? Social studies of science [Online] 26(2), 229.

Collins H M (1998) The Meaning of Data: Open and Closed Evidential Cultures in the Search for Gravitational Waves, American Journal of Sociology 104 (2) 293-338.

Collins H M and Evans R (2002) The Third Wave of Science Studies: Studies of Expertise and Experience, Social Studies of Science 32(2) 235-296.

Collins H M (1975) The seven sexes: A study in the sociology of a phenomenon, or the replication of experiments in physics, Sociology 9(2) 205-224.

Collins H M (1983) The sociology of scientific knowledge: studies of contemporary science, Annual Review of Sociology 265-285.

Collins R (1998) The sociology of philosophies: a global theory of intellectual change, Cambridge, Mass; London: Belknap Press of Harvard University Press.

Corley E A, Boardman P C and Bozeman B (2006) Design and the management of multiinstitutional research collaborations: Theoretical implications from two case studies, Research Policy 35(7) 975-993.

Cummings J, Finholt T A, Foster I, Kesselman C, and Lawrence K A (2008) Beyond Being There: A Blueprint for Advancing the Design, Development, and Evaluation of Virtual Organizations: Report from an NSF Workshop on Developing VIrtual Organizations.

Czarniawska (1997) A four times told tale: Combining narrative and scientific knowledge in organization studies, Organization The Interdisciplinary Journal of Organizationory and Society [Online] 4(1) 7.

David P A (2004) Can "open science" be protected from the evolving regime of IPR protections? Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft, 9-34.

Dreyfus H L, Dreyfus S E, and Zadeh L A (1987) Mind over machine: The power of human intuition and expertise in the era of the computer, IEEE Expert 2(2) 110-111.

De Roure D, Bechhofer S, Goble C, and Newman D (2011) Scientific Social Objects: The Social Objects and Multidimensional Network of the Myexperiment Website. Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, 1398-1402

De Roure D, Goble C, Aleksejevs S, Bechhofer S, Bhagat J, Cruickshank D, Fisher P, Hull D, Michaelides D, and Newman D (2010) Towards Open Science: The Myexperiment Approach, Concurrency and Computation: Practice and Experience 22(17) 2335-2353.

Dhami M K, and Wallsten T S (2005) Interpersonal comparison of subjective probabilities: Toward translating linguistic probabilities, Memory & cognition 33(6) 1057-1068.

Dreyfus H L (1980) Holism and Hermeneutics, The Review of Metaphysics 34(1) 3-23.

Dreyfus H L and Dreyfus S E (2005) Peripheral vision expertise in real world contexts, Organization studies 26(5) 779-792.

Drucker P E (1992) The society of organizations, Harvard business review 95-104.

Edge D O and Mulkay M J (1976) Astronomy Transformed. The Emergence of Radio Astronomy in Britain, New York: Wiley.

Edwards A (2005). Relational agency: Learning to be a resourceful practitioner. International Journal of Educational Research, 43(3), 168–182

Edwards A (2010) Being an expert professional practitioner: The relational turn in expertise (Vol. 3), Springer.

Edwards A (2011) Building common knowledge at the boundaries between professional practices: Relational agency and relational expertise in systems of distributed expertise, International Journal of Educational Research 50(1) 33-39.

Edwards A (2012) The role of common knowledge in achieving collaboration across practices, Learning, Culture and Social Interaction 1(1) 22-32.

Edwards P N (1996) The Closed World: Computers and the Politics of Discourse in Cold War America, Cambridge, MA: MITPress.

Edwards P N (2010) A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming, Cambridge, MA: MIT Press.

Edwards P N, Jackson S J, Chalmers M K, Bowker G C, Borgman C L, Ribes D, Burton M, and Calvert S (2013) Knowledge Infrastructures: Intellectual Frameworks and Research Challenges, Ann Arbor: Deep Blue

Edwards P N, Mayernik M S, Batcheller A L, Bowker G C and Borgman C L (2011) Science Friction: Data, Metadata, and Collaboration, Social Studies of Science 41(5) 667-690.

Eisenberg E M (2007) Strategic ambiguities: essays on communication, organization, and identity, Thousand Oaks: Sage Publications.

Ericsson K A and Smith J (Eds.) (1991) Toward a general theory of expertise: Prospects and limits, Cambridge University Press.

Ericcson K A, Krampe R T, and Tesche-Romer C (1993). The role of deliberate practice in the acquisition of expert performance. Psychological Review, 100, 363–406

Etzkowitz H (1998) The norms of entrepreneurial science: cognitive effects of the new university–industry linkages. Research Policy, 27(8), 823-833.

Evans J A (2008) Electronic publication and the narrowing of science and scholarship. science, 321(5887), 395-399.

Eynon R, Schroeder R, and Fry J (2009) New techniques in online research: Challenges for research ethics, Twenty-First Century Society 4(2) 187-199.

Ferber P, Foltz F and Pugliese R (2007) Cyberdemocracy and Online Politics: A New Model of Interactivity, Bulletin of Science Technology and Society 27(5) 391-400.

Fleck L (1935 Genesis and development of a scientific fact, Chicago IL: University of Chicago Press

Flyvbjerg B (2001) Making social science matter, Cambridge: Cambridge University Press.

Fortson L, Masters K, Nichol R, Borne K, Edmondson E, Lintott C, and Wallin J (2011) Galaxy Zoo: morphological classification and citizen science. arXiv preprint arXiv:1104.5513.

Foster I, Kesselman C, and Tuecke S (2001) The anatomy of the grid: Enabling scalable virtual organizations, International journal of high performance computing applications 15(3) 200-222.

Frickel S and Gross N (2005) A General Theory of Scientific/Intellectual Moveents. American Sociological Review, [Online] 70 (2)204-232.

Fry J (2006) Coordination and control of research practice across scientific fields: Implications for a differentiated e-science. New infrastructures for knowledge production: Understanding e-science, 167-87.

Fujigaki Y (1998) Filling the gap between discussions on science and scientists' everyday activities: applying the autopoiesis system theory to scientific knowledge. Social science information, 37(1), pp. 5.

Fuller S (1991) Social epistemology, Bloomington, IN: Indiana University.

Fuller S (2005) The intellectual, Thriplow: Icon.

Bell G, Gray J, and A Szalay A (2009) Petascale Computational Systems,IEEE Computer, pp. 110–112, vol. 39, 2006, doi: 10.1109/MC.2006.29.

Bell G, Hey T and Szalay A (2009) “Beyond the Data Deluge,” Science, 323(5919) 1297–1298, doi: 10.1126/science.1170411.

Galison P L (1996) Computer Simulations and the Trading Zone. In P. L. Galison & D. J. Stump (Eds.), The Disunity of Science: Boundaries, Contexts, and Power (pp. 118-157), Stanford: Stanford University Press.

Galison P L (1997) Image and Logic: A Material Culture of Microphysics. Chicago: University of Chicago Press.

Galison P and Stump D J (Eds.) (1996) The disunity of science: Boundaries, contexts, and power. Stanford University Press.

Geertz C (2000) Available Light: Anthropological Reflections on Philosophical Topics, New Jersey, Princeton University Press

Geertz (1968) Thinking as a moral act: Ethical dimensions of anthropological fieldwork in the new states. The Antioch review, [Online]. 28 (2), pp. 139.

Geertz (1982) The way we think now: Toward an ethnography of modern thought. Bulletin - American Academy of Arts and Sciences, [Online]. 35 (5), pp. 14.

Geertz (1992) Common sense as a cultural system. The Antioch review, [Online]. 50

Geertz (2002) An inconstant profession: The anthropological life in interesting times. Annual Review of Anthropology, [Online]. 31 , pp. 1.

Geertz C (1984) Distinguished Lecture: Anti Anti-Relativism. American Anthropologist, 86 (2), pp. 263-278

Gibbons M, Limoges C, Nowotny H, Trow M and Scott P (1994) The new production of knowledge : the dynamics of science and research in contemporary societies. London: Sage

Gibbons M (1999) Science's new social contract with society. Nature, 402, C81-C84.

Giere R N (1999) Using Models to Represent Reality. In L. Magnani, N. J. Nersessian & P. Thagard (Eds.), Model-Based Reasoning in Scienti"c Discovery (pp. 41-58). New York: Springer.

Goodman A A and Wong BringBringing the Night Sky Closer: Discoveries in the Data Deluge 2009) In (Eds) Hey, T., Tansley, S., & Tolle, K. (Eds.). (2009). The Fourth Paradigm: Data-Intensive Scienti"c Discovery. Redmond, WA: Microsoft Research.

Gourlay (2006) Conceptualizing knowledge creation: a critique of Nonaka's theory. The Journal of management studies, [Online]. 43 (7), pp. 1415.

Gowers T, and Nielsen M (2009) Massively collaborative mathematics. Nature, 461(7266), 879-881.

Greenwood M R C, and Riordan D G (2001) Civic scientist/civic duty. Science Communication, 23(1), 28-40.

Guba E G and Lincoln Y S (1994) Competing paradigms in qualitative research In N Hammersley M and Atkinson P (Eds.) (2007) Ethnography: Principles in Practice, London: Routledge

Haas P (1992) Introduction: Epistemic Communities and International Policy Coordination, International Organization 46 (1) 1–35

Hacking, I. (1992). The self-vindication of the laboratory sciences. Science as practice and culture, 30.

Hagendijk, R. P. (2004). The public understanding of science and public participation in regulated worlds. Minerva, 42(1), 41-59.

Hammersley M (1992) What's Wrong with Ethnography? Methodological Exploration.

Hammersley M and Atkinson P (2007) Ethnography: Principles in Practice, London: Routledge

Hartswood M, M Jirotka, R Procter, R Slack, A Voss and S. Lloyd (2005): Working IT out in e-Science: Experiences of Requirements Capture in Health Grid projects. In T. Solomnides, R. McClatchey, V. Breton, Y. Legre and S. Nørager (Eds.): Proceedings of HealthGrid 2005. From Grid to Healthgrid. Amsterdam: IOS Press.

Hasse C (2008) Learning and transition in a culture of playful physicists. European Journal of Psychology of Education, 23(2), 149-164.

Hasse C, and Trentemøller S (2011) Cultural work place patterns in Academia. Science Studies, 24(1), 6.

Hatano G (1982) Cognitive consequences of practice in culture specific procedural skills. Quarterly Newsletter of the Laboratory of Comparative Human Cognition, 4, 15-18.

Hatano, G. (2003). Forward. In A. J. Baroody & A. Dowker (Eds.), The development of arithmetic concepts and skills: Constructing adaptive expertise (pp. xi–xiii).Mahwah, NJ: Lawrence Erlbaum Associates.

Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Assume, K. Hakuta (Eds.), Child development and education in Japan (pp. 262–272). New York: Freeman.

Hatch J A (2002) Doing Qualitative Research in Education Settings, State University of New York Press

Haythornthwaite C (2006) Learning and knowledge networks in interdisciplinary collaborations. Journal of the American Society for Information Science and Technology, 57(8), 1079-1092.

HEFCE (2010) Research Excellence Framework, Research [Online] Extracted from <http://www.hefce.ac.uk/Research/ref/> On March, 01, 2010

Hemlin, (2006) The shift in academic quality control. Science, technology human values, [Online]. 31 (2), pp. 173.

Henkel M (2005) Academic identity and autonomy in a changing policy environment. Higher education, [Online]. 49 (1), pp. 155.

Hey T and Trefethen A E (2002) The UK e-Science core programme and the grid. Future Generation Computer Systems 18(8) 1017-1031.

Hey T, Tansley S and Tolle, K (Eds.) (2009) The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, WA: Microsoft Research.

Hine C (2007) Mutli-sited Ethnography as a Middle Range Methodology of Contemporary

Hine, C (2007) Journal of computer-mediated communication, [Online]. 12 (2), pp. 618.

Hine, C. (2006). Databases as scientific instruments and their role in the ordering of scientific work. Social Studies of Science, 36(2), 269-298

Holland J H (1992). Complex adaptive systems. Daedalus, 17-30.

Hughes C (2001) Shackled to stereotypes. Science and Public Affairs, 21-23

Irwin A (2006) The politics of talk. Social studies of science, 36 (2), pp. 299.

Irwin A (1995). Citizen science: a study of people, expertise, and sustainable development. Psychology Press

Irwin A (2001). Constructing the scientific citizen: science and democracy in the biosciences. Public understanding of science, 10(1), 1-18.

Irwin A (2006) The Politics of Talk Coming to Terms with the 'New'Scientific Governance. Social studies of science, 36(2), 299-320.

Irwin A (2014) From deficit to democracy (re-visited). Public Understanding of Science, 23(1), 71-76.

Jasanoff S (2004) States of Knowledge : The Co-Production of Science and Social Order. New York: Routledge.

Jasanoff S (2003a) 'Technologies of humility: Citizen participation in governing science', Minerva 41, 223–244

Jasanoff S (2003b) Accounting for expertise. Science and Public Policy 30 (3), 157–162

Jasanoff S (2014). A mirror for science. Public Understanding of Science, 23(1), 21-26.

Jirotka M, Lee C P and Olson G M (2013). Supporting Scientific Collaboration: Methods, Tools and Concepts. Computer Supported Cooperative Work (CSCW), 22(4-6), 667-715

Jirotka M, Procter R, Hartswood M, Slack R, Simpson A, Coopmans C, Hinds C, and Voss A (2005) Collaboration and trust in healthcare innovation: The eDiaMoND case study. Computer Supported Cooperative Work (CSCW), 14(4), 369-398.

K Denzin and Y K Lincoln (Eds), Handbook of Qualitative Research. Thousand Oaks

Kelling S, Gerbracht J, Fink D, Lagoze C, Wong W, Yu J, Damoulas, T, and Gomes C (2012). Ebird: A Human/ Computer Learning Network for Biodiversity Conservation and Research. Proceedings of the 24th Conference on Innovative Applications of Artificial Intelligence.

Kerr A, Cunningham-Burley S and Tutton R (2007) Shifting Subject Positions: Experts and Lay People in Public Dialogue. Social Studies of Science, 37 (3), pp. 385-411.

Knorr-Cetina K (1999) Epistemic cultures: How the sciences make knowledge. Cambridge,

Knorr-Cetina K (2001) Objectual Practice In Theodore R. Schatzki,Karin Knorr Cetina and Eike von Savigny (Eds.), The Practice Turn in Contemporary Theory. London and New York: Routledge, 175-188

Knorr-Cetina K (1982) Scientific communities or transepistemic arenas of research? A critique of quasi-economic models of science. Social studies of science, [Online]. 12 (1), pp. 101.

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 Reviews, 32(4), 361-375.

Korpela E, Werthimer D, Anderson D, Cobb J and Lebofsky, M (2001) SETI@home-Massively Distributed Computing for SETI, Computing in Science and Engineering 3(1) 78-83

Kowalczyk, S., and Shankar, K. (2011). Data sharing in the sciences. Annual review of information science and technology, 45(1), 247-294.

Kramer J (2007) Is abstraction the key to computing?. Communications of the ACM, 50(4), 36-42

Kraut R, Olson J, Banaji M, Bruckmann A, Cohen J and Couper M. (2004) Psychological research online: report of the board of scientific affairs' advisory group on the conduct of research on the internet. American Psychologist, 59(2): 105–117.

Kuhn T S (1996) The structure of scientific revolutions. 3rd ed. Chicago ; London: University of Chicago Press

Kvale S (1996) Interviews: an introduction to qualitative research interviewing. Sage:

Lakatos, I., and Musgrave, A. (1970). Criticism and the Growth of Knowledge. Cambridge: Cambridge University Press.

Lamb R, and Kling R (2003). Reconceptualizing users as social actors in information systems research. MIS quarterly, 197-236.

Latour B (2000) When things strike back: a possible contribution of' science studies' to the social sciences. The British journal of sociology, [Online]. 51 (1), pp. 107.

Latour B and S Woolgar (1979) Laboratory Life: The Social Construction of Scientific Facts. Beverly Hills: Sage

Latour B (1983). Give Me a Laboratory and I Will Raise the Earth. In K. Knorr-Cetina & M. Mulkay (Eds.), Science Observed (pp. 141-170). London: Sage Publications

Latour B (2004). Politics of Nature: How to Bring the Sciences into Democracy. Cambridge, MA: Harvard University Press.

Latour B (2005). Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford: Oxford University Press

Latour B and Woolgar S (1979). Laboratory Life: The Social Construction of Scientific Facts. London: Sage Publications.

Laws S, Harper C and Marcus R (2003) Research for Development: A Practical Guide. London: Sage

Leander K M, and McKim K K (2003). Tracing the everyday'sitings' of adolescents on the internet: A strategic adaptation of ethnography across online and offline spaces. Education, Communication & Information, 3(2), 211-240.

Lessig L (1999) The law of the horse: What cyberlaw might teach. Harvard law review 501-549. Chicago

Leydesdorff L (2003) The Triple Helix of university-industry-government relations. Scientometrics [Online]. 58 (2) pp. 191.

Leydesdorff L (2005) Science shops: A kaleidoscope of science-society collaborations in Europe. Public understanding of science 14 (4) pp. 353

Leydesdorff L (2009) The non-linear dynamics of meaning processing in social systems. Social science information 48 (1) pp. 5.

Liberman S and Wolf K B (1997) The flow of knowledge: Scientific contacts in formal meetings. Social Networks 19 (3) pp. 271-283.

Lindblom C E (1979). Still muddling not yet through. Public administration review 517-526.

Lintott C and Land K (2008) Eyeballing the universe Physicsworld http://physicsworld.com/cws/article/print/35626

Malone T W Laubacher R and Dellarocas C (2009) Harnessing Crowds: Mapping the Genome of Collective Intelligence. MIT Center for Collective Intelligence

March J G, Olsen J P & Christensen S (1976). Ambiguity and choice in organizations (pp. 69-81). Bergen: Universitetsforlaget.

Marcus G E (1995) Ethnography in/of the world system: the emergence of multi-sited ethnography Annual Review of Anthropology 24 95-11

Marcus G E (1995). Ethnography in/of the world system: the emergence of multi-sited ethnography. Annual review of anthropology 95-117.

Mazawi A E (2011). “Which Palestine Should We Teach?” Signatures Palimpsests and Struggles over School Textbooks. Studies in Philosophy and Education 30(2) 169-183.

McAdam R (2000) A critique of knowledge management: using a social constructionist model. New technology work and employment [Online]. 15 (2) pp. 155.

Merton R K (1972) Insiders and Outsiders: A Chapter in the Sociology of Knowledge. The American Journal of Sociology 78 (1 Varieties of Political Expression in Sociology) pp. 9-47.

Merz M (2006) Embedding digital infrastructure in epistemic culture. Hine CM New Infrastructure for Knowledge Production: Understanding E-Science. Hershey PA: Information Science Publishing. p 99-119. Chicago

Merz M and Knorr-Cetina K (1997). Deconstruction in a thinking 'science: Theoretical physicists at work. Social Studies of Science 27(1) 73-111.

Meyer E T, Bulger M, Kyriakidou-Zacharoudiou A, Power L, Williams P, Venters W, Terras M and Wyatt S (2011). Collaborative yet independent: Information practices in the physical sciences. Research Information Network (RIN) Report Series.

Miles M B and Huberman A M (1994) Qualitative data analysis: An expanded sourcebook London: Sage

Miller J (2004) Public understanding of and attitudes toward scientific research: What we know and what we need to know. Public understanding of science [Online] 13 (3) 273-294

Miller S (2001) Public understanding of science at the crossroads. Public understanding of science , 10 (1) pp. 115.

Miller S (2004) Public understanding of and attitudes toward scientific research: What we know and what we need to know. Public understanding of science 13 (3) pp. 273.

Miller J D (1983) “Scientific Literacy: a Conceptual and Empirical Review ” Daedalus 112(2): 29–48.

Miller T P (2003). Changing the subject. The Realms of Rhetoric: The Prospects for Rhetorical Education

Mortensen T and Walker J (2002) Blogging thoughts: personal publication as an online research tool. Researching ICTs in context. Intermedia. Oslo Norway University of Oslo.

Mulkay (1974) Methodology in the sociology of science: Some reflections on the study of radio astronomy, Social Science Information, 13 (2) 107-119

Mulkay M J (1975) Problem areas and research networks in science. Sociology, 9 (2) pp. 187.

Mulkay M J (1982) Accounting for error: how scientists construct their social world when they account for correct and incorrect belief. Sociology , 16 (2) pp. 165.

Mulkay M and Gilbert G N (1982) What is the Ultimate Question? Some Remarks in Defence of the Analysis of Scientific Discourse. Social Studies of Science , 12 (2) pp. 309-319.

Narayanan A. & Shmatikov V. (2008 May). Robust de-anonymization of large sparse datasets. In Security and Privacy 2008. SP 2008. IEEE Symposium on (pp. 111-125). IEEE.

National Science Foundation (2004) Science and Technology: Public Attitudes andUnderstanding [Online] Extracted from <http://www.nsf.gov/statistics/seind04/c7/c7i.htm> on March 01 2010

Nerland M and Jensen K (2010). Objectual practice and learning in professional work. In Learning Through Practice (pp. 82-103). Springer Netherlands.

Nielsen M A (2012) Reinventing Discovery: The New Era of Networked Science. Princeton NJ: Princeton University Press.

Nonaka I, Toyama R and Konno N (2000). SECI Ba and Leadership: a Unified Model of Dynamic Knowledge Creation. Long range planning 33(1) 5-34.

Nowotny H (2003) Democratising expertise and socially robust knowledge, Science and Public Policy, 30 (3) 151-156

Nowotny H (1999) The place of people in our knowledge. European Review , 7 pp. 247.

Nowotny H (2000). Transgressive competence the narrative of expertise. European Journal of Social Theory ,3(1) 5-21. Chicago

Nowotny H (2005) The Increase of Complexity and its Reduction Emergent Interfaces between the Natural Sciences Humanities and Social Sciences. Theory Culture & Society 22(5) 15-31.

Nowotny H (2014) Engaging with the political imaginaries of science: Near misses and future targets. Public Understanding of Science 23(1) 16-20.

Olson G M, Olson J S and Venolia G (2009). What Still Matters About Distance. Proceedings of the Human Computer Interaction Consortium.

Woolgar S and Coopmans (2006) Virtual Witnessing in a Virtual Age: A Prospectus for Social. New infrastructures for knowledge production: Understanding e-Science, 1

Paulus T Woodside M and Ziegler M (2008) Extending the Conversation: Qualitative Research as Dialogic Collaborative Process. The Qualitative report ,13 (2) 226- 243

Pea R (1997) Practices of distributed intelligence and designs for education In Salomon G (Ed.) Distributed cognitions - psychological and educational considerations. Cambridge: Cambridge University Press. 47-8 research. Thousand Oaks CA; London: Sage

Pentland B T (1995) Information systems and organizational learning: the social epistemology of organizational knowledge systems. Accounting Management and Information Technologies 5(1) 1-21.

Pickering A (1992). From science as knowledge to science as practice.Science as practice and culture 1-26.

Pfister H and Oehl M (2009) The impact of goal focus, task type and group size on synchronousnet-based collaborative learning discourses. Journal of Computer Assisted Learning, 25, 161–176.

Poliakoff E and Webb T L (2007. What factors predict scientists' intentions to participate in public engagement of science activities? Science communication, 29(2) 242-263.

Powell M C & Colin M.(2008). Meaningful Citizen Engagement in Science and Technology What Would it Really Take?. Science Communication , 30(1) 126-136.

Punch K F (1998) Introduction to social research: Quantitative and qualitative approaches. Thousand Oaks CA: Sage

Ribes and Jackson 2013; Ribes D. and Jackson S. J. (2013). Data Bite Man: The Work of Sustaining Long-Term Data Collection. In L. Gitelman (Ed.) Raw Data Is an Oxymoron. Cambridge MA: MIT Press.

Riesch H and Potter C (2013). Citizen science as seen by scientists: Methodological epistemological and ethical dimensions. Public Understanding of Science 0963662513497324.

Robson C (1993) Real world research. Cambridge: Blackwell

Rooker T (2001) The Pragmatism of Knowledge: Quantum Mechanics. Bulletin of science technology & society , 21 (2) 119- 128.

Rosner D K (2012 February). The material practices of collaboration. In Proceedings of the ACM 2012 conference on computer supported cooperative work (pp. 1155-1164). ACM

Rossman G B and Rallis S F (1998) Learning in the field: An introduction to qualitative research. Thousand Oaks CA; London: Sage

Rowe G and Frewer L (2005) A typology of public engagement mechanisms Science Technology and Human Values (30) pp251–290

Rowe G and Frewer L J (2004). Evaluating public-participation exercises: a research agenda. Science technology & human values 29(4) 512-556.

Royal Society (2006) Survey of Factors Affecting Science Communication by Scientists and Engineers Final Report London

Rubin H J and Rubin I S (1995) Qualitative interviewing: The art of hearing data. Thousand Oaks

Rupp A A , Gushta M, Mislevy R J and Shaffer D W (2010).Evidence-centered design of epistemic games: Measurement principles for complex learning environments. The Journal of Technology Learning and Assessment 8(4).

Samwald M, Jentzsch A, Bouton C, Kallesøe C S, Willighagen E Hajagos J, Marshall M S and Stephens S (2011). Linked open drug data for pharmaceutical research and development. Journal of cheminformatics 3(1) 19.

Schatzki T R, Knorr-Cetina K and von Savigny E (Eds.). (2001). The practice turn in contemporary theory. Psychology Press.

Schroeder R & Bailenson J (2008) “Research uses of multi-user virtual environments”. In Handbook of online research methods Edited by: Blank G. Lee R. and Fielding N. 327–342. London: Sage.

Scott P (1997) The changing role of the university in the production of new knowledge. Tertiary Education and Management, 3 (1) pp. 5.

Scribner S and Cole M (1981) The psychology of literacy. Cambridge MA :Harvard University Press

Seale C (1999) Quality in Qualitative Research Qualitative Inquiry Sage: 5 (4)

Schwartz D L (1995) The emergence of abstract representations in dyad problem solving.Journal of the Learning Sciences, 4, 321–354

Shaffer D W (2006). Epistemic frames for epistemic games. Computers & education 46(3) 223-234.

Shapin S(1995) Here and everywhere: Sociology of scientific knowledge. Annual Review of Sociology , 21 (1) pp. 289.

Shapin S and Scha!er S (1985) Leviathan and the Air-Pump: Hobbes Boyle and the Experimental Life. Princeton NJ: Princeton University Press.

Shilton K (2009) Four Billion Little Brothers? Privacy Mobile Phones and Ubiquitous Data Collection. Communications of the ACM 52(11) 48-53

Shirky C (2009) Here Comes Everybody: The Power of Organizing without Organizations. New York: Penguin Group USA.

Shirky C (2010). Cognitive Surplus: Creativity and Generosity in a Connected Age. New York: Penguin Press.

Simeone L (2013) The potential and shortcomings of strategic ambiguity as management practice in design labs: An ethnographic study of MIT SENSEable City Lab.

Simeone L & Ratti C (2011) Transdisciplinarity in action: An ethnographic study of MIT SENSEable City Lab’s organizational culture. Integral Leadership Review.

Smith A M Lynn S Sullivan M Lintott C J Nugent P E Botyanszki J and Walters R. (2011). Galaxy Zoo supernovae. Monthly Notices of the Royal Astronomical Society 412(2) 1309-1319.

Star S L and Ruhleder K (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. Information systems research 7(1) 111-134.

Stehr N (1994) Knowledge societies. London: Sage Publications.

Sternberg R J (1999) Intelligence as developing expertise. Contemporary educational psychology, 24(4) 359-375.

Sundberg M (2010a) Creating Convincing Simulations in Astrophysics. Science Technology & Human Values , 37(1) 64-8

Sundberg M. (2010b) Cultures of Simulations Vs. Cultures of Calculations? The Development of Simulation Practices in Meteorology and Astrophysics. Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics 41(3) 273-281.

Sundberg M (2011) The Dynamics of Coordinated Comparisons: How Simulationists in Astrophysics Oceanography and Meteorology Create Standards for Results. Social Studies of Science, 41(1) 107-125.

Sweeney L (2002) k-anonymity: A model for protecting privacy. International Journal of Uncertainty Fuzziness and Knowledge-Based Systems , 10(05) 557-570.

Thomas G and Durant J (1987) Why should we promote the public understanding of science. Scientific literacy papers, 1 1-14.

Tsoukas H (2005) How is New Knowledge Created in Organizations? A Dialogical University of Michigan Press

Tsoukas H and Hatch M J (2001) Complex thinking complex practice: The case for a narrative approach to organizational complexity. Human Relations 54(8) 979-1013.

Van Leeuwen T and Tijssen R (2000) Interdisciplinary dynamics of modern science: analysis of cross-disciplinary citation flows. Research Evaluation 9(3) 183-187.

Vann K and Bowker G C (2006) Interest in Production: On the Configuration of. New infrastructures for knowledge production: Understanding E-science 71

Vertesi J (2012) Seeing like a Rover: Visualization embodiment and interaction on the Mars Exploration Rover Mission. Social Studies of Science, 42(3) 393-414

Way M J, Scargle J D, Ali K M and Srivastava A N (Eds.). (2012). Advances in machine learning and data mining for astronomy. CRC Press.

Weick K E (2005) Organizing and the process of sensemaking. Organization Science [Online]. 16 (4) pp. 409.

Wellcome Trust (2000). The role of scientists in public debate. Full report. London

Whitley R D (1972) Black boxism and the sociology of science: a discussion of the major developments in the field. Pp. 61-92 in Paul Halmos (ed.) The Sociology of Science. The Sociological Review Monograph 18.

Woolley A W, Chabris C F, Pentland A, Hashmi N, and Malone T W (2010). Evidence for a collective intelligence factor in the performance of human groups. Science, 330(6004), 686-688.

Wouters P and Beaulieu A (2006). Imagining e-science beyond computation. New infrastructures for knowledge production: Understanding e-science 48-70.

Wuchty S (2007) The increasing dominance of teams in production of knowledge. Science 316 (5827) pp. 1036.

Wynne B (1996) May the Sheep Safely Graze? A Reflexive View of the Expert-Lay Knowledge Divide' in Scott Lash Bronislaw Szerszynski and Brian Wynne (eds) Risk Environment & Modernity: Towards a New Ecology London: Sage

Wynne B (2006). Public engagement as a means of restoring public trust in science–hitting the notes but missing the music?. Public Health Genomics 9(3) 211-220.

Zeilik M (1999) Conceptual astronomy. II. Replicating conceptual gains probing attitude changes across three semesters. American journal of physics [Online]. 67 pp. 923.

Ziman J (2001) Getting scientists to think about what they are doing. Science and engineering ethics , 7 (2) pp. 165.

Zimmerman A, Bos N, Judith S, Olson and Olson G M (2009).The promise of data in e-Research: Many challenges multiple solutions diverse outcomes. In N. Jankowski (ed): e-Research: Transformation in Scholarly Practice Routledge pp. 222–239.

Zittrain, J. (2009). Work the new digital sweatshops. Newsweek, December, 8

11. Appendices

Appendix A

INTERVIEW AND MEETING SCHEDULE WITH GALAXY ZOO TEAM MEMBERS

S no.	Dates	Location	Informant	Meeting Description
1	26.09.09	Oxford	PI 0	First contact made with the PI
2	8.02.10		PI1	Research approved by PI on behalf of the GZ Team
3	12.02.10		PI 2	
4	19.02.10		PI 3	
5	24.02.10		PI4	
6	01.03.10		PI 5	
7	11.03.10		PI 6	
	12.03.10		GZ1	
8	15.03.10		PI 7	
9	17.03.10		TL 1	
10	24.03.10		PI 8	
11	27.03.10		PI 9	
12	31.03.10		TL 2	Pre-NSF Zooniverse Advisory meeting
13	21.04.10	Chicago	GZ 2	Informal Meeting with the Coordinator of NSF Meetings
14	22.04.10		GZ 3	EVENT: 1st NSF Zooniverse Advisory Committee Meeting

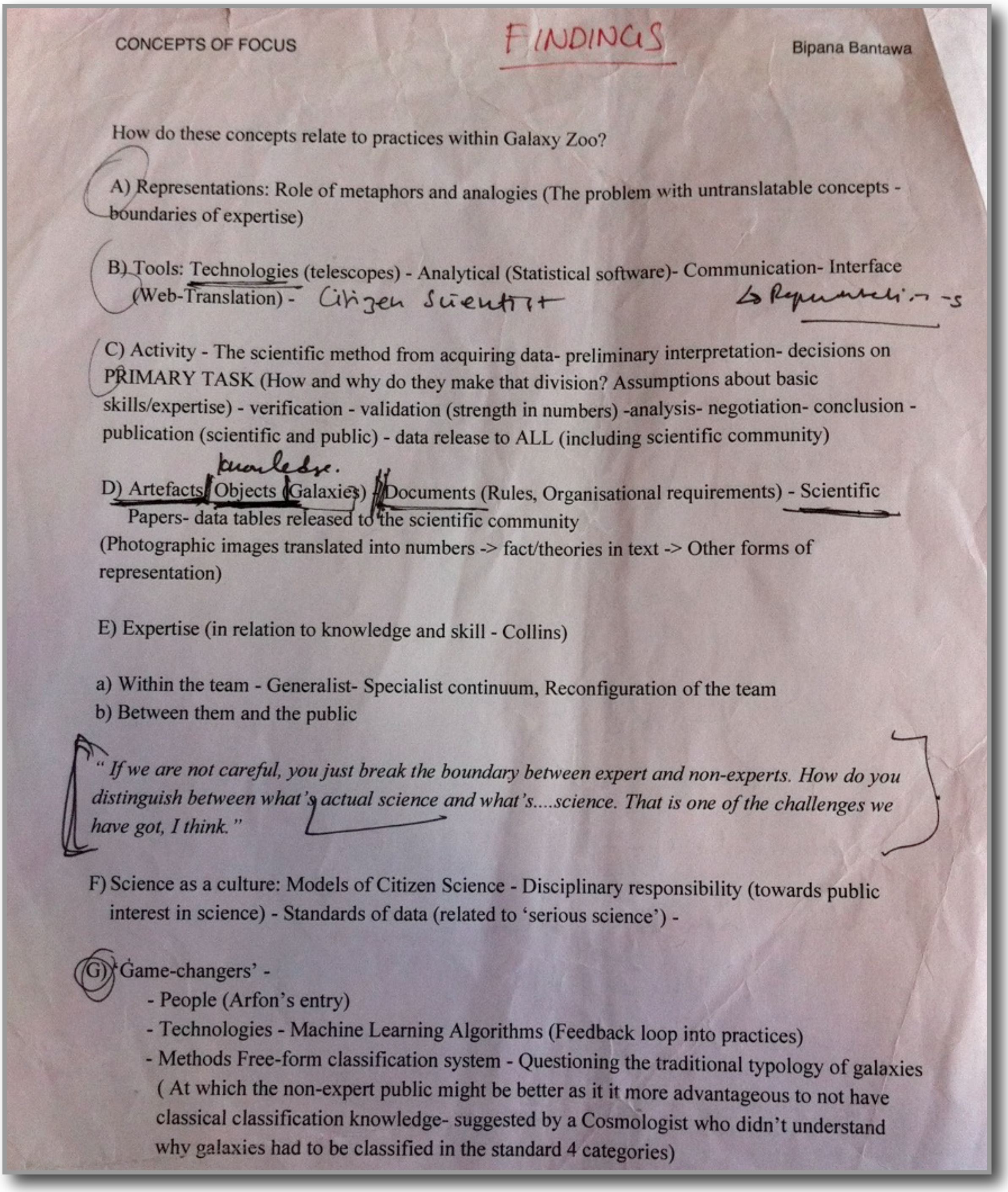
15	23.04.10			
16	4.05.10		PI 10	
17	28.05.10		TL3	Post NSF Zooniverse Meetings
18	4.06.10		PI 11	Post NSF Zooniverse Meetings
19	15.06.10		PI 12	
20	22.06.10		PI 13	
21	1.07.10	Oxford	PI 14	
22	4.07.10		PI 15	
23	11.08.10		TL 4	
24	23.10.10		PI 16	
25	25.10.10		PI 17	
26	15.12.10		TL 5	
27	17.12.10		PI 18	
28	17.12.10		GZ 4	
29	17.12.10	Chicago	GZ 5	
30	20.12.10		PI 19	
31	21.12.10		PI 20	
32	16.01.11	Oxford	AS 6	
33	16.02.11		AS 7	
34	4.04.11		GZ 6	EVENT: dotAstronomy, Oxford
35	8.04.11		PI 21	
36	18.04.11	Oxford	TL 8	
37	13.06.11		TL 9	
38	17.06.11		TL 10	

39	17.12.11		TL 11	
40	21.12.11		PI 22	
41	5.01.12		TL 12	
42	6.07.12		PI 23	
43	4.10.12		TL 13	
44	5.04.13		PI 24	Follow up and Closing Meeting

NOTE: Between the 23rd and the 24th of April, 2010, I had the opportunity to have informal conversations with some of the GZ Science and Education Team members as result of attending the meetings and being the only participant from Oxford and being asked to monitor the video conferencing technologies during the meeting.

Appendix B

Analytical Notes Taken During Fieldwork



CONCEPTS OF FOCUS FINDINGS Bipana Bantawa

How do these concepts relate to practices within Galaxy Zoo?

A) Representations: Role of metaphors and analogies (The problem with untranslatable concepts - boundaries of expertise)

B) Tools: Technologies (telescopes) - Analytical (Statistical software)- Communication- Interface (Web-Translation) - Citizen Scientist

C) Activity - The scientific method from acquiring data- preliminary interpretation- decisions on PRIMARY TASK (How and why do they make that division? Assumptions about basic skills/expertise) - verification - validation (strength in numbers) -analysis- negotiation- conclusion - publication (scientific and public) - data release to ALL (including scientific community)

D) Artefacts/ Objects (Galaxies) /Documents (Rules, Organisational requirements) - Scientific Papers- data tables released to the scientific community (Photographic images translated into numbers -> fact/theories in text -> Other forms of representation)

E) Expertise (in relation to knowledge and skill - Collins)

a) Within the team - Generalist- Specialist continuum, Reconfiguration of the team
b) Between them and the public

" If we are not careful, you just break the boundary between expert and non-experts. How do you distinguish between what's actual science and what's....science. That is one of the challenges we have got, I think."

F) Science as a culture: Models of Citizen Science - Disciplinary responsibility (towards public interest in science) - Standards of data (related to 'serious science') -

G) 'Game-changers' -
- People (Arfon's entry)
- Technologies - Machine Learning Algorithms (Feedback loop into practices)
- Methods Free-form classification system - Questioning the traditional typology of galaxies (At which the non-expert public might be better as it it more advantageous to not have classical classification knowledge- suggested by a Cosmologist who didn't understand why galaxies had to be classified in the standard 4 categories)

Exhibit A: An example of periodic analytical note taking during fieldwork

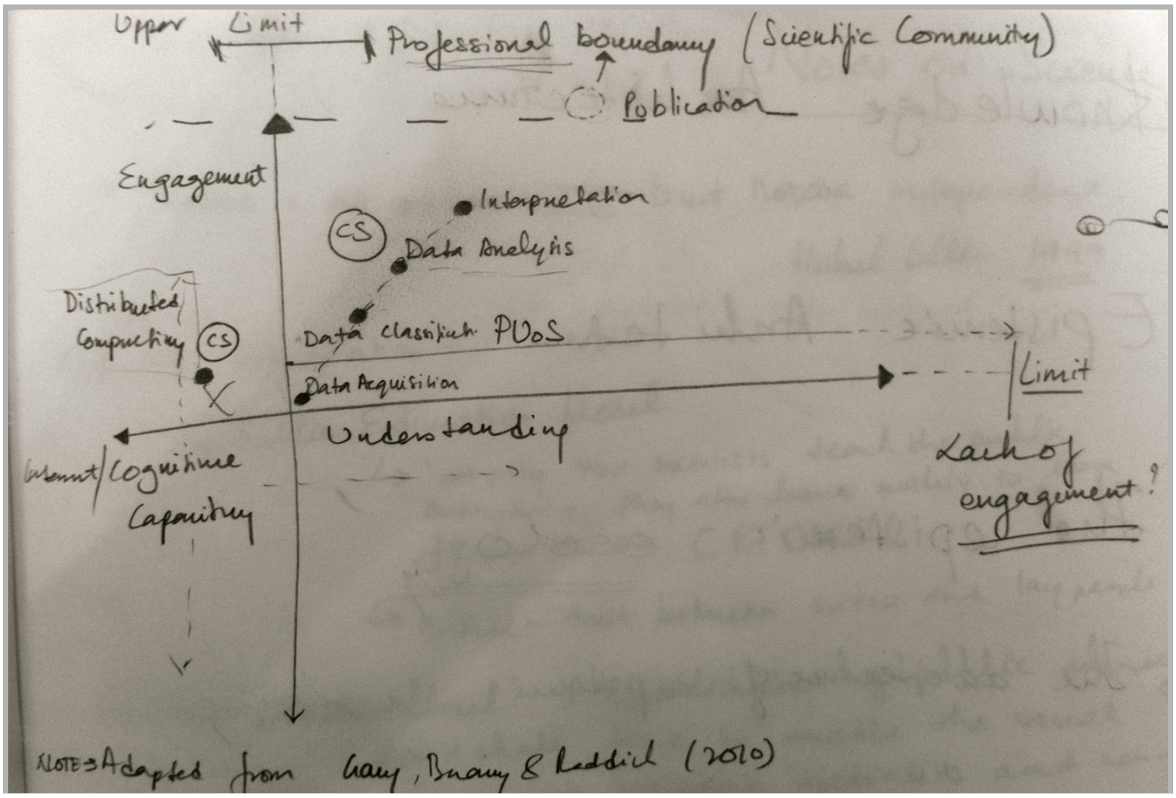


Exhibit B: Conceptualising The Role of Volunteer Expertise, Based on Reports Drafted by the Galaxy Zoo Education Team

Appendix C

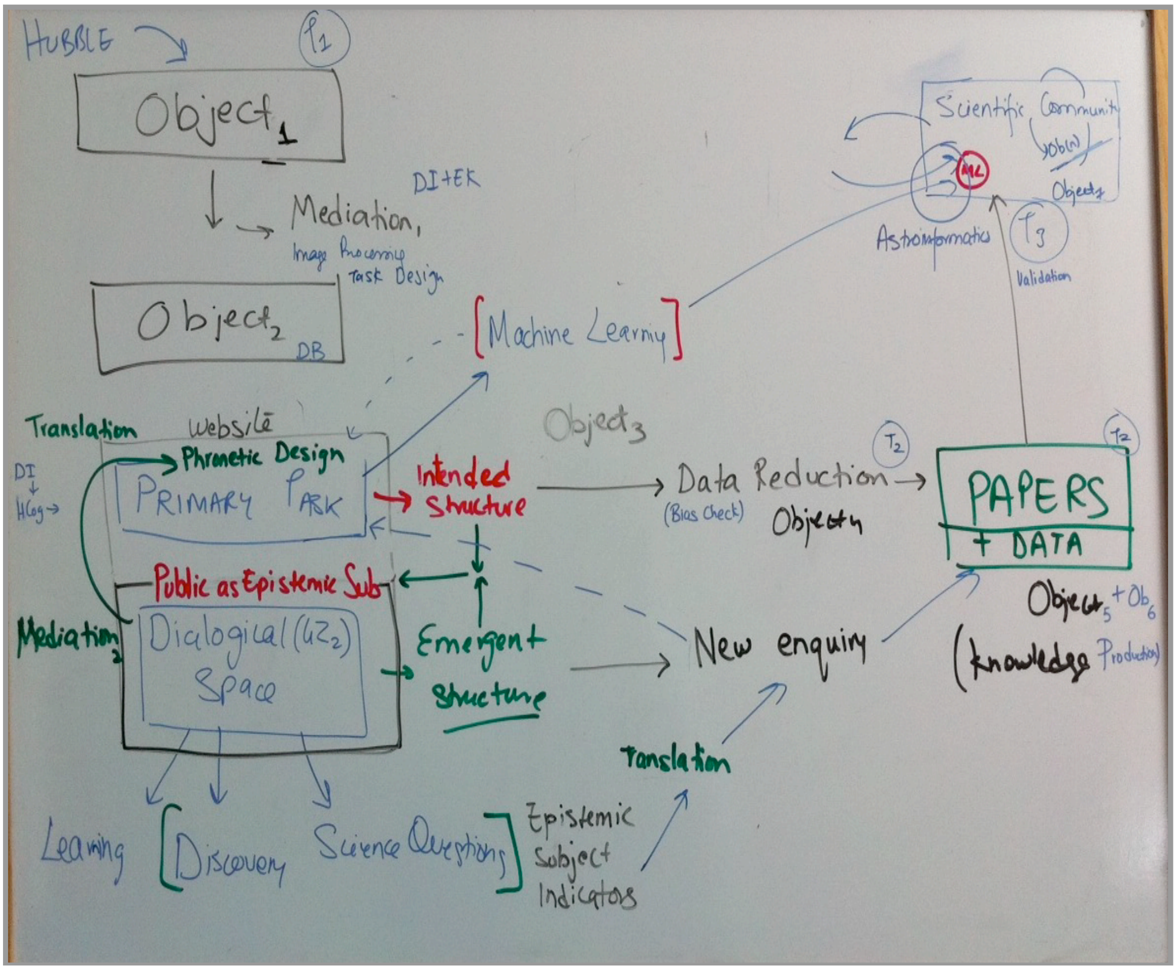
Events Attended by the Researcher That Helped Shape the Research

Event	Date	Notes
Talk: The Rise and Rise of Citizen Science: Galaxy Zoo and Beyond	26th September 2009	Established contact and requested for a meeting with research proposal
Approval from GZ Team for research	February 2010	
National Science Foundation Eager Workshop Web Methodologies Workshop by Invitation	7-8 April, 2010 RPI- NY	Presented preliminary and talk titled: Transdisciplinary Methodologies in Researching Citizen Science Initiatives
NSF Zooniverse Advisory Committee Meeting (Data gathering)	22-23rd April, Adler Planetarium Chicago	Participant Observer
Society for Social Studies of Science (4S) Conference, Tokyo	25th-28th Aug, 2010, Tokyo	Presentation titled: Methodological considerations in researching a citizen science initiative
"Working with the crowd : 21st century citizen science", by GZ's Principal Investigator	25th October, 2010 Martin School, Oxford	Talk by PI
Computational frontiers in scientific discovery	4th November, 2010 Royal Society	Seminar Attended
Lecture: Professor Alexander Szalay, 'Extreme Data Intensive Computing in Astrophysics'	21st January, 2011 Oxford Physics	Talk by speaker who was initially involved with GZ
A New Mandate: 21st Century Challenges for Research Policy, Royal Geographical Society	24th March, 2011	Workshop
Visualisation in the Age of Computerisation Conference, Oxford	25th-26th March, 2011	
dotAstronomy, New College, Oxford	4th-6th April, 2011 Oxford	GZ Team Organised event

Collective Intelligence Conference, MIT	April 18-20, 2012	Developed and refined the framework of Collective Intelligence for this study
Rigour and Openness in 21st Century Science, Oxford	April 11th-12th, 2012	
Science Online London 2011	Sepetember 2-3, 2011	
Science Online co hosted by the Nature Network, The Royal Institution and Mendeley	22 August 2009	

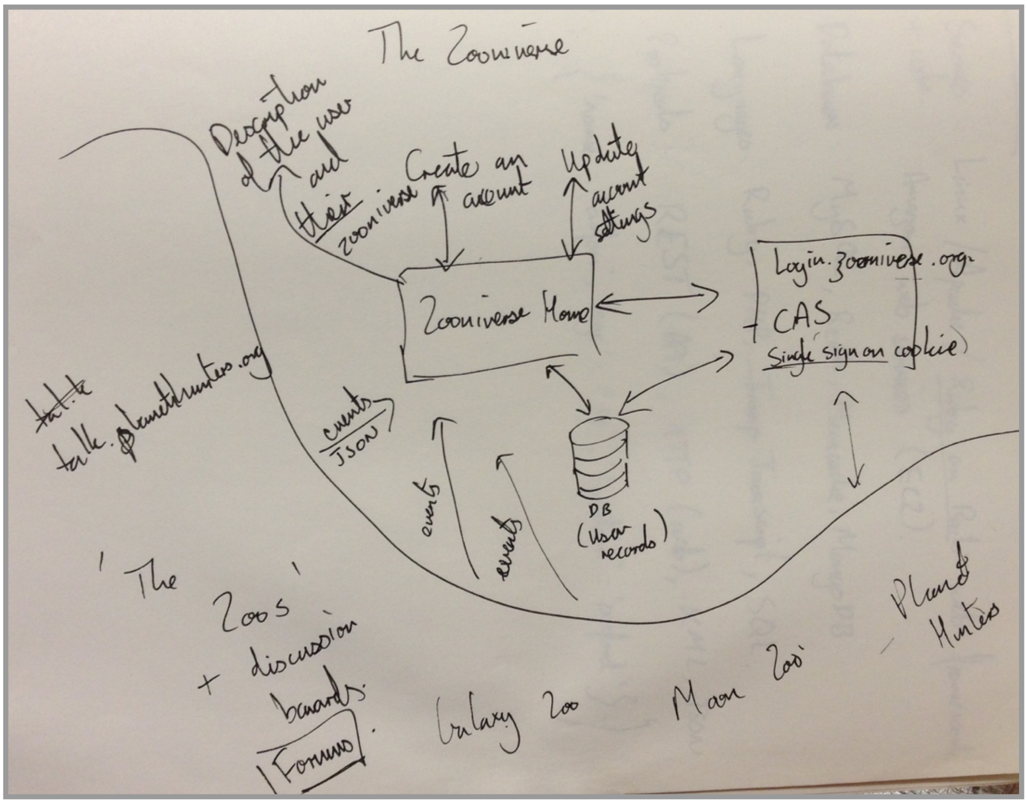
APPENDIX D

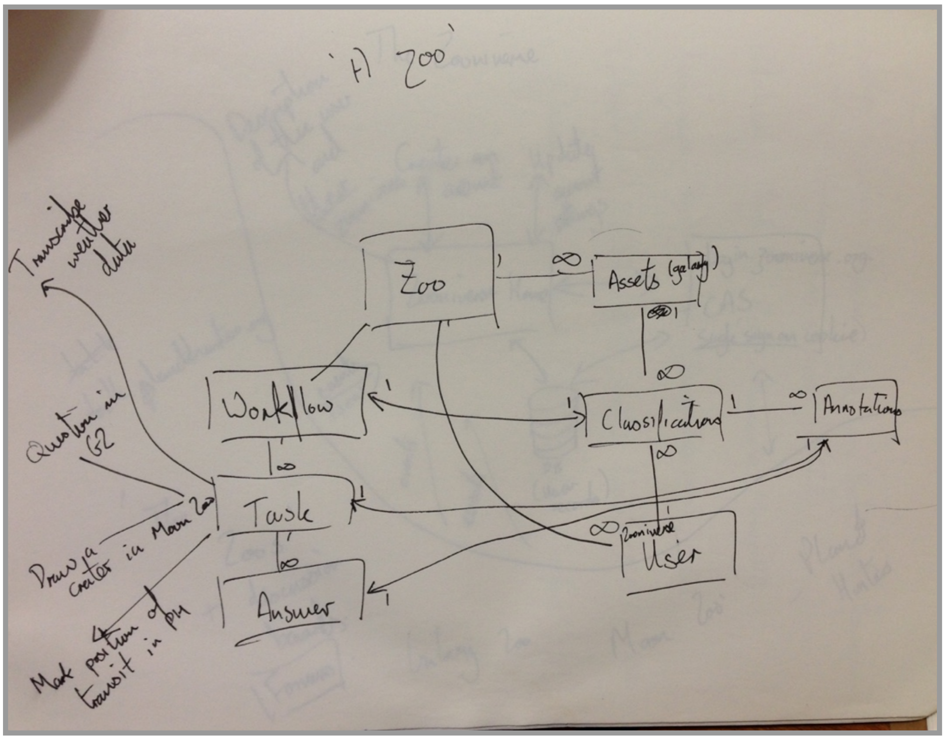
Identifying the Stages of Translation - Work in Progress (July-August, 2011)



Appendix E

Galaxy Zoo's Technical Architecture Conceptualised by the Technical Lead





Appendix F

Zooniverse's Ground Rules for Citizen Science by The Technical Lead

Exhibit A: Twitter post by GZ's Technical Lead

Arfon Smith (@arfon) on **September 9, 2011 at 2:55 pm** said:

Alice, as always thanks for the good read.

I wanted to follow up on your idea of the 'ethics of citizen science'. At the Zooniverse we have three simple rules that we test new ideas against that I thought might be useful to share:

1. Our community are collaborators not users
e always credit the contribution of our community to the research and when possible include individual community members as authors.

2. Citizen science should contribute to real research
I'm not sure that this should be a hard and fast rule for every citizen science project out there but for Zooniverse this is key. The primary goal of all of our projects is to deliver real academic research (i.e. papers) even though they may also be excellent tools for communicating science and scientific method.

3. Don't waste people's time
This sounds obvious but I think it's worth stating. If you could achieve equally good results by other means (for example by using an automated algorithm) then you shouldn't be asking people to do it.

These three rules are our 'contract' with our community. They're far from perfect but hopefully they give a good idea of what we thing some of the ground rules are for an 'ethical' citizen science endeavour.

Reply ↓

Exhibit B: Technical Lead's post on a public blog

Appendix G

A Blog Post on the official Zooniverse Blog by Galaxy Zoo's Technical Lead

In Defense of Zooniverse Design
March 12, 2013

Zooniverse The Zooniverse Blog
by arfon

Why SciStarter.com is Bad For Citizen Science

Preface: I'd like to begin by saying that I've met Darlene Cavalier at conferences in the past and I'm a big supporter of her efforts. Darlene is truly is a 'cheerleader' for citizen science, her enthusiasm is infectious and the citizen science domain is clearly a better place with her. I'm writing here about what I consider the bad practice of SciStarter.com and Science For Citizens LLC, their parent organisation. I have no idea whether the issues highlighted here are because of decisions that she has made. There was a time not so long ago when you needed a new account for pretty much everything you tried out on the web. Want to upload photos to Flickr? Then signup for a Yahoo! ID. Want a blog? Then give WordPress or Tumblr your details. Feeling social? Then FaceBook, Twitter or MySpace would pretty much want the same information. These days there are a number of solutions that allow you to log in to web-based services using things like your Facebook, Twitter or Google account. Under the hood these solutions typically rely on a couple of protocols such as OAuth and OpenID and often still request your email address when you sign in but the days of hundreds of accounts each with their own password to remember are coming to a close.

In many ways a request by an organision for your email address when signing up for a new service is completely reasonable. In exchange for handing over your email address and a few personal details these tools were often available for free – both parties win. There is of course the discussion around who or what is the product when you use these free services but let's not go into that here. Since launching the original Galaxy Zoo back in 2007 we've encouraged our volunteer community to register for an account with us, although for the vast majority of our projects (and all of our recent ones) this login/signup is an optional step. For the Zooniverse there are two main reasons for asking you to create an account:

1) When we publish a paper as a result of your efforts we feel extremely strongly about crediting you for your efforts. Experience has taught us that attempting to publish a paper with 170,000 authors on is somewhat frowned upon by the journals but if you take a look at any of the Zooniverse publications (http:/\www.zooniverse.org/publications) you'll find a link to an authors page such as here (http://authors.galaxyzoo.org), here (http://www.planethunters.org/authors) and here(http://www.milkywayproject.org/authors). We can only credit you if you share some personal information with us when you sign up.

For our research methods to work well, identifying an individual 'classifier' is pretty important. You can read more about this here (http://adsabs.harvard.edu/abs/2008MNRAS.389.1179L) (the original Galaxy Zoo paper) or here (http://adsabs.harvard.edu/abs/2012MNRAS.424.2442S) but in order to produce the best results possible we spend lots of time working out who is 'best' at a particular task and weighting their contributions accordingly. Being able to reliably identify an individual throughout the lifetime of a project (and even between projects) is most simple when someone has logged in. Over the past year or so I've become increasingly concerned by the behaviour of SciStarter.com – a website that indexes citizen science projects from across the web. The site does a pretty good job of cataloging citizen science projects you can contribute to – when you visit the site and search for example for 'bats' the Zooniverse project Bat Detective (http://www.batdetective.org/) is listed in the results. Selecting the result takes you to a brief summary of Bat Detective and offers you a link to 'get started now!' and this is where it goes wrong: Rather than taking you straight to the Bat Detective site you have to be 'logged in'. Sign up for what exactly? Am I signing up to take part in Bat Detective? No. You're actually just signing up for an account with SciStarter.com just so you can get a link to a project that SciStarter.com has nothing to do with.

Additionally, in a recent 'top 10′ blog (http://scistarter.com/blog/2012/12/top-12 citizenscience-projects-of-2012/) post of most successful citizen science projects of 2012, Bat Detective was highlighted. Did the link in this article send you straight to the Bat Detective website? Sadly not, it of course links to SciStarter's catalogue page about Bat Detective which requires account registration before you can access the URL.

To me this doesn't seem right and in many ways this is just exploiting people's lack of experience and understanding of the web. There's a reason that Facebook.com is in the consistently the most Googled terms – many people just don't quite understand how the web works and I think SciStarter.com are exploiting this. Conversly, for those who are a little more web savvy these tactics must seem very clumsy. Perhaps more importantly though, it's widely recognised that signup forms are a barrier to entry for many people and so by having people jump through this hoop SciStarter.com are actually holding potential citizen scientists back. I don't believe it's in anyone's interest other than Scistarter's to require you to sign up to follow a link through to a project. By mandating this step they are building an index of individuals interested in other people's projects when they don't have any of their own and they're risking confusing new community volunteers about what they have and haven't signed up for. All of this is made worse by the fact that SciStarter.com is a division of Science for Citizens LLC – a commercial company. So my challenge to SciStarter.com is this: If you're so committed to citizen science then why put up this artificial barrier to contribution? Crawling the internet for people's emails is one of the less tasteful aspects of the web and one I'd hoped we'd seen the end of. So how about it SciStarter?

Source: blog.zooniverse.org/2013/03/12/why-scistarter-com-is-bad-for-citizen-science/ 3/3