

DIGITAL ERA LOCAL GOVERNMENT IN ENGLAND

SERVICE REFORM AND THE INTERNET

Tom Nicholls
Green Templeton College
University of Oxford

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

85,163 words

3 November 2016

ABSTRACT

DIGITAL ERA LOCAL GOVERNMENT IN ENGLAND SERVICE REFORM AND THE INTERNET

TOM NICHOLLS
GREEN TEMPLETON COLLEGE
UNIVERSITY OF OXFORD

Local governments have been developing online services, aiming to better serve the public and reduce administrative costs. However, the impact of this work, and the links between governments' online and offline activities, remain uncertain.

Using a new full-population dataset of the UK government web, this thesis directly analyses the structure and content of government online. It argues that recent digital-centric public administration theories, typified by the Digital Era Governance quasi-paradigm, are not empirically supported by the UK local government experience. Instead, it argues for a positive but critical attitude to the value of technology in public administration.

New methods are developed for full-text empirical analysis of government web crawl data and the use of machine learning approaches for adding attributes to full-population data. Webometric and content analyses of the .gov.uk web are conducted, together with a time-series panel regression of English local government cost, service quality, satisfaction, and an expert assessment of web provision.

English local government online is highly heterogeneous and there is remarkably little intrinsic structure to the .gov.uk web. Geography is weakly visible in online government structures, but there is little evidence that online structures closely reflect the offline activities of government.

Empirically, English local governments' web development levels are found not to be associated with improvement in the cost of, quality of, or satisfaction with council services, challenging the merit of online approaches to service improvement.

Overall, the importance of digital-centric approaches to local government service improvement is challenged. A more nuanced theoretical approach to online service delivery is advocated, which discards utopian enthusiasm and focuses on concrete service improvements.

It is recommended that full-population quantitative analyses be used more broadly in political science, with awareness of both the opportunities and the challenges associated with them.

CONTENTS

I	INTRODUCTION	19
2	LITERATURE REVIEW	27
2.1	Introduction	27
2.2	Public sector service reform and e-government	30
2.2.1	New Public Management and challenges to it	30
2.2.2	Digital Era Governance as an alternative	34
2.3	Administrative performance and Internet use	40
2.3.1	Modernism, optimism, and technology in government	41
2.3.2	Maturity models	44
2.3.3	Is use of the Internet associated with improved government service delivery?	47
2.3.4	The web and the shape of government	52
2.4	Local government, service reform, and e-government	53
2.4.1	Local government structure in England	53
2.4.2	Why local government is distinct and interesting	57
2.4.3	The policy context	57
2.4.4	Informational, transactional, and democratic uses of the Internet	62
2.4.5	Published accounts of local government Internet use	64
2.4.6	Socitm's <i>Better Connected</i>	69
2.4.7	Digital Era Governance: a good fit for local government?	70
2.4.8	Homogeneity and heterogeneity in local government and its web provision	74

2.5	Big data, political science methods, and policymaking	75
2.6	Conclusions and research question	76
3	NEW METHODS FOR ONLINE PUBLIC ADMINISTRATION RESEARCH	81
3.1	Introduction	81
3.2	Research choices and opportunities	83
3.2.1	Case selection and generalisability	83
3.2.2	Webometrics and full-population approaches	85
3.2.3	Operationalisation of key concepts	89
3.3	Big data approaches in public administration	91
3.3.1	Big data in political science	91
3.3.2	Web crawling in political science	92
3.3.3	Link analysis, nodality and organisation-level insights	93
3.3.4	Online traces of governments and their service users	95
3.3.5	Scales of analysis – linking traces to macro-level phenomena	98
3.4	Web crawling: methodological issues and ways forward	100
3.4.1	Definition of terms	100
3.4.2	Web crawling concepts	101
3.4.3	Strengths and weaknesses for statistical analysis	104
3.4.4	Web crawling ethics	105
3.5	Textual methods in political science	108
3.5.1	Automated classifiers	110
3.5.2	Classifying web content	113
3.6	New methods developed	117
3.6.1	The Warctika framework	117
3.6.2	Wayback Machine as classification platform	120
3.7	Conclusions	122
4	THE STRUCTURE OF GOVERNMENT IN THE DIGITAL AGE	125
4.1	Introduction	125

4.1.1	Why structure is important	126
4.1.2	Network-centric measures of digital governance	128
4.1.3	Hypotheses	129
4.2	Process	131
4.2.1	Data choices	132
4.2.2	Data set selection and collection	134
4.2.3	Mapping from websites to organisations	140
4.3	Findings	141
4.3.1	Node centrality and importance	141
4.3.2	Network communities: Infomap link clustering	152
4.3.3	Network communities: geography and linking patterns	156
4.3.4	Content analysis of connections	161
4.3.5	Outliers	164
4.4	Discussion	166
4.4.1	What explains the structure?	166
4.5	Conclusions	168
4.5.1	Limitations and further research	169
5	LOCAL GOVERNMENT USE OF THE INTERNET: CONTENT AND CONTEXT	173
5.1	Introduction	173
5.1.1	Hypotheses	174
5.2	Process	175
5.2.1	Classification of government web content: service and non-service content	176
5.2.2	Topic modelling of the .gov.uk corpus	179
5.3	Findings	183
5.3.1	Number of pages	183
5.3.2	Page classifications	189

5.3.3	Classification of government web content: transactional and informational services	192
5.3.4	Topic modelling	195
5.4	Discussion	199
5.4.1	Homogeneity, heterogeneity, and prospects for a local.gov.uk	201
5.4.2	The value and limits of web data for studying government	202
5.5	Conclusions	204
6	COUNCIL PERFORMANCE AND USE OF THE INTERNET	207
6.1	Introduction	207
6.1.1	The Comprehensive Performance Assessment	208
6.1.2	Better Connected	214
6.2	Internet engagement and service quality	217
6.2.1	Regression modelling of BC and CPA results	219
6.3	Value for money	226
6.3.1	Cost effectiveness measures	227
6.3.2	Regression modelling of cost effectiveness and BC results	229
6.4	Citizen satisfaction	233
6.4.1	Best Value Performance Indicators and citizen satisfaction	234
6.4.2	Regression modelling of citizen satisfaction and BC results	236
6.5	What explains these relationships?	239
6.6	Conclusions	242
7	DISCUSSION: THE INTERNET AND PUBLIC ADMINISTRATION	245
7.1	Introduction	245
7.2	Empirical findings: an overall picture	246

7.3	Digital Era Local Governance?	250
7.3.1	Digital Era Governance in local practice	250
7.3.2	Local governance in digital context	253
7.3.3	Transformation, digitalisation and public services	254
7.3.4	Digitalisation: consequences for DEG and ways forward	256
7.4	Implications for public administration theory	260
7.4.1	Explanations for the limited impact of the Internet on effectiveness	261
7.4.2	Public administration beyond DEG and hyper-modernist enthusiasm	265
7.4.3	Towards a theory of Digital Era Local Governance	267
7.5	Full-population web crawl data in political science research	270
7.5.1	Full-population datasets	271
7.5.2	Units of analysis	274
7.5.3	The ethics of web data	276
7.5.4	The practical challenges of large-scale web data analysis	278
7.6	Conclusions	279
8	CONCLUSIONS	283
8.1	The structure of government in the digital age	283
8.2	Local government use of the Internet: content and context	284
8.3	Council performance and use of the Internet	285
8.4	Digital Era Local Government	286
8.5	Political science analysis of large-scale data	287
8.6	Limitations, future research and policy implications	287
A	APPENDIX: DATA TABLES	291
A.1	Isolated .gov.uk domains	291
A.2	Distribution of web page topics by council	293

A.3	Performance regression models including council dummy variables	311
-----	---	-----

B	BIBLIOGRAPHY	319
---	--------------	-----

LIST OF FIGURES

Figure 1	UK local government top-tier boundaries	56
Figure 2	Web crawling process	136
Figure 3	Data cleaning process	139
Figure 4	Domain ownership coding process	142
Figure 5	Domain classifications	143
Figure 6	Indegree centrality by domain type	146
Figure 7	Indegree centrality by size: English councils	150
Figure 8	Indegree centrality by size: other domains	151
Figure 9	Third-level domain diagram, coloured by cluster	155
Figure 10	Mapping British councils and their web links	158
Figure 11	Supervised classification: feature extraction and training process	180
Figure 12	Latent Dirichlet Allocation (LDA): feature extraction and fitting process	182
Figure 13	LDA: keyword extraction	183
Figure 14	Website size by type of domain registrant	184
Figure 15	Total English council web site size by population	185
Figure 16	Proportion of service-focused pages, by domain type	191
Figure 17	By council: proportions of pages in each topic	197
Figure 18	Comprehensive Performance Assessment (CPA) ratings by year – top-tier authorities	211
Figure 19	CPA-derived service scores by year – top-tier authorities	214
Figure 20	Better Connected (BC) grades by year, 2002–2008 – top-tier authorities	216
Figure 21	BC / CPA association by year	221

Figure 22	Spend in excess of standard, by year	230
Figure 23	Expenditure, by year	231
Figure 24	Best Value Performance Indicator (BVPI) 3 citizen satisfaction, by year and type of council	236

LIST OF TABLES

Table 1	Top tasks from Socitm (2013a), with their categories	73
Table 2	Multiple auto-generated representations of a single document, from a live web crawl	114
Table 3	Web crawl dataset: final content	140
Table 4	Domain classifications	144
Table 5	Node importance: top 40 domains by indegree	145
Table 6	Node importance: local authorities in England	147
Table 7	Median indegree by domain type	148
Table 8	Regression models: indegree by English council type, population, and number of pages	148
Table 9	Infomap network clusters: .gov.uk	154
Table 10	Logit regression model: geography as determinant of linking patterns	160
Table 11	Logit regression model: geography diagnostics	160
Table 12	Manual classifications of web links by link to / from English councils	163
Table 13	Automated service classification: validation confusion matrix	177
Table 14	Automated service classification: validation results	178
Table 15	Regression model: number of web pages by council type and population	186

Table 16	Proportion of service and non-service pages by domain type	190
Table 17	Manual classifications of pages as transactional	193
Table 18	Topic modelling: topic size summary statistics	196
Table 19	Topics by council: measures of association	198
Table 20	LDA: topic labels and selected keywords	199
Table 21	CPA coding and grade label changes	210
Table 22	CPA ratings by year	211
Table 23	BC grades by year 2002–2008: top-tier authorities	216
Table 24	Pilot data analysis: CPA and BC, 2008 only	218
Table 25	BC / CPA association by year	220
Table 26	Regression models: CPA and BC, 2002–2008	222
Table 27	Regression models: council spending, CPA, and BC, 2002–2007	232
Table 28	Regression models: citizen satisfaction and BC, 2004 / 2007	238
Table 29	Regression models: BC and CPA, including lagged dependent variables	240
Table 30	.gov.uk: isolates	292
Table 31	Distribution of web page topics by council	309
Table 32	Performance regression models including council dummy variables	318

LIST OF REGRESSION MODELS

Model 1	Indegree for council web domains by council type and population	148
Model 2	Indegree for council web domains by council type, population and site size	148

Model 3	Council geography and linking patterns	160
Model 4	Number of web pages by council type and population	186
Model 5	CPA and BC, 2008 only	218
Model 6	CPA services and BC	222
Model 7	CPA services and all BC categories	222
Model 8	CPA rating and BC	222
Model 9	Excess spending and BC Transactional or better	232
Model 10	Raw spending and BC Transactional or better	232
Model 11	BVPI 3 satisfaction and BC, all councils	238
Model 12	BVPI 3 satisfaction and BC, top-tier councils only	238
Model 13	BVPI 3 satisfaction and BC, top-tier councils, with controls	238
Model 14	CPA services scores and BC including lagged dependent variables	240
Model 15	Excess spending and BC including lagged dependent variables	240
Model 16	Satisfaction and BC including lagged dependent variables	240

LIST OF RESEARCH QUESTIONS

Research question	Do the form and impact of local e-government reflect the expectations of Digital Era Governance and of local e-government practitioners? 77
-------------------	---

Research sub-question 1	What is the structure, shape, and connectedness of the English local government Internet presence and what are the relationships between different bodies? What can this tell us about the structure of government in the digital age?	77
Research sub-question 2	What is the content of the English local government web? What can this tell us about local government and its approach to e-government?	78
Research sub-question 3	What is the relationship between council administrative performance, cost-effectiveness and use of the Internet? What explains this relationship?	78
Research sub-question 4	To what extent does the experience of English local government fit the Digital Era Governance quasi-paradigm?	78
Research sub-question 5	What opportunities for future political science research exist in large scale analysis of web content?	78

LIST OF HYPOTHESES

Hypothesis H1	Local authorities' network importance, measured by indegree centrality, is substantially explained by council type and population.	130
---------------	--	-----

Hypothesis H2	Organisations on the network will be logically clustered by the structure of government, not by shared function or common citizen needs. 130
Hypothesis H3	The government web represents a single link network; there will be no isolates. 130
Hypothesis H4	Web linking between local authorities is substantially explained by geographic proximity. 131
Hypothesis H5	Web links between government bodies primarily reflect structural ties rather than service signposting, as reflected by hand classifications and the proportion of front page to deep links. 131
Hypothesis H6	Local authorities' extent of online activity, measured by the number of web pages, is substantially explained by council type and population. 174
Hypothesis H7	Local authorities' online presence is heterogeneous, measured by the proportion of each council's pages classified as service-focused. 175
Hypothesis H8	Local authorities' online presence is heterogeneous, measured by the proportion of each council's pages classified as transactional. 175
Hypothesis H9	Local governments' online presence is heterogeneous, measured by the proportion of each council's pages in each topic. 175
Hypothesis H10	Local authority engagement with the internet, as measured by BC performance, is positively associated with service performance, as measured by CPA scores. 217

Hypothesis H11	Local authority engagement with the internet, as measured by BC performance, is negatively associated with the cost of delivering services, as measured by spending in excess of the government-established standard when controlling for service performance. 217
Hypothesis H12	Local authority engagement with the internet, as measured by BC performance, is positively associated with citizen satisfaction, measured by BVPI 3 scores. 236

ABBREVIATIONS

ADM	Alternative Document Model
BC	Better Connected
BVPI	Best Value Performance Indicator
CPA	Comprehensive Performance Assessment
CSS	Cascading Style Sheet
CSV	Comma Separated Values
DCLG	Department for Communities and Local Government
DEG	Digital Era Governance
dfes	Department for Education and Skills
dfid	Department for International Development
DNS	Domain Name System
DTLR	Department for Transport, Local Government and the Regions
EDGE	Essentially Digital Governance
fsQCA	fuzzy-set Qualitative Comparative Analysis
GDS	Government Digital Service

HMRC	Her Majesty's Revenue and Customs
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HTTPS	HTTP over TLS
ICT	Information and Communications Technology
IP	Internet Protocol
ISP	Internet Service Provider
IT	Information Technology
JSON	Javascript Object Notation
LDA	Latent Dirichlet Allocation
LGIU	Local Government Information Unit
LGOL	Local Government Online
LSDV	Least Squares Dummy Variable
NPM	New Public Management
ODPM	Office of the Deputy Prime Minister
OLS	Ordinary Least Squares
ONS	Office for National Statistics
OXIS	Oxford Internet Survey
PDF	Portable Document Format
QAP	Quadratic Assignment Procedure
RTF	Rich Text Format
SNA	Social Network Analysis
SVM	Support Vector Machine
TLS	Transport Level Security
URL	Uniform Resource Locator
WARC	Web Archive
WAT	Web Archive Transformation
WET	WARC Encapsulated Text
WNTD	Web Network Thick Description
WPS	Website Performance Survey

INTRODUCTION

Local government is a major part of England's democratic apparatus and is responsible for overseeing £91.2bn per year in spending (DCLG, 2016). Current austerity drives are hitting local authority budgets considerably harder than those of central government, with a consequent renewed emphasis on reducing the cost of services. This has heightened the need for new ways of delivering local government services and organising local government operations.

Since the 1990s, there have been bursts of enthusiasm for putting local government services online – not least the ambitious target to make 100% of services e-enabled by the end of 2005 (ODPM, 2003) and the subsequent push by the Cabinet Office for “Digital by Default” across UK public services. An estimated 414m visits are made each year to local authority websites in the UK (Socitm, 2013a, p.3).

Despite this, and despite the huge changes that have occurred in other industries as the Internet has changed the way business is done, the impact of widespread use of the Internet on local government remains limited and understudied by academic public administration and e-government scholars. Even within the practitioner sector, only one body (Socitm) regularly focuses on specific aspects of local e-government and e-services.

Local government is the principal site of contact between citizen and government in the UK. If digitally-focused models of public administration reform (such as those presented in Dunleavy et al., 2006a) are correct, Internet-enabled improvements in service delivery might have a great impact here. Furthermore, it has been an important service delivery vehicle for central government since the assignment of poor law and public health responsibilities in the 19th

Century (Wilson and Game, 2011, p.57)¹. Much of the rhetoric of the current administration in stressing localism and the importance of place speaks to local government's centrality, as does the drive to make social care a local authority responsibility.

Having said that, central governments' tendency since the 1980s to bypass the established local authority delivery structures in favour of a thin, "commissioning" role (Clarke and J. Stewart, 1994), speaks to distrust of councils as a separate administrative structure and reinforces the idea that local government is the forgotten part of the state, despite its huge expenditure.

Even more directly on topic are a series of present discussions about the merits of merging all of the local authority websites in England into a single local.gov.uk domain. This is intended to mirror for local government services the perceived success of the Government Digital Service (GDS)-operated gov.uk website. Under this line of thinking, as the services provided by local government are largely duplicative, a single better site could be provided with locally-federated content when required.

This raises a significant empirical question: are local websites really minor variations of each other, or does the much vaunted "sense of place" at the heart of local government really result in the provision of substantially different content tailored to local needs? Given the homogenising pressures of heavy central government influence, coupled with the presumably homogenising effects of the large use of existing technology, there is a reasonable assumption of the former.

¹ These assignments were originally to special-purpose authorities – Poor Law Unions and local sanitary districts. The history of local government in England goes back continuously to the Anglo-Saxon period; until a series of reforming and systematising acts in the 1800s structures and responsibilities varied across the country, with much local service provision in the hands of specially-chartered ad-hoc bodies. The modern elected general-purpose local authority is a creature of the late 19th and early 20th Centuries, which saw the sanitary districts re-created into a unified and coherent system of authorities with tax-raising powers and a wide range of responsibilities including education, housing, the regulation of trading standards, planning, provision of social services, economic development, highways, leisure, waste collection and disposal, registration of marriages, and many others.

As a result of this unique positioning and the ongoing debate around local.gov.uk, local government in England makes a rewarding area of study for looking at the Internet and its interactions with public administration. While central government web services have had repeated changes of direction (from early departmental sites, through the work of the Office of the e-Envoy, through the attempt to centralise services through DirectGov, and most recently through the establishment of the Government Digital Service) local authorities have evolved separately, albeit with strong central government encouragement (Beynon-Davies and Martin, 2004). Analysing the level of web development over time, and its interactions with service quality, offers opportunities to look at the underlying relationships, with enough data available to control for the effects of authority role, deprivation, previous performance ratings and other potentially confounding factors.

This thesis therefore aims to contribute to debates within public administration on the effectiveness of e-government approaches, where there is a disconnect between a pervasive optimism and the limited empirical evidence of benefits actually obtained from greater technology use. It offers an opportunity to answer important questions about council performance and the use of online services: what shape of provision has been obtained, by fifteen years' worth of time, energy and money spent on e-government? Is implementation uniform within and between authorities, or heterogeneous? What explains this? To what extent have web developments been associated with improvements in local authority administrative performance and cost-effectiveness? To the broader public administration literature, it aims to see how well the experience of English local government fits the Digital Era Governance (DEG) quasi-paradigm's focus on reintegration, holistic services for citizens, and implementing thorough-going digital changes in administration (Margetts and Dunleavy, 2013).

The use of English local authorities as the unit of analysis allows a full-population analysis of empirical data. Comparisons between different British

jurisdictions are complicated by quite different local government structures and legal frameworks, especially in Northern Ireland and Scotland, which compromises the comparability of data where government statistics are being relied upon. Literature on local governments outside England will, however, inform the study.

Empirically-minded public administration scholars have been increasingly interested in the Internet and e-government in recent years, albeit from a very low base. Unfortunately, the field is still relatively underdeveloped. Heeks and Bailur (2007, p.243) report that research “draws mainly from a weak or confused positivism and is dominated by over-optimistic, a-theoretical work that has done little to accumulate either knowledge or practical guidance for e-government”. Empirical work has historically been particularly weak: “Of the empirical projects that have looked at e-government, most have limited their analysis to single American states or small numbers of Web sites, weakening the generalizability of the findings.” (West, 2004, p.17).

Direct empirical work based on measurable process or outcome data is rare: “the majority of the studies are based on surveys (65%), followed by case studies (36%), and other methods (7%)” (Andersen et al., 2010, p.567). Although survey-based and more qualitative approaches are valuable for understanding the motivations and work situations of participants, they are not capable in themselves of answering many harder-edged questions about performance improvement which we may wish to ask. This methodological limitation is also true of local government work, which is even less prominent within the public administration literature and which often features $n = 1$ case studies to the exclusion of other empirical methods².

2 As Meijer and Zouridis (2004, p.2) note: “We are dazzled by the success stories and case studies, sometimes told by the project champions (consultants and civil servants). Their principle message is: ‘In our country we used to have paper government in this policy sector (education, tax, social security), but now as you can see (via a direct Internet connection) we now have e-government’. However, as impressive these case studies may sound, their resemblance sometimes becomes boring.”

In addition, new tools are now available to study public administration involving the use of census-scale web data (Dunleavy, 2010a). There has, however, been no systematic use of the data available *from* the Internet to assess local government service reform. There are also extensive tool sets available from computational social science, corpus linguistics and computer science to analyse the networks formed by local government web content, the content of local government information online and the extent to which local government web provision is homogeneous.

In addition, this thesis develops novel approaches to the study of public administration, using big data approaches to model aspects of local government public administration at a population level. In particular, large scale text analysis and webometric analysis of connectivity is used to explore and demonstrate aspects of local government activity which have not been sufficiently explored by existing work.

This will be a predominantly quantitative thesis. The lines of enquiry have been suggested by the limitations and gaps in existing empirical work, and by theory, particularly the Digital Era Governance work by Dunleavy et al. (2006a) and currents of service improvement thought in the local government literatures. Nevertheless, due to the state of evidence in the field, the initial part of the work will be exploratory in structure. Direct large-scale collection of new local authority web presence data will allow the analysis of network-centric measurements of scale, connectedness and density. This data will also support the testing of hypotheses about the nature of the government web. It will “investigate[...] the reality of what has happened, rather than accepting the wild predictions of politicians and commentators” (Margetts, 1999, p.xv).

No one academic field has a sharp focus on local government service reform and use of the Internet, so this thesis is necessarily multidisciplinary. “E-government can be seen as sitting at the cross-roads between a number of other research domains, particularly computer science, information systems, public administration and political science” (Heeks and Bailur, 2007, p.252).

Each field takes a different perspective and emphasises different parts of the problem.

Key areas of interest in this thesis include both narrative and analytical topics. Firstly, what does English local government use of the web look like? How has it changed, how is it structured and connected, how much is it used? What are the online relationships between the different bodies involved in local government and what can this tell us about the structure of government in the digital age? More analytically, what relationship is there between Internet use by local authorities and performance and cost-effectiveness? Is widespread use of the Internet, as has been suggested, a key driver of efficiency gains in a time of austerity? And to what extent does the experience of English local government fit the DEG quasi-paradigm?

The contributions of this thesis are both substantive and methodological. Methodologically, innovative techniques are developed for using full-population government data and a comprehensive web crawl for answering some of the questions normally handled by convenience sampled surveying and manual content coding of websites and documents. In particular, solutions are offered to some of the issues intrinsic to running large-scale web crawls in political science and public administration, along with new software tools to handle the large volume of data created by them and to extract useful information from it.

Substantively, a key contribution is a robust full-population analysis of local government performance in light of e-government efforts, showing that there are no demonstrable gains in service quality or cost which can be attributed to greater web development (see Chapter 6). In addition, the development of local e-government is mapped as both a set of narratives and as a set of direct measures of online web networks, and the extent to which councils have conformed to the New Public Management (NPM) and DEG approaches is measured. The findings of all these analyses inform a critique of existing

models of e-government and the development of new theory around electronic government and service improvement.

Chapter 2 reviews the existing literature surrounding local government, service performance and the Internet. It focuses on the conflict between techno-optimist models of service improvement and the uncertain returns of previous rounds of informatisation and establishes a research question and a set of sub-questions.

Chapter 3 argues for the use of webometrics and full-population machine learning approaches as a means of answering the research question set. It develops a set of new approaches to the study of web crawl data which not only enable the answering of the research question but also have wider potential for use in political science and public administration.

Chapter 4 analyses the structure of government revealed by a new web crawl of the .gov.uk domain. Using large-scale webometric approaches, it examines how this structure can be explained. Weak geographical associations are revealed to underlie the structure of the UK government web, but the structure of web linking otherwise yields surprisingly little information. A key finding is that online and offline government structures differ, which has significant implications for the future use of web-based approaches in public administration.

Chapter 5 extends the analysis of Chapter 4 to the content of the web as published by government bodies. Using innovative machine learning approaches, the textual content of the local government web is used to assess the extent of online service delivery by local government and to reveal the high volume of content which is not, in fact, related to any form of service delivery. A case is made for the importance of informational, as opposed to transactional, uses of the Internet in local government.

Chapter 6 conducts a panel data analysis of local government web development against service performance, cost, and customer satisfaction measures. It concludes that there is no evidence that the mass of work undertaken on

local e-government has produced positive results, and links this finding back to previous theoretical debates on the effectiveness of informatisation.

Chapter 7 argues for improved theories of public administration and greater use of full-population methods in political science. It challenges DEG and utopian theories of digital service improvement in light of the empirical and theoretical findings of this research. It argues for a more critical and nuanced view of the opportunities of technology in public administration and cautious optimism about the possibilities for future improvements. The chapter concludes by outlining the methodological contributions of the research and advocating for wider use of large-scale full population data analysis in political science, and for an understanding of the opportunities and challenges that they offer the researcher.

A final major research output of this study is the .gov.uk web crawl itself, which is available for general reuse by political scientists interested in the Internet. A full link graph has been extracted, together with the classifications developed as part of the project. For more information, see Section 4.2.2.

LITERATURE REVIEW

“We do not cite empirical works that validate claims made in cyber-optimist writings because, after an extensive review of the e-government literature, we have not been able to find any”

— Norris and Reddick (2013, p.167)

2.1 INTRODUCTION

The dominant theoretical framework for governments seeking to improve public administration has, since the 1980s, been New Public Management (NPM), an approach based on private sector competition approaches. In recent years, practitioners’ public administration and public management focus has been on service quality and value for money, with NPM and private sector-derived approaches dominant (Van Ryzin, 2013).

Nevertheless, the NPM paradigm has run out of steam. There are extensive academic critiques of NPM on both theoretical (e.g. Parsons, 2002) and practical (e.g. Andrews and Van de Walle, 2013) grounds. Some of the basic assumptions of NPM turn out not to be valid in practice: for example, NPM disaggregation and performance management techniques are only very weakly associated with the more customer-focused culture that they were supposed to engender (Wynen and Verhoest, 2013). In a comprehensive overview of 30 years’ worth of largely NPM-driven reform in UK central government, Hood and Dixon

(2015) were highly critical of the whole approach, finding that costs rose and service deteriorated (or at best stayed the same).

However, despite the development of new approaches based on a rediscovery of the distinctiveness of public administration (Moore, 1997), networks of policy actors (Lecy, Mergel and Schmitz, 2014), and others, no one major paradigm has emerged to replace NPM as the public administration toolkit reached for by most public sector policymakers. As a result, we have seen the development of what has been described as “zombie NPM”, where the approach is intellectually dead but keeps shambling along for want of being given a final burial (Dunleavy, 2010b).

Crucially, NPM has little to say about digitalisation, and the opportunities now afforded by the Internet to deliver public services. In light of the huge efforts made by governments to use the Internet as a service delivery channel in recent years, this has led to new models which take e-government seriously; the most prominent of these is Digital Era Governance (DEG), pitched as both a descriptive explanation of how digital developments are changing governance, and also as a toolkit to use for the improvement of public services in the future:

“The toolkit for public management reform has shifted away from a ‘new public management’ (NPM) approach stressing fragmentation, competition and incentivization and towards a ‘digital era governance’ (DEG) one, focusing on reintegrating services, providing holistic services for citizens, and implementing thorough-going digital changes in administration.” (Margetts and Dunleavy, 2013, p.1)

Despite this, there are substantial questions as to the efficacy of electronic tools for improving service delivery. Despite huge enthusiasm from practitioners and reports of huge improvements made (Foresheew-Cain, 2015), a long-running strand of information systems research has been reporting less-than-predicted returns on investment for informatisation projects since the 1970s, and there are some signs that this also applies to e-government work. After a long career studying e-government in both the public and private

sectors, Kraemer and King offer a resolutely pessimistic view of the utility of IT as a tool for government service reform:

“On the contrary, we believe even more strongly than before that IT is not a catalyst for administrative reform in government. Machiavelli’s fifteenth century admonition about the perils of administrative reform is as applicable today as then. IT continues to be a useful instrument of administrative and incremental change, but it is no more capable today of bringing about institutional change and administrative reform than it was fifteen years ago.”
Kraemer and King (2003, p.5)

This thesis is particularly concerned with local government in England. Understudied by both public administration and electronic government scholars, specific issues arise from the greater use of technology in local government. In addition to the general question of how to spend its huge budgets to deliver the best services and value for money, the nature of local government as a set of geographically-rooted bodies sits in tension with the universalising and globalising tendencies of electronic approaches to public administration. Put bluntly, if local government services are being delivered through websites to citizens who could be connecting from anywhere, is it of value to continue to break up the delivery apparatus into several hundred separate bodies?

This debate touches on the traditional local government studies question of whether councils are best seen as a homogeneous set of local service delivery agencies or as a heterogeneous group of bodies that distinctively respond to the local context. It has come to a head with practitioner debates in the UK around the possible creation of a single local.gov.uk: an attempt to expand the single government service delivery website gov.uk to encompass local authority services (Czerniawski, 2014; Socitm, 2014; Copeland, Fink and Scott, 2014).

These areas of interest have attracted ongoing research in several different fields: “E-government can be seen as sitting at the cross-roads between a number of other research domains, particularly computer science, information systems, public administration and political science” (Heeks and Bailur, 2007, p.252). Even within the major discipline studying the field, work on the Internet

and informatisation is quite separate and has “only had a limited impact on mainstream debates in Public Administration” (Meijer, 2007, p.234). As a result, this literature review will be thematic rather than disciplinary in structure.

This chapter reviews the literatures around these questions and develops a research question and a set of sub-questions (in Section 2.6) that motivate the rest of the research project. An analysis of the methods literature, including the development of quantitative methods suitable for answering the research question set here, is in Chapter 3.

2.2 PUBLIC SECTOR SERVICE REFORM AND E-GOVERNMENT

2.2.1 *New Public Management and challenges to it*

The dominant theoretical framework for governments seeking to improve public administration is NPM, an approach which applies classically private-sector market-driven approaches to attempt to improve government efficiency. It has been the main public administration approach in practice since the 1980s, and has been the subject of a “long-running battle between the true believers in the efficacy of new managerialism (and related remedies for performance problems in government and public services) and the academic and other sceptics” (Hood and Dixon, 2015, p.179).

The evidence for its success, however, is much weaker than the enthusiastic arguments made for it in the early days would lead one to suspect. As noted above, the theoretical and practical critiques of NPM have been extensive. The analysis by Hood and Dixon (2015) of the United Kingdom, which is one of the stand-out enthusiasts for NPM in global terms, is one of the most empirically-founded comprehensive critiques. It notes a “remarkable lack of proof” of the

empirical benefits (p.178), and particularly challenges the idea that large-scale outsourced technology projects have been beneficial to the UK¹.

In addition to the lack of evidence of actual cost reductions or service improvements identified by Hood and Dixon, the rigidity of the NPM model has also been criticised, particularly given the enthusiasm with which its notions of disaggregation and competition have been adopted by the UK and other countries. Performance management and accountability-by-contract have been subsequently criticised by private sector management theorists and the appropriateness of such tools for much of the public sector is heavily challenged:

“NPM failed (and continues to fail) because it substituted a monolithic, fixed private sector delivery model for a monolithic, fixed public sector delivery model.” (Fishenden and Thompson, 2013, p.998)

Where NPM methods are currently defended by researchers, there is a focus on applying them in situations that would clearly benefit from NPM prescriptions and in contexts where they could be effective (rather than as the universal panacea for all government ills). This is the much more nuanced claim that “there is nothing inherently wrong with NPM tools as long as adequate implementation and suitable local conditions are in place” (Dan and Pollitt, 2015, p.1307, described by its authors as an ‘optimistic’ review). While some tools may have limited empirical support in certain environments, though, it is extremely difficult to defend NPM as a paradigm, given the evidence against it (Drechsler and Randma-Liiv, 2015).

Although dominant in practitioner discourse, NPM is by no means the only body of theory. It is particularly weak in the e-government / informatisation / Internet government space, largely pre-dating the digital developments of the 1990s and subsequently; consequently alternative theories are common.

¹ Technology is not a central focus of NPM, which prioritised managerialism, but a sense of technology as a modernist deliverer of rational outcomes has led it to be adopted by some NPM enthusiasts as part of reform efforts (Margetts and Dunleavy, 2013).

Indeed, upwards of a quarter of published e-government papers focus on discussing the conceptualisation of e-government and its terms (Joseph, 2013, p.435).

Since 2000, most prescriptions for modernisation of public administration have included technology, alongside other innovative approaches (Bekkers, 2007). Alternatives proposed to NPM have included both descriptive and normative models based on networks of public and voluntary sector actors (Kapucu, Hu and Khosa, 2014), and amorphous all-encompassing models based on a broad notion of “governance” rather than the more focused “government” (Larsson and Grönlund, 2014; Trotta et al., 2011). Despite a sustained and effective critique of the validity of NPM prescriptions, none of the proposed replacements have been widely adopted as a replacement framework and hence a “zombie NPM” approach (Dunleavy, 2010b) carries on.

Dunleavy et al. (2006b) critique NPM’s model of public sector IT outsourcing and highlight the challenges of informatisation and the opportunities presented by Internet technologies to improve public administration. Although well-cited², there is not a substantial body of competing theory in this space: the work that uses it seems to do so as part of an overview of the field (e.g. Morgeson and Mithas, 2009; O’Neill, 2009). Several of its concepts do inform other studies without necessarily being explicitly cited as a theoretical base: for example Quigley (2004) draws on Margetts (1999) in support of the weakness of the central government IT function in the UK.

Other empirical observations are shared by other critics. That the supposedly free-market doctrine of NPM has in Britain ironically created a closed market of oligopoly IT suppliers is also noted by Fishenden and Thompson (2013, p.982) and others.

Other alternatives to NPM also have a significant place in the public administration literature although the links between these approaches and e-

² Dunleavy et al. (2006b) alone has 286 citations in the Web of Science as of October 2016.

government have not been fully explored. The framework of Public Value was a relatively early attempt to displace NPM's focus on private-sector values as an underpinning (Moore, 1997). More recently, there has been an explosion of work looking at networks of public sector actors as a more effective explanatory frame for governmental motivation and activity:

“Networks are used both as a metaphor to describe and explain intergovernmental, cross-sector, and interorganizational relationships (Koliba, Meek, & Zia, 2010) and as methodological tools to examine relational data (Choi & Kim, 2007; Kapucu, 2006a ; Kapucu & Demiroz, 2011; J. Lee & Kim, 2011; Wachhaus, 2009)” (Kapucu, Hu and Khosa, 2014, p.2)

This work has the characteristic of being an explanation of process rather than a complete public administration theory per se. The central insight of the field is that the relationships between different actors are important; no single set of propositions for how to reform public administration in light of this insight has yet emerged to rival NPM. This network-centric public administration research is heavily focused on the nuances of interaction between actors on a small scale: its position in the wider network analysis field is in small-scale Social Network Analysis (SNA) with a sociological bent, rather than in the larger and more quantitative approaches popular in computer science and physics which are discussed in the context of Internet research in Section 2.3.4 below.

This thesis takes as a starting point Digital Era Governance (DEG) (Dunleavy et al., 2006a,b; Margetts and Dunleavy, 2013)³, which attempts to provide a replacement for NPM as a quasi-paradigm for public service reform. It is part of a wider body of literature criticising NPM and proposing alternative methods of improving public administration.

The main reason to prefer DEG as a framework to these other approaches is the digital dimension. With digital service transformation being such an important part of government rhetoric and practice, some kind of response by

3 The 2013 paper offering a revised variant snappily subtitled DEG₂

the academy is necessary. The DEG angle is to say “yes, this is important”, and to put it in the context of changing IT delivery arrangements and approaches to service aggregation.

The amount of critical and data-driven analysis investigating the effectiveness of digital public administration is still limited, however, as highlighted in the introduction and discussed in detail in Section 2.3.3.

2.2.2 *Digital Era Governance as an alternative*

DEG is one of a significant number of new models of public administration offered as a replacement for NPM over the last thirty years.

The original quasi-paradigm of DEG was set out by Dunleavy et al. in their book *Digital Era Governance: IT corporations, the state, and e-government* (Dunleavy et al., 2006a). In 2013 a follow-up publication emerged, describing a putative “second wave” of DEG representing a more developed version of the first (Margetts and Dunleavy, 2013).

The DEG case is that governments are, and should be, moving away from NPM approaches towards three new themes:

- A. Reintegration, where the disaggregation tendencies of NPM are reversed in order to exploit digital-era technology opportunities;
- B. Needs-based holism, in which the relationships between agencies and citizens are simplified and processes are ‘re-engineered’ on an end-to-end basis; and
- C. Digitalisation, with electronic delivery channels becoming the transformative central approach, rather than being supplementary – agencies ‘becoming their websites’ (Dunleavy et al., 2006a, p.227)

The qualitative picture painted by the research, that there are substantial problems with multiple countries’ IT procurement approaches, is cogent and well-argued. It is unambiguously clear, for example, that British central government bodies have a long history of failed IT system procurement, and

Dunleavy et al. identify several reasons for this, compared with other more successful countries.

What this does not go so far as to answer, though, is the link between this large-scale but contained empirical finding and the authors' desire to present a new quasi-paradigm for public administration in the later parts of the same volume. The empirical findings are necessarily an "is", and one about a relatively small problem in the public administration literature at that.

A claim for a new approach to public administration based upon having a less problematic IT procurement policy would be one thing, but DEG ambitions are much wider: no less than to replace NPM as the dominant narrative of public administration reform, making central not only the value of digital service delivery but also values of needs-based holism and reintegration. It is not clear that the evidence presented goes this far, though the theoretical discussion in Dunleavy et al. (2006a) does attempt to draw in this direction.

The principal tool used in Dunleavy et al. (2006a, chapters 3–5) to generalise from the particular cases studied⁴ is fuzzy-set Qualitative Comparative Analysis (fsQCA). This originates from Ragin (2000), and is based on assigning cases to (fuzzy) membership of sets, then using set-based operations to assess whether various set memberships appear to be associated with particular outcomes of note.

The use of fsQCA methods here are somewhat problematic. Despite their current popularity as a way of bridging the divide between low-N qualitative and high-N quantitative studies, the foundations of the fuzzy-set approach as a way of drawing firm conclusions from medium-N studies are quantitatively unappealing. In particular, the method relies on a multi-dimensional comparison of all units which is analogous to running repeated statistical tests, with the inherent Type I error problems which arise from repeated tests on the same data (Braumoeller, 2015). Results from fsQCA have also been shown

⁴ Australia, Canada, Japan, the Netherlands, New Zealand, the United States, and the United Kingdom

to be highly sensitive to minor variations in model specification (Krogslund, Choi and Poertner, 2014).

The major tension within DEG as a presented theory is its status as a mixed normative / empirical approach. The critiques of existing approaches to government IT are based on an empirical research base assembled for a set of propositions about IT contracting and e-government performance. However, DEG itself is in the nature of an ideal type, or as the authors state a ‘quasi-paradigm’ (Margetts and Dunleavy, 2013). This makes analysis difficult, as no government has actually implemented DEG in its pure form, even in the UK which has clearly been continuing its aspirations towards thoroughgoing digitalisation:

“The United Kingdom is currently witnessing a coalition government exhibiting a rhetoric and aspiration that is clearly strongly rooted in DEG, but the evidence on its ability to drive effective implementation remains deeply mixed.” (Fishenden and Thompson, 2013, p.997)

The extent to which DEG is a viable model in practice therefore depends both on empirical evidence of the effectiveness of the three themes (reintegration, needs-based holism, and digitalisation) as a means of improving government, and also on a normative discussion as to whether the end state of a largely digital approach to public administration is desirable. The empirical and normative aspects of DEG will be addressed in turn.

There are a number of proposed mechanisms by which DEG approaches can improve service delivery. Parts of these, such as a focus on better information technology contracting and management practices, are largely outside the scope of this research, despite the obvious importance of moving past the chaotic and ineffective approaches to technology management that have characterised the UK government in particular.

There is some acknowledgement that the changing conditions of the IT market and government approaches to Internet-based service delivery sit well within a DEG framework. Fishenden and Thompson (2013), for example,

consider that the disaggregation of governmental IT systems and their rebuilding around a black-boxed Open Architecture make possible DEG approaches. They criticise continued use of monolithic contracting approaches within central government as making the delivery of DEG more difficult (Fishenden and Thompson, 2013, p.979).

DEG requires innovation beyond that available in traditional outsourced system procurement, especially the development of interoperability, open standards and the building of software ecosystems rather than monoliths:

“It is notable that very few of the above features of DEG can be delivered via a traditional “command economy” approach to governance. This is because the great majority of DEG features, even those involving reintegration, involve new technologies, incentivization mechanisms, and resulting behaviors that do not currently exist – either within the existing public domain or within existing outsourcing arrangements” (Fishenden and Thompson, 2013, p.981)

The Government Digital Service (GDS) is a new part of British central government, added to the Cabinet Office by the coalition government in 2011 to drive improvement in digital provision. It has been responsible for the setting up of the new cross-government gov.uk domain to bring services to the public together in one place, as well as reforming the means by which government procures and delivers Information and Communications Technology (ICT). Notably, it has been given the power to disapprove proposals for IT contracts from other departments and works much more through an in-house team of coders and designers than the huge-scale and fully-outsourced model criticised so heavily by Dunleavy et al. (2006a).

The innovations identified above might, therefore, be thought of as happening within the GDS and in a few other parts of government, but it is far from clear that the culture of Whitehall or indeed local authorities is such to allow this approach to be quickly widespread.

However, it is not certain that DEG is as general a model for change as suggested by its proponents: wholesale digitising of services is much easier

for simpler transactional services than for those developing complex policy in politically contentious areas (Laffin and Ormston, 2013). It is also not clear whether DEG is simply the latest in a long line of alternative frameworks proposed by NPM-averse public administration scholars, rather than a genuine movement within government service itself.

The mark of a ground-breaking theory of public administration, I suggest, is that it is actually adopted by governments and used as a means by which public administration can be reformed. NPM, for all its obvious empirical and theoretical flaws, passes that test. It is, as yet, unclear whether DEG will do so.

Clearly DEG is not being universally applied as a guide to action by government in the UK. Much ICT activity continues under legacy contracts such as at Her Majesty's Revenue and Customs (HMRC). Some new projects continue to be structured as large monolithic contracts with a single large provider: Universal Credit is the latest example of a failed government IT project on the old model. Nevertheless, there is some indication, especially through the coalition government's work with the GDS, that new models of contracting are replacing the British-style NPM approaches that had previously dominated.

A key unknown variable is the effect of the UK government's austerity program. Three scenarios are outlined by Margetts and Dunleavy (2013) for the further implementation of DEG:

- “a systematic revival of NPM approaches, and abandonment of or crisis for the DEG quasi-paradigm. So far, this seems to have been limited to short-lived spasms in the UK and elsewhere;
- an austerity squeeze beginning a ‘pause’ of DEG changes that perhaps opens up into an escalating ‘de-modernization’ or permanent lag in the organization of public services. This scenario remains a live one; and
- no more than a short-lived austerity pause in DEG developments, with long-run changes returning to trend, reflecting the need to continuously boost government productivity and curtail relative price increases, and to keep the administration of liberal democratic governments in close touch with digital-related changes in business and civil society.”

(Margetts and Dunleavy, 2013, p.14)

With the benefit of hindsight, this assessment seems over-optimistic. Austerity itself has been far more than short-lived, and the austerity-driven reduction in local government resources has been sufficient to result in the outright closure of valued services such as public libraries and major cuts being made in other areas. All kinds of service redesign take both initial capital and management attention to carry out, and local authorities are short of both. The wave of nationally-focused activities carried out in the 2000s by local authorities has not been repeated, and there is little evidence of enthusiasm for structural re-integration in particular. Rather, budgetary pressures have resulted in the spinning-out of services by local authorities to co-operatives, social enterprises, and volunteers. This looks more like NPM-recommended disaggregation than any kind of holistic provision, and ruling out the first scenario above would be premature.

Theoretically, there are reasons to believe that some of the empirical evidence assembled against existing IT provision applies with less force in English local government than for the central government bodies studied. It is not clear that the large scale historic outsourcing of ICT functions applied in the same way in local government in England, for example, at least in part because of local political resistance to the ideal of privatisation. Unlike in central government, local administrations have (by virtue of a certain amount of political autonomy) the ability to row in a different direction to that which is being taken by their peers and that suggested by central government. This autonomy is, to be sure, limited by both central government power to control funding and also legal restraint on the activities of local government (which still have no formal power of general competence) (Wilson and Game, 2011).

In addition to these theoretical concerns, a number of empirical questions remain to be answered. The key question is whether digitalisation of services actually delivers improvements in cost or quality of service. As explored in Section 2.3.1 below, this is a troubled question which has been the subject of debate by public administration and IT scholars for several decades. Further,

do the same factors which drive central government IT performance also drive local government?

The aspect of DEG to be explored and empirically tested in this research is that which describes (and advocates) a thoroughgoing digital reform in administration. Section 2.3.1 reviews the literature on administrative performance and use of technology and suggests a research sub-question to test this proposition in light of the local government experience in England.

2.3 ADMINISTRATIVE PERFORMANCE AND INTERNET USE

There have been many different terms used, confusingly, for public administration and information systems research into government attempts to improve services by using IT. Various authors and fields of research have described “e-government”, “digital government”, “Internet government”, “G2C”⁵, “digitalisation”, “informatisation”, and other terms for the broad concept of service delivery (and sometimes democratic engagement) via the Internet. Although these different terms each have their own history and flavour, they will be treated as near-synonyms in this research.

There has been surprisingly little attention within the e-government literature to what has been pithily described as “the theoretical ideal of information technology as [an] instrument of administrative change” (Kraemer and King, 2003, p.2). In other words, does IT act as a catalyst or enabler in the process of public administration improvement, or not?

5 “Government to Citizen”, by contrast with other x2y template identifiers where x and y are any of “Government”, “Business”, “Citizen”, or “Customer”.

2.3.1 *Modernism, optimism, and technology in government*

The program of technology-supported public service improvement has a strong flavour of modernism. The enthusiastic pursuit of e-government as a clean, efficient, and futuristic method of delivering public services by the Blair government was almost teleological: the nature of a modern Britain requires a modern government; e-government is therefore beneficial because it feels modern, almost irrespective of any concrete benefits that might accrue as a result of its implementation.

Optimistic technologists and enthusiastic administrators have long seen IT as a means for the rational improvement of administration. Margetts (1999, pp.164-178) places IT directly within the conflicts of modernism, with enthusiastic hyper-modernists seeing technology in technologically deterministic terms as positively transformative of the state, anti-modernists focusing on the negative possibilities of increased government control, and post-modernists positing technology as a central element of an emancipation from the bureaucratic. More compelling, though, are the arguments made for the tradition of critical-modernist writings – those which recognise the complexity and uncertainty inherent in any development of a more computerised society and are “more careful and realistic” about the actual impacts of technology on government (Margetts, 1999, p.169).

When looking at the impact of the Internet, rather than informatisation in general, the optimistic approaches of the hyper- and post-modernists are even more salient, because of parallel techno-utopian traditions amongst the technologists who were influential in the early development of the Internet. *The Declaration of the Independence of Cyberspace* (Barlow, 1996) is perhaps the highpoint of the post-modern trend, declaring “Governments of the Industrial World” as “weary giants of flesh and steel” and advocating their technologically-enabled obsolescence. The same West Coast computer culture that bred Barlow’s

techno-enthusiasm also sustained the early development of the commercial web.

This notion of e-commerce as a shiny, modern, and efficient way of doing business was not only extolled by Al Gore in the United States as a model for government, but directly influenced the Blair administration's e-government drive in the UK (discussed in Section 2.4.3 below). The influence of technoutopian thinking hasn't entirely disappeared even within public administration: the *Declaration* was quoted non-ironically as a foundation of rights-based online public administration principles by Dunleavy and Margetts (2015, p.19) in setting out Essentially Digital Governance (EDGE), a paradigm for digital public administration which builds upon the digitalisation aspects of DEG.

Critically, the approach of the Internet pioneers was self-validating in terms of empirical success: the Internet as a means of connecting computer scientists across the country was wildly and obviously successful, and the early commercial pioneers found ready markets and (until the 2001 .com crash) no shortage of enthusiastic investors. As has been repeatedly demonstrated over the history of informatisation research, though, reforming existing organisations using technology is neither simple nor a sure thing. In contrast to the optimistic enthusiasm that characterised early UK local e-government efforts, therefore, we must take a much more critical eye to examine the benefits or otherwise of the work that was done.

There is little evidence that use of the Internet has done much to fundamentally change the underlying power relationships within government. The disputes within the literature between theorists of weak or strong technological determinism, and those arguing that organisations are complex socio-technical systems for which technology is only one variable, have generally been resolved in favour of complexity and against a reductionist view of technology as the sole driver (Nograšek and Vintar, 2014). All the way back to early theoretical and empirical work on informatisation such as Danziger et al. (1982), the complexities of implementing information systems, and the extent to which

organisational variables change the outcomes of technical projects, have been well established. Such work that there has been on the effects on participation and democracy have generally suggested no strong link (Andersen et al., 2010).

This work has changed our view of ICTs and how they interact with their organisational environment and remains hugely important. This focus continues in later work such as Fountain (2001), which is again focused on technological and service design and how they interact. Less considered, by contrast, are studies focused on a later part of the implementation chain using technology as an independent variable: having implemented (or not) online services, to what extent are they associated with improved local government performance?

Looking at this stream of work more broadly, even where models of government IT use emerge they are often focused on micro-interactions within agencies, and generally ignore local government (see, for example, Cordella and Iannacci, 2010, in which the underlying case is the UK criminal prosecution system and the model aims to explore the values embedded in information systems, not their takeup, use, or effectiveness). As with much public administration work, though, the underlying case study supporting the theory is of one sector of government – UK criminal prosecution. It is unclear whether the factors identified can be generalised to other areas of public administration.

Public organisations, and local governments in particular, have distinctive uses for the Internet (as an information-processing and -aggregating system) that are much more useful than the incremental instrumental efficiencies to be had from the automation of existing processes. Not only is much of local government work about understanding citizens so as to deliver services and improve localities, but much of the information to understand what is really needed is not available to bureaucrats as a matter of course. In short, local government is information-constrained.

Many public organizations, whose task is the administration of public welfare or the monitoring and policing of a society, are chiefly “intelligence organisations”, where human actors work together in order to store and to process information and to

produce information output. The “production processes” of these organisations at their operational level are mostly informational, with “decisions” or informational services as their output. (Lenk, 2007, p.209)

2.3.2 *Maturity models*

Perhaps the most prevalent expression of ordered and rational modernism in IT development has been the persistence of maturity modelling approaches. These model the development of a particular technology or business method against an ordered progression with a defined end-state.

Consultancies and academics have introduced various overlapping stages-of-development models to measure the maturity of e-government services (see, for example, Layne and Lee, 2001). This builds upon a long tradition of maturity models for computer use in the private sector, beginning with Nolan (1973). Nolan expressly claimed his model as normative, but more recent attempts have tended towards an uncritical presentation of such models as in some way descriptive of a general trajectory in informatisation. The criticism of such models also has a long history: Benbasat et al. (1984) provide an empirical critique of stage models in IT which still looks powerful as a critique of today’s e-government models.

Stages models are, in some cases, conflated with the broader political science concern for democratic engagement. West (2004), for example, holds “interactive democracy” to be the fourth stage of his four-stage development model. These issues are clearly interesting, but do not fit well with the current English context, where the focus is strongly on services and cost reduction, and public interest in any form of democratic engagement with local authorities is at very low levels. A three-stage model offered by Andersen et al. (2010, p.572) is perhaps more robust and typical:

- Information provision and dissemination (the “billboard” stage)
- Transactions and service delivery.

- Participation and collaboration

At the time of their writing, most governmental websites in the US were focusing on the first stage of this model, though some authorities are beyond this: Reinwald and Kraemmergaard (2012) report a case study of a Danish authority which they claim has reached the “transformational government” stage of the Siau and Long (2005) five-stage maturity model.

The implicit notion of technology as a one-dimensional route from “un-developed” to “developed” is easy to criticise; technology can be used for many purposes, and it is trite that the mere availability of technology doesn’t determine outcomes in a social system. This one-dimension staged maturity model long predates the Internet as a fascination of local government technocrats; Kraemer, Dutton and Northrop (1981, p.17) note that the computing literature as long ago as the 1970s suggested that the computer environment in an organisation should “approach the most advanced technological stage appropriate for a given purpose”.

A particular problem with maturity modelling is in those models which promote interactive democracy as the highest form of development, above transactional service delivery. This seems difficult to accept simply because there is no causal or temporal relationship between an authority’s degree of democratic engagement and its quality of service delivery: it is quite possible to have thriving local democratic politics but dreadfully poor services, or the domination of a single elite group but efficient service delivery. The same is presumptively true with councils’ online presences, so it would be better to see the development of democratic website features as being on a different axis of development to that of services.

Maturity models in general are attractive to practitioners as they offer a goal for service development. However, the latter stages of the model are invariably speculative: the end state may be desired, but with much of the sector at the “early” stages of the model, there is no robust empirical evidence that a simple linear development to the indicated end-state is feasible (Brown, 2007).

There is also a clear normative underpinning that “the more e-government the better” (Reinwald and Kraemmergaard, 2012, p.134). The very notion that policy and outcomes progress through ordered stages of development is also dubious, with the reality being “less linear and less straightforwardly rational” (Laffin and Ormston, 2013, p.186), with local authorities in particular potentially having less uniform development patterns than central government (Shackleton, Fisher and Dawson, 2004).

Scholars interested in the democratic and interactive aspects of online public administration have generated alternative typologies of web pages, focusing on the degree of interaction evidenced. The following is a simple typology which may have utility for assessing the extent to which web pages have been developed: “retrieval, feedback or participation (or 1-way, 2-ways and 3-ways communications channels).” (Baldersheim and Øgård, 2008, p.127). Again, this type of typology has an intrinsic bias towards “more communicative” pages as desirable.

The maturity modelling approach combines two items: a descriptive coding of the present state of the technology being modelled, coupled with a normative set of assumptions about future development. Even if the latter is rejected, the coding frame from the first part is often used by researchers who are interested purely in the descriptive aspects (e.g. Manoharan, Fudge and Zheng, 2014, in their “e-governance scale”).

Ultimately, the maturity models offered in these parts of the literature are limited by their combination of empirical and prescriptive content mixed up in a single model. The “best practices” suggested by them are observational rather than robustly underpinned by properly controlled scientific work and therefore stand on assertion. Consequently, they are not a viable theoretical basis for explaining e-government.

2.3.3 *Is use of the Internet associated with improved government service delivery?*

There has been much work on service reform through the development of ICT in general (for a review, see Willcocks, 1994), including an extensive grey literature. In addition, there is a fairly extensive e-government literature, looking at the impact of information technology on the delivery of services and its utility, or otherwise, in the ongoing reform of public services (Nograšek and Vintar, 2014). However, much of the literature is technically- and service-focused, and centred on the evolution of existing practices and managerial reform. Although IT has a long pedigree in local government, it is only with the availability of the Internet that e-government has been seen as having major applications external to the government body itself (Moon and Norris, 2005, p.44). Work looking at the effect of government Internet use on service outcomes, or on cost-effectiveness, is notably scarce save for some US studies founded on surveys of government staff (e.g. Oliveira and Welch, 2013) and not independent data. This focus on surveys of senior managers as a proxy for solid performance measures is unfortunately common across the public administration field (see, for example Wynen and Verhoest, 2013), despite the obvious disconnect between the actual measurement (reported belief) and the underlying variable of interest (performance).

There are related reasons to be sceptical about the benefits when looked at through the lens of Blair-era public administration. In addition to the clear desire of the government to be seen as modern, there was a genuine desire to realise the benefits of online activity that were beginning to become apparent from early private sector adoption. Digital government reforms, then, can be seen as part of a continuum of NPM-influenced activity. Unfortunately, there is limited evidence in central government at least that costs have been reduced or services improved:

“There is a case to answer because we have shown that NPM and associated digital-age developments over much of UK central

government's operations seem to have had little dramatic long-term effect in the area for which the strongest claims were made by both advocates and critics about its likely impacts (i.e., cutting running costs) even in the country often claimed to be the originator of NPM and undoubtedly one of its most enthusiastic cheerleaders." (Hood and Dixon, 2015, p.192)

There are, by contrast, theoretical reasons to be sceptical about the benefits that the web can offer government. There is an extensive literature studying informatisation in both the public and private sectors which finds no measurable return on investment for major IT projects (Strassmann, 1997). Kraemer and King (2003) analyse the effectiveness of IT as a catalyst for reform and report their findings that it in fact has a tendency to reinforce existing trends in organisation and its effects are highly contingent – a long way away from the techno-optimists' claims that IT results in changes to more effective, focused, and cheaper organisations. Despite the acknowledgement of Kraemer and King that in the right hands technologically-supported change is possible, it is difficult to square this (empirically-driven) sceptical analysis with the positive analysis of the benefits of technology inherent in digital models of public administration like DEG. Indeed, this is at the "other end" of the spectrum of Internet enthusiasm from DEG (Andersen et al., 2010, p.564).

Similar logic could well apply here, particularly as website development in the public sector in the UK has been driven by top-down central government enthusiasm at least as much as a pragmatic business-case based rational development process.

By contrast, there is a related literature on the "productivity paradox" exemplified by Brynjolfsson and Hitt (2003) which uses alternative, firm-level metrics, to argue that if the paradox existed in any real sense in the early days of informatisation, it had largely disappeared by 1991, and that the gains from increased investment in information systems now outweigh those from the deployment of other forms of capital.

A further prominent strand of literature, summarised by DeLone and McLean looks at the success of information systems in terms of both process and outcome measures.

“We, therefore continue to advocate the inclusion of “System Use” as a critical dimension of its success measurement” (DeLone and McLean, 2003, p.27)

This requirement is a serious practical constraint in data-driven evaluation of government web use. Direct measurement of use, of a kind which would be directly comparable with census measures of web activity, is not available on a general basis and could only be approximated by survey or by obtaining usage measures from a sample of organisations⁶. One of the other key instruments, net benefits, is potentially easier to measure with organisational measures than they are by survey or other means. These constraints are discussed in more detail in Section 3.3.4.

More fundamentally, this part of the literature is focused on the processes of implementation of web technology rather than mainly on changes to the public face of organisations that result. When looking at British local government web use, the decision (to implement wide-scale use of the web) and the approach (maximal, with an aim of 100% of services) were both driven from above and so less variation would be expected than in less constrained environments.

Even if there are gains to be had from greater informatisation, it is still by no means clear that service advantages can be had simply from constructing a website in the absence of care and attention given to the services that sit behind it. Indeed, at least one paper (Gulati, Williams and Yates, 2014) treats good (national) public sector performance as a precursor to the delivery of online services and uses the latter as a dependent variable.

One major attempt to demonstrate the value of online services for productivity is *Growing the Productivity of Government Services*. Not only do Dunleavy

⁶ Or, theoretically, by instrumenting some of the Internet’s routing or Domain Name System (DNS) capacity to see web page requests in transit. This is much more problematic for encrypted HTTP over TLS (HTTPS) connections for which the target page is not available.

and Carrera conduct a Total Factor Productivity analysis of several areas of UK government activity (including customs, tax collection, and the National Health Service), but they discuss the value of using data *from* the web at census scale to analyse government online (Dunleavy and Carrera, 2013, pp.226–231). Total Factor Productivity allows a measurement of the overall productivity of organisations with disparate non-fungible outputs by weighting the output measures by the cost of their production, assuming that organisations' costs of production reflect the value they place in the consequent outputs. One conclusion of the study is that ICT use is indeed positively related to productivity.

A significant methodological issue with the approach taken by Dunleavy and Carrera is the difficulty of separating out the effects of ICT expenditures from other impact factors. For each of the unitary agencies such as HMRC, there are a large number of changes made over time which make the extraction of any one factor from the analysis impossible. Although the authors demonstrate increasing ICT investments together with increasing unit productivity, this does not demonstrate causation.

Their National Health Service analysis, on the other hand, is expressly accounted for on a trust-by-trust basis, which makes a cross-sectional analysis of input factors possible. The Ordinary Least Squares (OLS) regression models built to estimate labour productivity of health trusts feature an “ICT use” independent variable, together with others tracking management practices and the nature and location of the hospital. The conclusion is that the ICT use variable is statistically significant, though the effect size is small⁷ (Dunleavy and Carrera, 2013, p.255).

There are reasons to be sceptical of this finding. The quality of the data available to the researchers to measure ICT use is unfortunately fairly limited: the index they develop largely looks for the presence or absence of individual

⁷ The whole model, including management, hospital type, etc. only explains 16% of the variation between trusts.

features on the trust's website as a proxy for ICT use (for example "site map is provided", "pop-up web survey is provided", and "online donations to hospital charity are possible"). As there is no plausible mechanism of action for a customer-focused online site map actually delivering greater efficiency in the staff-intensive business of medical care, the index is really measuring the extent to which the trust cares about (or, possibly, is skilled in) the use of ICTs. This may, or may not, translate to better information systems and decision-making processes within the hospital. At best, the evidence is suggestive.

Furthermore, the model is cross-sectional in design and is based upon a single-year's data, so the results are sensitive to model specification error. It is possible, for example, that variations in web services reflect variations in affluence between the areas served by each trust (with more affluent areas providing more sophisticated online services to reflect higher patient Internet use) which also have an effect on clinical outcomes and hence efficiency measurements.

Although using web quality measurements as a proxy for ICT use inside organisations seems troubling, it does potentially have great potential for measuring the use of the Internet itself by the body concerned. For parts of government where there is opportunity for actual service delivery via the Internet (either in informational or transactional form) it is more plausible to see this kind of measurement as a valid instrumental variable.

It is still not especially clear to what extent local government websites are well used or, indeed, well-designed. Partly this is because usage figures for individual parts of council websites are held internally and not centrally published (though see small scale logfile-related research such as Lambert, 2013). There has been some research in various countries looking at usability and takeup (though often, as with Detlor et al. (2013), only relatively superficially and on a small arbitrary selection of local sites).

Consequently, although Dunleavy and Carrera highlight an interesting way forward for measuring IT's position as an enabler of service improvement, they

do not provide a conclusive answer. Unfortunately, taking a similar approach in local government would require reliable and comparable output metrics for every key service within local government. Such data are not available⁸, and therefore the approach is not easily scalable to the whole local government context.

2.3.4 *The web and the shape of government*

Some recent work looks at these questions for parts of central government by examining the structures and connectedness of government websites (Petricek et al., 2006; Henman, Ackland and Graham, 2014b; Hale, Yasseri et al., 2014). By gathering a snapshot of part of the web encompassing a number of different government bodies and agencies, it is possible to look both at structural similarities and differences between different agencies' provision and also look at the extent to which government is as joined-up as it claims, especially where the customer is concerned. This work necessarily makes an underlying assumption that either the structure of an organisation or its service can be discerned from a website or (more loosely) that such information can at least tell the observer something about the public-facing edges of an organisation or the face which it puts out for the world to see. This seems not unreasonable:

“The full realisation of the potentials of digitalisation of local government operations will require thorough changes at the front and back office levels of the municipal organisations as well as in roles and relationships among groups of actors. Therefore, a well designed and well functioning web site is an indication of capability of deep-going organisational change.” (Baldersheim and Øgård, 2008, p.126)

Unfortunately, this question is difficult to answer in the local e-government field. There is a shortage of large-scale data collection and systematic analysis,

⁸ The number of sets of data collected from all local authorities is now much more limited than was the case in the past (see DCLG, 2013) and do not in any way constitute a complete set of local government output measures.

with $n = 1$ case studies particularly common, especially in the practitioner literature (e.g. Socitm, Improvement and Development Agency and Local Government Association, 2003; Alfano, 2011)⁹. There is, however, an extensive literature on service reform and transformation within the field of local government studies, though with less overlap than would be expected with the broader currents of public administration scholarship. Particular focuses are on the politics and policy of change management, including the heterogeneity of reform processes and results.

Small scale and $n = 1$ qualitative work certainly has a place in the e-government literature. A textured understanding of the underlying motivations of key political and administrative personnel, for example, emerges best from this kind of work. The best case studies are insightful, locally comprehensive, and have an eye to universal lessons to draw (e.g. Margetts, 1999). In this way they are at least complementary to wider-scale quantitative analyses.

Nevertheless, the inherent availability of the government web online is a major opportunity for comparative, quantitative work. The strengths and weaknesses of web datasets are explored further in Chapter 3.

2.4 LOCAL GOVERNMENT, SERVICE REFORM, AND E-GOVERNMENT

2.4.1 *Local government structure in England*

As a unitary state, power in the UK is theoretically held centrally by the Crown in Parliament and all other governments are constitutionally subordinate. Nevertheless, there are a large number of other government bodies set out by statute with geographic scope and with enumerated powers. The system is complicated, having developed piecemeal over a thousand years. Nevertheless,

⁹ This is not a new problem. Kraemer, Dutton and Northrop (1981, p.33) note that even in the 1970s studies of technology in local government tended to be case studies of “a handful of organizations, chosen on the basis of access, convenience, or interest”.

local government bodies in England can generally be gathered into three tiers by scope of operation.

The largest scale part of local government is known as the group of “top-tier” authorities. Some parts of England are covered by unitary authorities of a number of closely related forms¹⁰, meaning that all principal local government services are delivered by one council and the top tier is also a single tier. Councils have very limited legislative powers¹¹ but these authorities have very wide service delivery responsibilities¹². In the remaining areas (colloquially known as the “shire counties”) there is a two-tier system: more strategic powers such as health, education, and transport are in the hands of the County council, while more local services are operated by the tier below. There are 152 English top-tier authorities as of 2016 (gov.uk, 2016)¹³.

The lower tier of local government proper is made up of the district and borough councils in shire areas¹⁴, which have the more local of the services covered by unitary councils in single-tier areas. District councils have services such as economic development, trading standards, waste collection (but not disposal, which is done by counties), and leisure. They are significantly smaller than county councils (and most unitaries) and each shire county has several

10 London borough councils in London, metropolitan borough councils in major conurbations outside London, district or borough councils with unitary powers in some other urban areas, and county councils with unitary powers in a few more rural areas. The City of London and the Isles of Scilly have *sui generis* councils, each of which are only loosely comparable to the other top-tier authorities.

11 There are a number of byelaw making powers available to local governments, but these are exercised only with the consent of the Secretary of State and are very limited in scope. For more substantial local legislation, councils must petition Westminster for a local Act of Parliament.

12 These include social services, primary and secondary education, trading standards regulation, provision of leisure services, town planning, libraries, public toilets, building and maintenance of non-strategic roads, registration of births, marriages, and deaths, housing, waste collection and disposal, housing benefits, and many others. It is estimated that the full scope of services amounts to more than 700 separate lines of business (Wilson and Game, 2011, p.35).

13 This comprises 32 London borough councils, 36 metropolitan borough councils, 55 unitary councils, and 27 county councils, plus the Common Council of the City of London and the Council of the Isles of Scilly.

14 Though be aware that many councils called districts and especially boroughs are actually top-tier unitary authorities. Confusingly, city status (and the corresponding “City Council” title) can be held by bodies with the legal status of unitary, shire district, or even parish councils. As with much of the rest of the British state, local government structure and naming is rife with historical exceptions.

district councils that cover different parts of their area. Top-tier councils and shire districts are collectively known as the “principal authorities”. There are 201 shire districts in England as of 2016 (gov.uk, 2016). The number of shire districts has been slowly decreasing as central government replaces two-tier county/district councils with one-tier unitary authorities.

The lowest tier is not conventionally considered part of local government *per se*. These authorities are known as parish councils (though some are called town or city councils). Despite the name, parish councils are civil and not ecclesiastical bodies. A parish council will generally cover an area of approximately one village or small town in size. They have few formal responsibilities, but some discretionary powers in the planning process and may choose to deliver some local services such as playgrounds. Not all areas of England have parish councils; some (particularly in urban areas) are “unparished”.

In addition to these general-purpose authorities, there remain a large number of specialist quasi-local government bodies with particular remits, including national park authorities (with concurrent jurisdiction over planning and some services in national parks), fire authorities in areas where the fire service is not a county council responsibility, police and crime commissioners, and some more obscure and place-specific bodies such as internal drainage boards and waste disposal authorities.

Northern Ireland, Wales, and Scotland have different systems, with a national “devolved government” operating many services delivered by central government in England, together with their own systems of local government. An excellent overview of local government in the United Kingdom and its complexities is given by Wilson and Game (2011).

This research deals almost entirely with the group of principal authorities in England (that is county, district, and unitary councils). A map of local authority boundaries is at Figure 1¹⁵.

15 The map is based on Ordnance Survey Boundary-Line and Ordnance Survey Northern Ireland Open Data Largescale Boundaries (Ordnance Survey, 2014; Ordnance Survey Northern

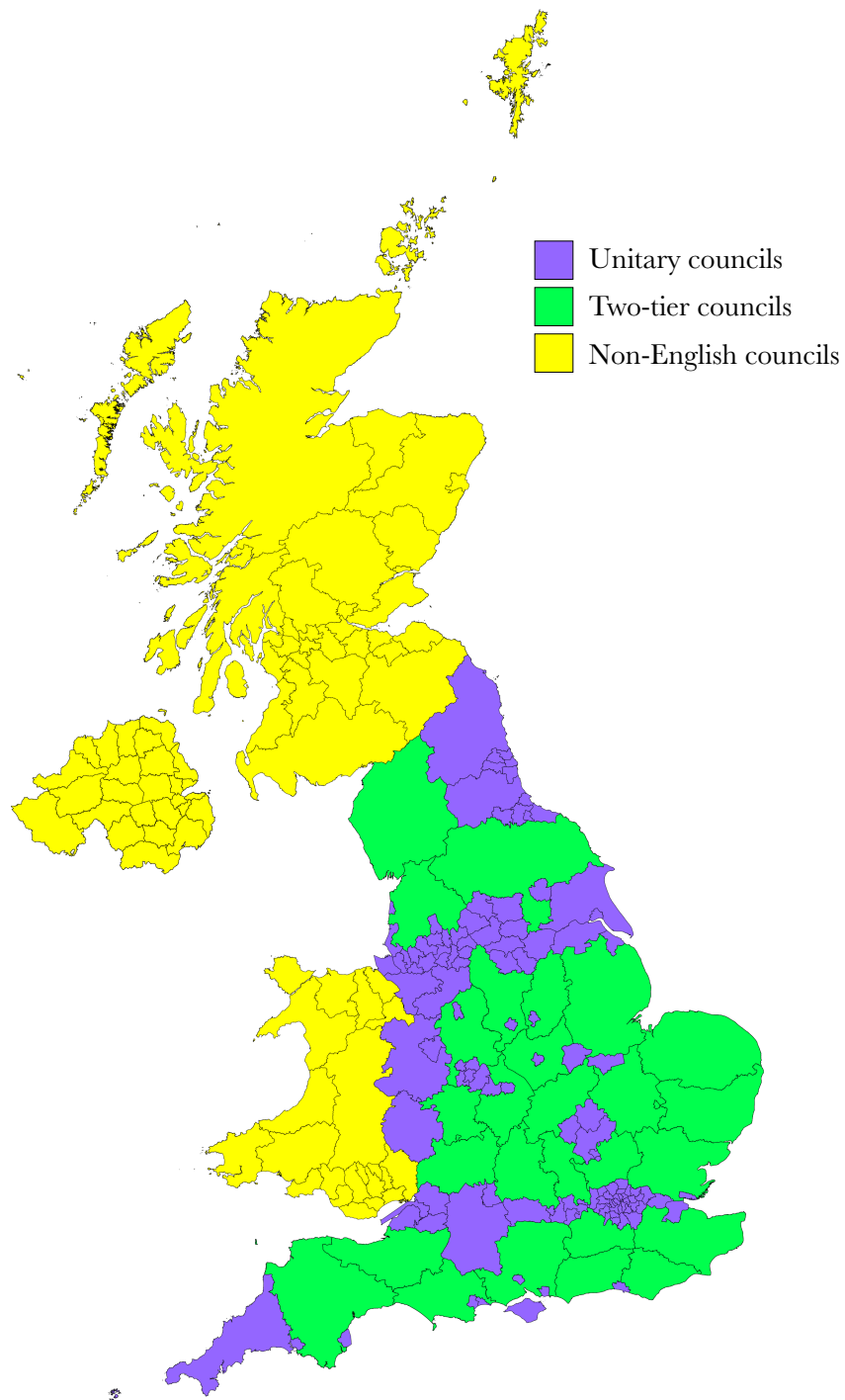


Figure 1: UK local government top-tier boundaries

Ireland, 2016, © Crown copyright and database right 2014–2016), using the boundaries current at the time of the data collection described in Chapter 4. These data are available under the Open Government Licence.

2.4.2 *Why local government is distinct and interesting*

In common with the broader e-government literature, there is a good deal of unqualified enthusiasm for the benefits of the Internet as a way of improving local administrative performance and cost-effectiveness, both in the UK and abroad (Tat-Kei Ho, 2002). This is particularly marked in portrayals of the impact of social media and online collaboration tools, which have been seen as the solution to foster collective action and allow public goods to be co-produced with and by citizens themselves (Linders, 2012). There is also, however, a suggestion that some of the politicians' rhetoric in this area is the product of a desire to cover the effects of public spending reductions rather than a genuine belief in citizen empowerment. This suggestion has been reinforced by the relative shortage of citizen empowerment activity coming from central government since the coalition took over, especially compared to the "big society" ideas focused on before the election.

2.4.3 *The policy context*

The context of local government's approach to the Internet is inextricably linked with the election of the Labour government in 1997. Tony Blair's focus on public sector reform and service delivery is attested in the 1999 paper *Modernising Government* (Cabinet Office, 1999b) which set out an ambitious plan to improve the way government operated. Local government was targeted alongside central departments, initially under the rhetoric of being an equal partner (Cabinet Office, 1999a) although the subsequent relationship was much more hierarchical. One of the key themes of Blair-era reform was opportunities for the use of new technologies. A whole chapter of *Modernising Government* was dedicated to "information age government", explicitly drawing inspiration from the private sector and espousing utopian themes: "We have seen a revolution over the past decade in the way leading companies across the world

do business” (Cabinet Office, 1999b, p.45) and “The pace of technological change is fast and exhilarating.” (Cabinet Office, 1999b, p.52).

Although the development of e-government was one of the themes of the paper, the majority of the space was devoted to more traditional public administration concerns: management reform, strategy, procurement and the oversight and motivation of civil servants. Even the e-government sections were long on ambition and rhetoric, and short on evidence and detail.

“Is it being unfair to summarise this, and everything else that has so far been said about this White Paper, into saying that what government is for is better government, and better government is government that is more efficient and effective because it is government in the age of information technology?” (Chapman, 1999, p.12)

Indeed, the whole of the Labour government’s service improvement programme has been described as drawing on no single theory of how to improve organisational importance, and indeed to be a “theoretical smorgasbord” (Walker and Boyne, 2006, p.375). The technocratic imperative to focus on “what works” did not marry especially well with theoretical coherence in policymaking.

Nevertheless, there was extensive enthusiasm for the more technocratic aspects of service improvement using the Internet. Goals set out in the 1999 white paper *Modernising Government* ranged from the ambitious but ill-defined, such as “A new target of all dealings with government being deliverable electronically by 2008” (later brought forward to 2005: Cabinet Office, 1999b, p.6) to more serious and detailed establishment of principles:

“Information age government: we will use new technology to meet the needs of citizens and business, and not trail behind technological developments. We will:

- develop an IT strategy for Government which will establish cross-government co-ordination machinery and frameworks on such issues as use of digital signatures and smart cards, websites and call centres.
- benchmark progress against targets for electronic services.”

(Cabinet Office, 1999b, p.7)

These enthusiasms spread through to the Department for Transport, Local Government and the Regions (DTLR). The coverage of “new technologies” in the subsequent local government white paper was the usual modernist model of its kind: uncritically positive, brief, and tacked on the end of a lengthy and challenging program of structural, service and managerial improvements (DTLR, 2001, §§2.56-2.59). At no point could it be claimed that e-government was the only or principal route envisaged by government for service improvements; the impression given is that the Internet is simply an integral part of the bright modern future being built, as well as a means of building it.

Local government practitioner enthusiasm for these developments varied, but most of the published practitioner literature is cautiously optimistic, while noting the scale of the management and service changes required to make the transformation to online delivery work (e.g. Socitm and Improvement and Development Agency, 2002). The rhetoric is notably less lofty than that of central government policy papers, perhaps because the scale of the challenge is more obvious to those actually required to implement it.

Central government’s enthusiasm for e-government didn’t subside. By 2005, the Cabinet Office was confidently claiming that “Twenty First Century Government is enabled by technology – policy is inspired by it, business change is delivered by it, customer and corporate services are dependent on it, and democratic engagement is exploring it” (Cabinet Office, 2005, p.3) This implicitly placed technology not just at the heart of the public sector reform process, but as a pervasive part of all government activities.

The tone of early period reform literature is one of service improvement, though the desire to improve value for money is also evident. Importantly, there was no suggestion of electronic access replacing telephone, face-to-face or paper systems, rather that they should “take their place alongside more innovative uses” of those methods (Cabinet Office, 1999b, p.46).

The focus was also very definitely supply-side, with often limited usage of the services put online. As Dunleavy et al. (2006a, p.80) note, “the government campaign presented the task as being about service availability and not about citizen usage of services”. This omission seems strange; an unused service online is costing development resources and providing no public benefit. A consequence is that there is no government-released data about e-government *use*, unlike the copious Comprehensive Performance Assessment (CPA)-driven metrics of everything else the government considered a public priority.

Although websites were mentioned, early ambitions were modest: “To bring about a more coherent approach to the use of web sites for giving information and eventually delivering services” (Cabinet Office, 1999b, p.50). Other early aspirations quietly disappeared, such as the “committment[sic] for people to be able to notify different parts of government of details such as a change of address simply and electronically in one transaction.” (Cabinet Office, 1999b, p.33) and wider use of digital signatures and smartcards (Cabinet Office, 1999b, p.50). An early focus on interactive television was also replaced by the use of the Internet as the principal communication method, as Internet penetration increased and success with interactive TV proved difficult.

Movement on e-government and Internet work was encouraged by various departments and agencies: the Office of the Deputy Prime Minister (ODPM) and its successor the Department for Communities and Local Government (DCLG), the Improvement and Development Agency for Local Government (later part of the Local Government Association), and the local government IT officers’ body Socitm. The Local Government Online (LGOL) project was given £670m by central government between 2000 and 2005, to help councils to make service improvements and cost improvements (Local Government Association, 2014, p.13).

Later, government priorities moved towards efficiency instead of service improvement. The “Gershon agenda” was laid out in a paper by Sir Peter Gershon (Gershon, 2004) and aimed to rein in the increasing levels of spending

on government services. Justifications for e-government projects subsequent to this have needed to have a greater focus on cost-savings or measurable service improvements for the same price.

From 2008 onwards, government pressure has been oriented overwhelmingly towards cost reduction. Fiscal settlements for local government have been tighter than for almost all other elements of the public sector; combined with the inability of authorities to control their own tax levels (progressively weakened since the 1980s) this has led to heavy downward pressure on local government budgets (23% reduced in real terms in England between 2007-08 and 2014-15¹⁶). As a result, e-government projects are increasingly justified on cost-reduction grounds and not on improvements to service or broader measures of increased productivity.

The Better Connected (BC) series of reports from Socitm (Socitm, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013a) provide a useful annual update on the concerns and focus of the local e-government practitioner community during the period under study. These will be discussed in more detail in Section 2.4.6.

In addition to the ratings of websites which are the main purpose of the reports, the accompanying advice and (particularly) the decisions made about which parts of the council web are important enough to be measured in any given year give an understanding of the local e-government policy context.

“Market research commissioned during 1998 by the Cabinet Office shows that electronic government services offer at least four potential benefits:

- Improving the speed of carrying out transactions
- Providing greater access and convenience
- Offering more flexible options in delivering services (e.g. extended hours of service)
- Bringing services closer to the public

(Socitm, 1999, p.9)

16 Total English local authority spending outturn for 2007-08 (excluding specific grants) was £101.8bn nominal (£118.1bn inflation-adjusted to 2014-15 values) (DCLG, 2010b) and in 2014-15 was £91.2bn (DCLG, 2016).

Note that this reflects the early Labour priority of improving service quality, rather than reducing their cost. In principle, these propositions admit of an empirical analysis: did the development of e-government actually deliver such advantages?

A useful sector-wide review of recent attitudes by local government to the Internet and digital transformation was published by the Local Government Association in 2014 – albeit as a curated series of case studies with an exhortation to “share best practice”, not any kind of structured evaluation of effectiveness (Local Government Association, 2014). By this point, enthusiasm has moderated sufficiently to allow an observation that “‘digital’ [...] is not a panacea for all ills, and should complement other approaches such as demand management, lean and systems thinking, and collaborative procurement” (Local Government Association, 2014, p.9). An unkind observer would see in the list of other approaches several other concepts offered as magic bullets for service improvement over the years, and which are likewise somewhat lacking in systemic evaluation for effectiveness in the sector.

A digitally-optimistic approach to service transformation continues. The GDS claims to have saved £1.7bn in 2014/15 through better IT service procurement, the delivery of gov.uk and digitally-enabled service transformation (Foresheew-Cain, 2015). Whether this reflects an overall saving from IT in government, or simply positive presentation of the successful part of a mixed-success programme, is unclear.

2.4.4 *Informational, transactional, and democratic uses of the Internet*

There is a significant division within the existing e-government literature between work looking at the delivery of governmental services using online tools, and work looking at opportunities for electronic democracy and citizen engagement. This research falls firmly on the former side, investigating service delivery uses of the Internet. In truth, there have been few attempts in England

to use Internet approaches for democratic purposes, and there is little published work showing successful attempts – even as single case studies. This reflects the prevailing focus of the government over the same period:

“Despite New Labour’s stated desire for ‘democratic renewal’ [...] and its apparent interest in ‘e-government’, both the Blair and Brown administrations followed their Conservative predecessors in their ambitions primarily to improve local *services*.” (Ellison and Hardey, 2014, p.22, emphasis in original)

Within the broader e-government literature, there have been helpful distinctions drawn between ideas of e-participation, e-politics and e-government, e-services, and e-management (see, for example Chadwick and May, 2003, on “managerial”, “consultative” and “participatory” ideal types). This adds clarity to an otherwise huge and confusing topic. There has been some work done on the possibilities for the Internet for reinvigorating democracy (e.g Parvez, 2006). Indeed, many local authorities are actively using social media, though not always with strongly democratic intent (Ellison and Hardey, 2014). Ellison and Hardey also explore uses of Internet technologies for “double-devolution” of power through local authorities to local citizens. Nevertheless, the pursuit of effective service delivery is quite a separate challenge. It is the questions asked by the e-government field, such as how governments can use the Internet in service delivery, how its use has changed in recent years, and what opportunities exist for reinventing government, that this thesis is concerned with.

The key expressed goal of e-government work in English local government has been towards the delivery of transactional services. The original target by the Blair government was to put all services online; the key drivers of cost reduction and service improvement were seen as best realised through the conducting of transactions online (see the BC reports in Section 2.4.6 below).

This drive has been subsequently reflected in the central government work done by the GDS, which concentrated initially on the delivery of high-volume transactions and only seems to have taken on the task of delivering regula-

tory and back office information in order to support the quicker abolition of separate departmental websites.

It is by no means clear that this focus on transactions is justified. Meijer and Thaens (2009), for example, make a coherent case for the high value of making government information available to citizens. The value of informational uses for local governments is discussed in more detail below in the context of DEG, in Section 2.4.7.

2.4.5 *Published accounts of local government Internet use*

There is no published academic overview of local government Internet use in the UK during this period, despite the importance of e-government to narratives of service performance. Indeed, the overall amount of work done on local government Internet use in England is not high. The local government studies and public administration literatures are quite distinct: while the public administration literature largely neglects local government, the local government literature has even less focus on IT.

The most prominent narrative of English local government Internet use since 1997 has been a general and ill-defined optimism that technology will in some way allow government to be better, without going into too much detail about the hows and whys. For example, in the middle of a lengthy and wide-ranging article on improving government, the former Cabinet Secretary simply handwaved the idea that “digital” would improve services:

“There is little doubt that further savings can be made while improving public services by taking on board the latest developments in digital delivery.” (O’Donnell, 2013, p.382)

Similar enthusiasm is present too in the attitudes of local government. The Local Government Information Unit (LGIU), a local government-owned thinktank, put out a paper arguing for the benefits of such ill-defined concepts as “big data” and “the cloud” as means of fundamentally transforming services

(LGIU, 2014). The paper was sponsored, as such papers often are, by one of the big outsourcing vendors who benefit from selling technology and consulting to local government (in this case Hewlett Packard).

Arguably, some of these things have been achieved: (some) policy is certainly inspired by seeing what other organisations are doing with computers, and offices would cease to function without electronic systems. Nevertheless, it is clear that technology alone does not deliver business change (Duivenboden and Thaens, 2008) and electronic democratic engagement as an aim of central government is as marginal as ever.

Despite this, there is a significant literature (e.g. Bonsón et al., 2012) measuring purely technical aspects of municipal web use (such as the presence of sitemaps or particular kinds of content) and inferring democratic intentions from it. This is not a valid approach, given that technology is neither sufficient nor, perhaps, necessary, for democratic or service transformation.

In fact, the amount of empirical e-government work supporting this kind of democratic transformation through technology is very low. As Norris and Reddick explain: “We do not cite empirical works that validate claims made in cyber-optimist writings because, after an extensive review of the e-government literature, we have not been able to find any” (Norris and Reddick, 2013, p.167).

This lack of empirical evidence for optimist claims about technology in government is not new. Kraemer, Dutton and Northrop (1981) found very uneven returns from local government informatisation in the 1970s United States, long before the Internet was seen as the enabling mechanism. They showed that the relationship between technology introduction and performance in public authorities is complex and the results highly variable.

Similar themes persist through the academic IT literature in the 1980s and 1990s. Researchers reported that the benefits of informatisation work were often overestimated and the gains difficult to realise (Strassmann, 1990; Willcocks, 1992).

Although there are a number of individual studies on particular aspects of local government and the Internet (such as Cullen and Houghton, 2000; Yannas and Lappas, 2007; Rufin, Medina and Sánchez Figueroa, 2012) these are often focused on other countries, one-authority case studies or focused on e-democracy rather than electronic service delivery. Quite a bit of work (e.g. Ganapati and Reddick, 2012) is focused on the US state governments, which do not map well onto the British system: many of their functions are those which would be carried out by central government in the UK, and even the smaller state governments are much bigger than the largest UK local authorities.

Internet-focused work is in any case generally scarce¹⁷. There is some literature on the co-production of local public services (e.g. Parks et al., 1981; Pestoff, Osborne and Brandsen, 2006) which focuses on third-sector organisations (possibly made up of groups of citizens) as co-agents in the production of public services (Brandsen and Pestoff, 2006), but not on the Internet as a medium for new ways of co-producing informational goods. A comprehensive overview of the history, shape and extent of local government Internet use has not yet been written. This research will go some way to fulfilling this need.

Nevertheless, we can learn from the more extensive work done in e-government and related fields overseas. Theorists and indeed governments have speculated on the correct categorisations of different kinds of government transactions (Carter and Bélanger, 2005, p.6) and the best ways to encourage take-up of them. Carter and Bélanger, for example, use Davis' Technology Acceptance Model and Rogers' Diffusion of Innovation theory as the basis of a survey-informed model explaining why citizens adopt e-government services.

Unfortunately, much of this work is either purely theoretical or relies on surveys of citizens to say whether they have visited or intend to visit government

17 For example, a search of the Social Sciences Citation Index for 'Title=(“local govern*” AND internet) OR Title=(“local authorit*” AND internet)' generates only 6 results in total, some of which cover democratic uses of the Internet, not administrative ones.

websites and use their transactional features (e.g. Bélanger and Carter, 2008). In principle, this data should be available (traffic data is collected by council IT teams) and so the models could be empirically validated. Despite this empirical tractability, it is unclear whether developments in e-government are, in fact, causally related to increased performance (Heeks and Bailur, 2007).

To bring order to these pile of disparate studies, it is helpful to consider the focus of the papers. Much of the literature looks at the perceptions of local managers as to e-government success, from the classic works (e.g. Danziger et al., 1982) to recent empirical work (e.g. Andrews, Boyne and Walker, 2006; Ganapati and Reddick, 2012; Norris and Reddick, 2013). Although in principle this survey- and interview-based work is often trying to understand the impacts of technology more widely than simply its impacts on the officials supplying the information, it is inherently a provider-centric means of study. Less satisfactory papers rely entirely on survey responses from managers and do not spend time probing whether their perceptions in fact correlate with any reality outside the agency.

This provider perception-centric approach, although limited, does have real merit. The discourses that have surrounded the development of local e-government may differ from the perspectives of users or whatever passes for measurable external reality, but they have played a significant part in the choices made by practitioners in developing e-government. When perceptions are widely shared, they have an important reality of their own (Yıldız and Saylam, 2013).

A second approach is to attempt to measure external reality directly, to look at websites either as the traces left behind by local authorities online reflecting some underlying organisation (e.g. Holmberg, 2009, looking at Finnish local authorities) or trying to measure quality using some kind of objective assessment, be it expert assessment (as in the *Better Connected* series, Socitm, 1999, et seq.) or by iterative enquiry (Irani et al., 2005).

This approach is fundamentally harder to do at a whole-sector level than for individual councils or agencies, simply because there is so much published by local authorities online and each have their own approach and their own websites. This is equally true of central government in some senses, given the number of disparate agencies there. Previous work has all tended to concentrate on a few websites.

What is largely missing from the literature is the third perspective, that of the users of the service. This is highlighted by Yıldız and Saylam (2013), a discourse analysis based upon Turkish newspapers. The researchers observe that user perception is as important as reality when trying to use the Internet as a tool to improve government, as an unpopular service will not be used, however objectively competent it is, or how successful its managers believe it to be¹⁸.

One reason why questions of local government Internet use remain problematic is the difficulty of collecting data sufficient to demonstrate the wide applicability of findings. The technical studies discussed above don't aim to answer fundamental questions of public administration, while the more small scale studies necessarily do not fully generalise. Likewise, studies based on surveys of government officials can give insight into intentions, not into the success or failure of whole approaches to government reform.

Consequently, more comprehensive datasets are needed. For some research questions, such as those focused on the reality of government behaviour online, sets of trace web data are suitable, as advocated by Dunleavy and Carrera (2013). For others, good quality probability sampled or full-population data of a more traditional kind are suitable. One such dataset, essentially unused for academic purposes but of good quality, is the Better Connected (BC) dataset.

18 This has interesting parallels with the GDS goal to produce "Digital services so good that people prefer to use them" (Government Digital Service, 2014).

2.4.6 *Socitm's Better Connected*

Better Connected is a good quality empirical data series for local government use of the Internet published by Socitm, the professional body for IT managers in local government. It comprises a series of annual assessments since 1999 of each local government website across the UK and a few sites from related sector organisations who subscribe to Socitm's *Insight* service (Socitm, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013a). Its purpose is to identify and share good practice to support improvement, reflecting the importance to Socitm and its members of online customer self-service.

The data are an expert assessment of the quality of each website and its customer experience. The methodology has developed over the years that the data has been collected, as have the reporting categories, but scores are awarded based upon the ease of completing particular tasks which vary by year¹⁹, conformance with a set of usability criteria and also on an overall reviewer rating (Socitm, 2013a). Although the methodology is intended to be as objective as possible, the authors note that "Website assessments are not an exact science, however much we try to be analytical and objective" (Socitm, 2013a, p.10). Indeed, what constitutes "quality" in the context of government websites is fairly contested even in the academic literature: "In the research literature, website quality is not a consistently defined term either, even though it is often used to report or evaluate websites" (Sorum, Medaglia and Andersen, 2009, p.200). The BC approach, of having all websites judged by an expert panel rather than relying on an automated sift to select the better ones for manual study, is amongst the more sophisticated of those methods.

Access to the data is restricted to *Insight* subscribers (and in later editions non-subscribers are not in any case identified in the printed reports, presum-

19 Since 2012 the assessments have explicitly focused on "top tasks", but did so implicitly before then.

ably to limit the utility of samizdat copies, though all councils are assessed and included). As a result, the BC data has not been previously used in an academic context, despite its obvious utility in studies of local e-government²⁰. With the kind permission of Socitm the full set of data is, however, available for use in this research. This allows the use of a quality of development metric which is independent of the researcher's own judgement.

Socitm's supporting analyses focus heavily both on individual case studies but also on the prerequisites for successful e-government delivery. Chief amongst these is the development of a strong digital capability. This involves having substantial in-house levels of expertise (c.f. the NPM tendency to out-source skills) and a great deal of nuts-and-bolts hands-on work around the provision of a solid technical foundation and a lot of management work to establish digital as part of staff roles and ensure development and accountability of the website being constructed.

This is particularly relevant to the context of this study, because it situates the difficulties of electronic delivery strongly in the existing management and administration of an authority, and not in an abstract notion of digitally sweeping aside old problems. Indeed, this is one of the key themes that emerges from studies of the pre-Internet phases of government informatisation, where early hopes of solely transformative technology were thwarted by cultural and management issues, along with insufficient attention paid to details of process and adoption.

2.4.7 *Digital Era Governance: a good fit for local government?*

There are reasons to suspect that some of the factors identified for poor central government IT performance in Dunleavy et al. (2006a) may not apply to the

²⁰ A few researchers, such as (Senyucel, 2007) record reading a few of the published reports as part of reviews of the practitioner literature, but not making use of the judgements contained therein.

same extent in local government as they do for the central government in Britain. Leaving aside the question of whether we have moved past NPM (Lodge and Gill, 2011), outsourcing has been pursued at quite variable rates within local government, due to differing local political cultures. A greater proportion of the existing IT expertise has also probably been retained by local government, potentially making contract management and re-contracting more effective. Similarly, the variety of size and approach between local authorities has led to greater heterogeneity in contracting, with smaller contracts and a broader supplier market than the massive contracts common in central government.

A broader concern is in the differing opportunities for savings and service improvements through digitisation that exist in local government as opposed to central government agencies. Unlike large information-transaction departments such as HMRC, the largest and most expensive parts of local government services are delivered face-to-face to local citizens in such a way as to make it difficult to see how these could be successfully transformed by the application of online methods.

This is reflected in the profile of visits actually seen by local authorities participating in Socitm's Website Performance Survey (wps), where 54% of web visits are for information. By contrast, the various categories of transactional activity were much less common²¹.

It has been assumed from the very beginnings of informatisation that cost savings occur "mainly in large-volume, highly repetitive information-processing tasks" (Kraemer, Dutton and Northrop, 1981, p.57). Indeed, there is a drive towards "top tasks" as a central concern of local government web teams, challenging the long-tail nature of current local government web provision (Socitm, 2013a).

English local government, however, only has a few high-volume information-processing tasks: finance departments' accounts received and payable, includ-

21 6.9% making payments, 1.8% making bookings, 5.8% reporting problems, 5.3% requesting a service, 7.5% making an application, 18.8% other/ nothing in particular (Socitm, 2013b).

ing council tax collection, is probably the highest volume. Other tasks, even those of a routine kind, occur in smaller volumes as a result of the large number of disparate services that make up the local government portfolio: see the list of “top tasks” identified by Socitm (2013a) as particularly important for council web teams’ future development in Table 1²².

This list is interesting from a DEG perspective, because a significant number of the services on it are not only not recommended to be transactional at present by Socitm but are intrinsically informational in scope: “Find out how to apply for housing” is catered for, not “Apply for housing” online. The opening hours for the council tip and the local swimming pool are inherently informational rather than transactional interactions. The disposal of refuse cannot actually take place electronically, any more than a swimming session can. These tasks are not only local in scope, but also inherently delivered offline.

This is not to say that the informational and transactional parts of council services are completely divorced from each other. Indeed, many services have a significant informational need underlying a transaction or face-to-face transaction. School admissions, for example, are a transactional service (sometimes delivered online, but also via a paper application form) which is backed up with a wealth of information on school performance available online. But it is likely that the having easy access to the school performance information online is more valuable to both the citizen and the council than the automation of the application form transaction itself.

Although these are among the more frequent reasons to contact your council, it isn’t clear that large cost savings will amount from, say, online objecting to planning applications, given the relative infrequency with which a given citizen will want to do such a thing: the implementation costs of digital services can

22 It is to be noted that these do not necessarily overlap with the tasks which citizens are most likely to want to carry out. Starting a business, for example, may be top in importance but is unlikely to be top in volume.

Task	Category
Pay parking fine	Transactional
Find out how to apply for housing	Informational
Find out opening times for council tip	Informational
Apply for older person's bus pass	Transactional
Apply for a council job	Transactional
Reserve a library book	Informational
Find out about getting help at home	Informational
Find out about my councillor	Informational
Find opening times for local swimming pool	Informational
Find help in starting new business	Informational
Report pothole	Transactional
Dispose of old bed	Transactional
Find out about free nursery places	Informational
Object to a planning application	Transactional

Table 1: Top tasks from Socitm (2013a), with their categories

be expected to be mainly fixed, while the benefits could generally be expected to be in proportion to usage. For low-usage services, the implementation costs could readily be greater than the benefits.

The largest services, such as schools and social services (the two largest components in local authority budgets) have the further difficulty of inherently relying on face-to-face interaction with clients and on-the-spot judgements by expert staff. In that sense, they perhaps more closely reflect services such as the police than they do a traditional public administration. In each case, it is clear that there is a value to having information on the web for service users and intermediaries (such as carers) but that this supports rather than replaces the in-person delivery of services.

This in itself is liable to reduce both the impact of the digitalisation which is one of the three key themes of DEG (Margetts and Dunleavy, 2013). Although it is possible to see some opportunities for Internet-supported service transformation, such as a joined-up holistic arrangement for customers to interact with the service providers, and efficiency improvements enabled by the use of mobile technology by staff, short of radical developments in educational

technology and practice savings in the schools budget are constrained by the requirement to employ teachers to put in front of children.

Savings and service improvements from the strictly informational side of web development, though, seem easier to realise. By putting up and signposting properly the opening hours for the household refuse centre (“tip”), the number of telephone calls to be handled by human operators could very plausibly be reduced. This does, perhaps, call into question the current drive towards transactional services as a goal of government web policy, at least for local government.

Other factors perhaps support the hypothesis: local authorities have been pursuing the “reintegration” and “needs-based holism” argued for by DEG (Margetts and Dunleavy, 2013, p.6) for many years, as a way of tackling a perceived “silo mentality” (Rashman and Radnor, 2005). On the other side, there is no doubt that IT in local authorities is seen as a support service, rather than as part of the strategic centre, despite the local government reform agenda. These many differences in practice both within and between local authorities raise the question of whether DEG fits the experience of local government, or whether a different conception of service reform may be needed. Empirical evidence for this is particularly difficult to find in the public administration literature, with most arguments on the subject strictly theoretical (e.g. Laffin and Ormston, 2013, which argues against DEG in local government by analogy with the impact of austerity cuts).

2.4.8 *Homogeneity and heterogeneity in local government and its web provision*

A key concern that does arise from the local government literature is that of homogeneity vs heterogeneity. Local government’s claim to political and organisational autonomy rests in part on the idea that different places and their residents have different needs and aspirations – as Clarke and J. Stewart (1994, p.163) put it, “local ‘choice’ and local ‘voice’”. There has been a tension

in recent years between central government's desire to see local government offer local vision and leadership, and its desire to tightly control the delivery of local government services and secure value for money (Martin, 2002).

These competing challenges, towards diverse solutions for local needs and the rapid adoption of national best practice, underpin much comparative research in local government studies. Is local government best seen as a collection of comparable service delivery organisations, or an archipelago of disparate organisations pursuing different goals in (legitimately) different ways? There is a current debate within the policy community as to whether the model of a single centralised local government website would be desirable (Copeland, Fink and Scott, 2014) or ill-conceived (Socitm, 2014). The terms of the debate are on exactly this subject: whether the structure of local government is sufficiently homogeneous to map onto a single framework of devolved national provision, or whether local sites would provide a direct democratic engagement with councillors (and not cost substantially more).

In e-government terms, can we model the sector as a homogeneous group, or do the variations in policy, scope and activity preclude a local government-wide approach to getting the best from the Internet?

2.5 BIG DATA, POLITICAL SCIENCE METHODS, AND POLICYMAKING

There has been an increasing focus in policy circles in recent years on the possible benefits to be obtained by using so-called "big data" approaches to gain useful intelligence from the mass of transactional data which government handles (Dunleavy, 2016). Similarly, there has been a move amongst researchers to exploit this data to answer those academically interesting questions which are tractable by quantitative and textual methods and for which trace and transactional data provides new opportunities.

The questions which arise for answer in this research can be addressed using a number of large-scale quantitative methods. More than this, though,

they permit the evaluation of the value of using these new approaches in a public administration focused policy and academic context.

These issues, and the literatures that discuss them, are discussed in more detail in Section 3.3.

2.6 CONCLUSIONS AND RESEARCH QUESTION

Small scale case studies and surveys of government staff suggest that there could be substantial gains in instrumental efficiency possible as a result of greater use of e-government (Andersen et al., 2010). Nevertheless, hard evidence that the attention paid to the Internet by English local authorities has paid overall dividends is much harder to come by.

The consequence of the scope of the local government literature for our theoretical understanding of e-government is two-fold. Firstly, local government perspectives are poorly joined up with wider e-government narratives within the UK literature at least. This is perhaps because of the unfashionable nature of local government as an object of study, coupled with the relatively small community of scholars studying online government service delivery at all. The Socitm studies are admirable, but they have a narrow focus on current priorities at the time of each survey. Their very focus on local government and lack of engagement with the wider scholarly literature also limits the extent to which they help join up with central government-focused accounts.

Secondly, the wider e-government literature systematically neglects the experiences of local governments in implementing and delivering online services as part of public sector reform efforts. This is doubly disappointing, as local government potentially provides an excellent environment for comparative public administration study. Multiple semi-independent agencies are operating to similar priorities and heavily monitored by central government, allowing the creation of datasets which are both wide and deep. This can complement the inherently fuzzy quantitative analyses of comparative central government (as

seen in the fsQCA in Dunleavy et al., 2006a) and the richness of interview-based and other qualitative methods.

It is apparent that there has been little effort by researchers in public administration to understand the overall scope of Internet use by local authorities. This is particularly disappointing given the claims that have been made (and are continuing to be made) by practitioners and academics for the merits of Internet-facilitated approaches to service improvement in the public sector.

The overall research question for this thesis examines whether local e-government has met these expectations, of DEG and of its practitioners, in England:

Research question: Do the form and impact of local e-government reflect the expectations of Digital Era Governance and of local e-government practitioners?

The overall research question is broken down into five sub-questions, each focusing on a part of the wider whole.

Sub-questions 1 and 2 are motivated by the need to understand the nature of local government web use as actually implemented and understand the extent to which we can learn about governments from their online presence. These treat e-government largely as a dependent variable. Following on from the discussions in Section 2.2, they examine the shape, scope, and use of the English local authority web, the relationship between the types and sizes of local authorities and their online presences, and the particular distinctivenesses of local (as opposed to central) government structure and content online. Importantly, answering these sub-questions will provide a better understanding of how closely local government empirically approaches the DEG ideal type.

Research sub-question 1: What is the structure, shape, and connectedness of the English local government Internet presence and what are the relationships between different bodies? What can this tell us about the structure of government in the digital age?

Research sub-question 2: What is the content of the English local government web? What can this tell us about local government and its approach to e-government?

Sub-question 3 directly addresses the uncertainties around the effectiveness of digital approaches to service improvement outlined in Section 2.3.3. It treats e-government as an independent variable, and assesses the extent to which changes in e-government use are associated with improvements in performance (as digitalisation allows the provision of better-quality services), cost-effectiveness (as an increasing provision of services and information online allows for a reduction in staffing levels for a similar level of quality), or customer satisfaction (as citizens find the experience of accessing services online, or the effects of improved performance or cost-effectiveness, more satisfactory than less-digitalised arrangements). This, together with the findings of sub-questions 1 and 2 provides the empirical underpinnings for a discussion of the DEG quasi-paradigm in the context of local government. Sub-question 4 assesses whether DEG is a good fit for local government in theory and in practice.

Research sub-question 3: What is the relationship between council administrative performance, cost-effectiveness and use of the Internet? What explains this relationship?

Research sub-question 4: To what extent does the experience of English local government fit the Digital Era Governance quasi-paradigm?

Finally, sub-question 5 broadens the scope of analysis to look at the wider use of large scale web datasets in political science introduced in Section 2.5 and explored in more depth in Chapter 3.

Research sub-question 5: What opportunities for future political science research exist in large scale analysis of web content?

Sub-question 1 is addressed in Chapter 4, sub-question 2 in Chapter 5, sub-question 3 in Chapter 6, and sub-questions 4 and 5 in Chapter 7. Chapter 8 summarises the conclusions as a whole.

NEW METHODS FOR ONLINE PUBLIC ADMINISTRATION RESEARCH

“[T]here are some obvious limitations in relying on practitioner testimony, ‘expert surveys’, or population-wide attitude survey data to assess the cost-reduction and quality-improvement claims of NPM or its numerous competitors”

— Hood and Dixon (2015, p.194)

3.1 INTRODUCTION

Political science has traditionally relied on sample-based data collection methods: surveys of individuals and organisations, use of whatever samples of data have been collected by official statisticians and others for governmental purposes, and more informally sampled qualitative approaches such as interviews and case-studies.

For many areas of interest, data have been hard to come by and insufficient for really robust conclusions to be drawn. In electoral studies, for example, the best data source available are the results of elections themselves. These generally happen only every few years; it is difficult to draw effective explanatory and predictive models on the basis of such data, especially as the population of voters under study and their attitudes change significantly as the decades-long data collection process is ongoing.

Social science attitudes to large-scale data are changing, though. There is an increasing development of wide scope, data supported, and computationally intensive research methods under such headings as computational social science¹ (Conte et al., 2012), Big Data (Monroe, 2013), and data science. These new approaches are multidisciplinary in nature, drawing people, tools, and insights from political science, physics, sociology, computer science, and other fields. This is explored in more detail in Section 3.3 below.

Government and politics are increasingly happening online; social movements are often expressed online if not largely formed there, and the study of digital government is now in many ways the study of government itself. As such, new methods for studying digital politics, such as the web-scale census approaches advocated by Dunleavy (2010a), help to lay the foundations for our better understanding of political phenomena.

Government is increasingly happening online; the study of digital government is now in many ways the study of government itself. As such, new methods for studying digital government, such as the web-scale census approaches advocated for by Dunleavy (2010a) and further developed in this research, help to lay the foundations for the future study of public administration. It is in the development of these methods, as well as in the substantive findings, that this research project provides value.

This chapter outlines how full-population approaches can help to answer the research question posed. It sets out some of the challenges posed by these approaches when applied to political science research, and offers some new solutions to these difficulties. In particular, solutions to some of the issues intrinsic to running large-scale web crawls are offered, along with some new software tools to handle the large volume of data created by them and to extract useful information from it.

¹ Focusing on agent-based approaches to understanding complexity

More detailed accounts of how these new methods are applied to the research sub-questions are in the subsequent chapters; the aim of this chapter is to be a high-level overview and justification of the approach, and to demonstrate the value of new methods for future political science research.

3.2 RESEARCH CHOICES AND OPPORTUNITIES

The research sub-questions being tackled have been established from the literature review and set out in Section 2.6. Each of these questions could be tackled in various ways. This section therefore sets out to justify two choices: English local government as a scope of study, and the use of webometrics and full-population approaches as a principal research direction.

It argues that English local authorities have both practical and theoretical features of interest, and (without making formal claims to out-of-sample generalisability) can provide insight into wider questions of e-government use and effectiveness. It also argues that collecting full-population web crawl data is a novel, viable, and powerful approach to answering the questions posed.

3.2.1 *Case selection and generalisability*

The field of analysis for this research is local authorities in England. This choice of case has both practical and theoretical purpose.

The choice of coverage of local rather than national governments for this research is purposive. The nature of sub-national administration is that there are a number of subunits in each country which are far more comparable than are national administrations with their different datasets, scope of action and constitutional restrictions. By focusing on local government entities, it is possible to assemble a true population dataset in a way which allows quantita-

tive as well as qualitative comparison without some of the compromises which are inherent in comparative national analyses².

A second substantial advantage of focus on local governments is their relative unfashionableness. As mentioned in Chapter 1, local government is a major part of the state, overseeing over £90bn/year of services in England alone. In addition to this importance, the constraints under which local governments operate are quite different from those of national government: in England they are creatures of statute, subordinate to a distant cabinet minister and without significant control over revenues (Wilson and Game, 2011) and yet with a sizeable degree of local discretion around how to operate services. The resulting variations in structure, political attitudes and service configuration leads both to commonality and variation in local government web provision.

Practically, English local government has the valuable quality of being pushed towards greater use of digital technology as part of a strong central push, which was accompanied by one of the most comprehensive and intrusive inspection and performance monitoring regimes in the world (J. Stewart, 2003). This means that an unusually large volume of uniform performance data is available (Revelli, 2010) to accompany the web analysis, making possible the performance analysis in Chapter 6.

Theoretically, there are few reasons to suspect that the experience of English local authorities are categorically different from those of other developed countries with reasonable Internet penetration. Although the size of authorities in England is more uniform than those of, say France (there is no single local authority as large as Paris, and none as small as those communes with handfuls of residents), they shouldn't be radically different than those of a similar size.

2 Although there are inherent differences between the different parts of local government systems, these are less than those between different central government bodies. In particular, the shared national political environment and oversight arrangements mean that local authorities are generally focusing on similar issues and have many common data collection standards, while central governments are often out of step and only certain data are comparably reported at international level. Some of the differences that do arise (for example in political culture, size, and demographics) can be more readily controlled for either using full-coverage Office for National Statistics (ONS) data, or by using panel-based approaches as in Chapter 6.

For the universal theoretical propositions being tested, such as those from Digital Era Governance (DEG), the very specific focus of this work is an advantage; by testing the empirical claims of the theory under different conditions from those originally envisaged, the universality of the theory is challenged. Indeed, it is very much in the desire to extend existing theories of electronic government to the local level that the usual concerns about generalisability are to some extent reversed.

It is obvious that not all the findings from this research will generalise beyond the particular authorities studied. To some extent, this is unproblematic; as identified in the introduction, local government in England is a hugely costly and important area of government and worthy of study in its own right. Nevertheless, it is clear that many of the underlying effects being studied are not geographically constrained. Although the particular history and environment of local government in England is specific, the affordances of new technology and the basic structure of local government are similar in many parts of the world.

3.2.2 *Webometrics and full-population approaches*

A key choice in this research is the decision to use full-population web crawling as the primary source of data rather than sample-based methods or smaller-scale qualitative approaches. As discussed in Section 2.4.5, the different methods choices made by public administration scholars result in understanding of different parts of an overall picture.

Local government Internet use might traditionally have been assessed by doing a document analysis of local government e-government rhetoric, or attempting to disaggregate the costs of Internet work from wider budgets. Survey approaches are also common, as are the assessment of websites for the presence of particular features of theoretical interest. The time-consuming nature of such approaches strongly incentivises small-scale studies, which would

not necessarily generalise to the wider population or add to the permanent stock of disciplinary knowledge. With the aid of web content, however, direct measurements can be made of many of the features of interest.

Web crawls have several key benefits for collecting research data: they gather direct traces of online activity, they allow the use of full-text methods such as automated content classification and corpus linguistics methods, and they allow direct analysis of the structure of links and connections between online content.

The research sub-questions outlined in Section 2.6 are not all straightforward to study from a traditional political science perspective. Research sub-question 1 in particular is problematic, as trying to view the shape of a network from a sample is profoundly difficult: as the network becomes sparser, the number of edges falls off faster than the number of nodes. Political scientists have certainly studied the interactions between small groups of agents using a network approach, but this doesn't scale to the whole system studied here. To sample a handful of local authorities and study them by hand would be making assumptions about the homogeneity of local government activity which would prejudice some of the analysis inherent in the question.

Fortunately, it is possible to collect something approximating the full scope of public-facing government presence on the web³ and analyse it holistically. This approach, taking the traces left by government activity online and directly analysing them, provides opportunities for both looking at government as a whole (by looking at high-level summaries and network analyses) or at specific actors and phenomena in particular (as the data can be classified, subsetting and looked at in detail).

³ When dealing with government publishing documents to and interacting with the public online, I shall use the words "web" and "Internet" interchangeably. Although the web is only one of the services carried by the Internet, it is by far the dominant publishing platform. The only other service with comparable reach is email; in practice, even email interactions with government are now generally triggered by use of web forms or looking up institutional email addresses using a web directory.

Similar considerations arise with the structural questions about government and the part which local government plays in it. In this case a large and comprehensive crawl of the .gov.uk web, highlighting the view of government that it offers to its customers and citizens, is of great interest and provides a powerful new angle on the question.

If a web dataset is to be created, sampling approaches are one possibility. However, there is no suitable sampling frame for undertaking a random sample at page level, as there is no published list of pages in the government web while there is a published list of domains (Cabinet Office, 2013). Any sampling would therefore have to be at domain or agency level. This introduces a sizeable issue of granularity into the collection process, but one which is not insurmountable.

More problematic is the consequent difficulty in running analyses of the link network that would arise. Any gaps in connectivity, whether caused by accidental under-collection or by purposive sampling strategies, make analysis tools which rely on measuring the connectivity between parts of the web unreliable: they are sensitive to the gaps that would be created. Apart from the high volume of data that needs to be processed for a full-population crawl, therefore, there are few reasons to want to undertake a sampled approach.

As discussed in Section 2.4.5, each method used collects a different part of the picture and direct observation of the results of government action online is currently underused. There is, however, precedent for the use of webometric approaches in public administration. Taking this work forward though the construction of a full-population dataset would be a significant step forwards for our understanding of local e-government.

The study published by the National Audit Office (2007), for example, includes an analysis of the scope and connectedness of government using web crawl data from 26 major central government departments and agencies, but not including local authorities (or, indeed, the rest of the central government web presence). The aim of this exercise was:

“to identify various characteristics of government websites: their size and nature of content; the ease with which information can be found within sites; their visibility to search engines via patterns of hyperlinks; their deep linking provision and the extent to which they direct users outwards to other information sources.” (National Audit Office, 2007, p.40)

The authors were at pains to point out the wide range of metrics available and thus the sensitivity of their analysis to particular research choices made.

Much of this part of the report consists of descriptive statistics for the comparative size and scope of different central government web sites, but these covered only the handful selected and not the full population of interest. More analytical approaches included a shortest-path analysis for links between different sites and measurements of the size of the strongly-connected component (that is, the pages on a site from which it is possible to navigate via a series of hyperlinks both to and from the core of the site). The variation in performance on these metrics between sites was clearly shown, though there is a sense that the research was identifying variations in metrics rather than explaining underlying phenomena.

More interesting, perhaps, is the measurement of inlinks – to some extent a measure of both visibility and accessibility of information into the deeper parts of the government web. This analysis was facilitated by the use of Yahoo! Site Explorer, a tool which is sadly no longer extant. The outward-facing nature of government was explored with outlinks, collected as part of the web crawl. Measurements were made for links between government sites as well as to other domains outside the public sector. Again, there was huge variation here between outward-facing sites like those for Department for International Development (dfid) and the Department for Education and Skills (dfe), and those with much more limited outward-facing links such as the central Civil Service site.

3.2.3 *Operationalisation of key concepts*

It is necessary to operationalise the key concepts used in this research, particularly cost-effectiveness, performance, and customer satisfaction. Each of these choices has an impact on the structure of the analysis and of the findings that can be drawn.

Numerous approaches to operationalising cost effectiveness have been explored in public administration, as discussed in Section 2.3.3. By its very nature, cost-effectiveness measures require robust assessment of both cost (which is relatively easy to measure in our case as budgetary outturn data is available for all local authorities) and also of effectiveness. The most fulsome expression of this approach is perhaps Total Factor Productivity (see Dunleavy and Carrera, 2013, for example), comparing each individual service output against organisational costs, which for the reasons discussed in Section 2.3.3 is quite infeasible at the scale of a local authority.

A much easier, and surprisingly frequent, approach in the existing local e-government literature has been to use surveys of city managers to self-report administrative gains. Unfortunately, results from such research are fairly weak, and self-reporting may not be consistent. In Moon and Norris (2005), for example, the combined adjusted R^2 of their model is only 0.11, with the authors reporting that due to using existing survey data, “only a limited number of variables were used as surrogates for conceptual variables.” (Moon and Norris, 2005, p.57). More problematically, there are obvious incentives for city managers to report success in such research, and no obvious way to account for the restricted perspective of the managers as internal to the process being studied.

By contrast, government performance data from the Comprehensive Performance Assessment (CPA) program is used in this research to operationalise performance and the effectiveness side of cost-effectiveness. The choice of government performance data to conduct this analysis operationalises a par-

ticular view of what performance is, by assuming that the government metrics are a reasonable proxy for effectiveness. The nuances of the data, and the consequences of this on a fine level, are discussed in Section 6.1.1.

It is certainly possible to conceive of other approaches to performance which rely on non-government data, or on a comparative assessment of locally expressed goals against their successful realisation. Such approaches would tie the research less into a top-down central government-centric view of what is an effective authority, and focus more on local goals.

What is not possible with alternative measures, and why this data is so valuable despite its weaknesses, is to get a complete view of all authorities, across all significant services, across a period of time, across a broad and defensible measure. By enabling a panel modelling approach, this allows a much more sensitive and robust statistical analysis than is possible with any less complete data series, which in turn provides a probably unique opportunity to look at the association of performance and Internet use across a large group of real-world government bodies. Consequently, CPA data is considered here to be the best choice.

For customer satisfaction, the government's own customer satisfaction data is available, collected on a full-population panel basis for several years as part of the performance measurement process that provides the service quality data discussed above. This data, which is discussed in detail in Section 6.4.1 is uniform, collected as part of a good quality survey, and directly addresses the key underlying question ("how satisfied are you?"). For a study more focused on the specifics of citizen technology use, a more qualitative approach coupled to a broader operationalisation of satisfaction would be appropriate. For the kind of quantitative approach discussed above, though, the summary satisfaction data available here is ideal.

3.3 BIG DATA APPROACHES IN PUBLIC ADMINISTRATION

3.3.1 *Big data in political science*

The phrase “big data” is both highly contested and somewhat unhelpful. The trivial, but unhelpful definition is simply any huge dataset (“If you can store it on your laptop’s harddrive, it’s not Big Data”; aphorism cited by Monroe, 2013) to a radically different approach to social enquiry relying on atheoretical computational approaches and the attitude that domain-specific knowledge is actively harmful (“Your decades of specialist knowledge are not only useless, they’re actually unhelpful” Aldhous and Howard, 2012).

Within political science, the boundaries between political methodologists, data scientists, statisticians, and computer scientists are still being drawn, and there is no consensus about the best approaches to maximise validity. The standard “Three Vs of Big Data” have been extended to five in the context of political science by Monroe (2013): Volume (high), Variety (many different types and structures of data, not just survey tables), Velocity (huge quantities of data coming in, fast algorithms needed to handle it in sensible time), Vinculation (binding diverse datasets together), and Validity (just what do the data mean and how can they be reliably handled).

In this research, the key methodological focus is on the validity question, where we attempt to extend the use of large volumes of web crawl data to answer specific political science questions in a valid way. Interpretation of web data is particularly tricky and it is not yet clear the extent to which web graphs can be reliably applied to government.

“The evaluation of any measurement is generally based on its reliability (can it be repeated?) and validity (is it right?). Embedded within the complex notion of validity are interpretation (what does it mean?) and application (does it ‘work’?).” (Quinn et al., 2010, p.215)

3.3.2 *Web crawling in political science*

Research sub-question 1 ideally requires direct observation and measurement of the local government web; it therefore depends in part on web crawling and analysis of the resulting content and networks. This suggests a crawl of .gov.uk which is as complete and accurate as possible. This is possible within the state of the research art, given work done by social scientists and others on the construction of web-scale network datasets, but has not been undertaken before.

There is a continuum of scale in web crawling from a focused (and probably repeated) scraping of a small number of sites, with the crawl carefully constructed to extract particular text or data of interest from the target pages, to a huge-scale crawl which intends exhaustive coverage of a much wider scope. To the former belong work such as Hale, Margetts and Yasseri (2013) and Bright and Nicholls (2014); the latter are uncommon in political science, possibly because they require a significant crawling infrastructure and expertise in network analysis. Work exploring these larger-scale crawls is concentrated in web science and computer science (e.g. Lehmberg, Meusel and Bizer, 2014).

Some questions in political science and public administration, particularly broad-scale questions on size, structure, connectivity and rhetoric, are best served with the kind of broad-scale full-population data that can be fetched from a full web crawl, in addition to those questions which can be adequately answered with smaller-scale scraping. Previous political science forays into the broader domain have been limited by the restricted scope of the crawl data collected. For example, Escher et al. (2006) looks at a handful of individual government agencies in each of multiple countries, not the full scope of those governments' activities online, while National Audit Office (2007) covers a larger, but still incomplete, set of UK government agencies' websites.

This section discusses previous large-scale full-population work in political science, and offers a case for extending such methods to study of government web presences using the methods described below.

3.3.3 *Link analysis, nodality and organisation-level insights*

There is a long history of work in various academic fields exploring the use of hyperlink structures to create or discover knowledge (Park and Thelwall, 2003). By making various assumptions about the underlying meaning of links between web pages, an analysis of the link structure can be used to reveal relationships between the actors creating websites and/or used to intuit organisational structures and linkages.

In principle, “[l]inks are created for some reason to fulfil a certain task and that is why they may reflect some offline phenomena.” (Holmberg, 2009, p.3). Nevertheless, the meanings that may be embedded in a particular web link are fluid, variable and contestable. It is therefore particularly necessary in this field to tread cautiously in reading meaning into measurable phenomena, and to ensure claims made on the basis of webometric data are triangulated with other approaches.

One of the intellectual precursors of web link analysis is bibliometrics, a field aiming to extract meaning from citation networks within academic papers. Citations have obvious parallels to web links: they are directed links – one paper makes a unilateral link to another – and reflect some kind of relationship between the two objects. This insight has spurred the development of the related discipline of webometrics, trying to understand the nature of web links and extract meaning from them.

A different approach to the problem comes from the Social Network Analysis (SNA) researchers. This approach builds upon sociological understandings of social ties and attempts to understand the Internet as a social network, with web links reflecting relationships between those who placed them. Although

the underlying ideas are different, there is much overlap in practice between the kinds of approaches used by webometrics and SNA researchers.

Possible motivations for the creation of web links are discussed by Holmberg, who notes that at the time of writing “no reliable automated classification schemes are available” (Holmberg, 2009, p.98)

The key theoretical issue is what do links mean and what, therefore, are we measuring when we analyse them. To answer the research sub-questions above, what can they tell us about English local authorities? This is a question not previously answered by political scientists.

Although there is a thriving literature on the analysis of networks in sociology, physics and computer science, it is still a relatively uncommon research method in political science. One reason is that it is reasonably difficult to interpret what links in a network actually correspond to in a political context (Brügger, 2013). For example, a link from a citizen to a politician’s tweet may represent endorsement or mockery; without careful analysis of the context of the link as well as the link itself, it is difficult to map the link graph to the real-world phenomena of interest.

Within the public administration context, however, uncertainty of link purpose is less of an issue. Links within public services delivery sites can be reasonably interpreted to be functional – that is to be intended as a way of giving access to information or services to the browser. The minimum necessary interpretation of a link is simply that the linker is aware of the linkee, and judged that it would be useful (either to the creating organisation or to the user of the web page) to make a link to their content.

There are some precedents in the e-government literature for the analysis of web link networks to measure various aspects of government online. Of particular interest are those measurements identified by Petricek et al. (2006) as of relevance to political science understandings of e-government: navigability, measured by connectedness of components, node distances, path lengths, indegree, and outdegree; and nodality as measured by hubness and authority.

These metrics are primarily internal, which reflects the single agency per country comparative design of the Petricek et al. study. By contrast, the relative extent to which geographically- and structurally-related bodies have interlinked is of particular interest to a study of local public administration, as it speaks to the extent to which authorities are collaborating by directing web visitors to each others' services. In this case, an extension of the approach of Holmberg (2009) would be apposite: not only looking at the raw propensity for outlinking, which is measured by outdegree, but assessing *to which* other bodies connections are made and the extent to which this reflects our assumptions of geographic and functional connectivity.

One key concept which might be measurable using network analysis is that of nodality – the extent to which governments are at the centre of social and informational networks⁴ (Escher et al., 2006). This last observation is particularly important in the study of local government, whose role as the sole governmental organisation with local democratic legitimacy is held as central to its operations – which were increasingly through partnerships, influence and contract rather than by direct delivery (J. Stewart, 2003).

3.3.4 *Online traces of governments and their service users*

There are two separate sets of full-population data which are available on the web: the pages with their associated text plus their link graphs, which are traces left by government activity in the public sphere, and the traces of individuals accessing them and using the services which they contain.

The former are readily available to web crawlers, notwithstanding the practical difficulties inherent in doing so reliably and on a wide scale which

⁴ Nodality is theorised as one of the four “tools of government”, the means by which government can achieve its goals, in the so-called NATO model. The other tools are Authority, Treasure, and Organisation (Hood and Margetts, 2007). One open question is whether government can maintain its nodal position in a world where Google is used as a hub for the world's information. Without it, the position of government potentially becomes a lot weaker.

are outlined in section 3.4.2. The latter give evidence as to which parts of the government web are used, as opposed to merely being created by government and forgotten. Citizen usage data is harder to obtain, as the traces are not left in public, but rather in server log files which are accessible to the server owning organisations and not to researchers.

In principle government agencies could use Google Analytics or a similar service providing real-time analysis of user activities across all their websites to get a hugely powerful synoptic view of citizen interactions with government. This is not yet in place across government agencies, though it is being worked on in the United States (General Services Administration, 2014), and even if it were it is unlikely that access would be granted for research purposes.

Nevertheless, such information would be hugely helpful to scholars of public administration in particular. There are, after all, two views of government activity: that from the perspective of the government itself as seen through its officers, processes and management, and that of the public who are the users of the supplied services and for whom the government organisation (at least in theory) works. The methods of analysis outlined above necessarily give a very organisational-centric view of public administration, as the activities and views of the public are not considered directly. Although that gap can be filled in part by the use of survey data on citizen use of government technology (e.g. Dutton, Helsper and Gerber, 2009), a more complete and detailed view can be had from the online traces left by citizens directly.

The most obvious route to obtaining this data is to obtain it directly from the government organisations under study. Some trace data, on the other hand, can be collected by the researcher without needing the co-operation of the government being studied. One method for getting detailed information is to set up a controlled online experiment (Margetts and Stoker, 2010), with the experimental subjects using the web site under study and their responses, routes through the site, and timings providing detailed evidence of users' perspectives.

Other possibilities include obtaining traces indirectly. Rough comparative estimates of visitors to web sites could be obtained by getting access to the logs of a Domain Name System (DNS) server with a reasonably large user population. As computer systems wishing to access a website by name (e.g. “direct.gov.uk”) need to perform a name-to-number lookup using the DNS, rough volumes could be estimated with the co-operation of either an Internet Service Provider (ISP) or a university IT department⁵.

Some conventional longitudinal measures of web use are also potentially available, in the form of metadata attached to the release of Socitm’s Website Performance Survey (WPS). The WPS is a pop-up customer survey for local authority websites, familiar to many of us who have accessed online services at any point in the last few years. It is an optional service organised by Socitm and used by only a fraction of its membership – nevertheless, at the point of the 2013 data analysis, 100 English councils used the service and thereby provided comparative data (Socitm, 2013b).

The main focus of the WPS is on measures of customer satisfaction, but the framework also collects estimates of the number of visits, and visitors, to the websites in order to provide normalised measures. This data would provide an unusual opportunity to get “within the firewall” and access information which would normally require individualised negotiation with a large number of different holders⁶.

However obtained, visitor and visit numbers could be used to provide additional context as to the development of web sites by local authorities, crucially linking the customer’s perspective of demand with the supply-side perspective of the local authorities involved. For those authorities not covered by the data release, it would be possible to estimate visits and visitors by means of a re-

⁵ As computers subsequently cache the DNS response, this would be an estimate of the number of *users* visiting each site, rather than the volumes of pages fetched – though it may be possible to estimate the latter with the co-operation of an ISP which runs a web proxy.

⁶ A campaign of mass Freedom of Information Act requests would be a less-preferable method, running the risk of annoying stakeholders who may be of greater interest as willing participants in future research.

gression model. By assuming that those authorities which participate in the survey are representative of all authorities⁷ and analysing the outturn numbers alongside possible independent variables such as demographic factors, local penetration of Internet access, and use of the same to access government services (these latter are available in the Oxford Internet Survey (oxis) series: Dutton, Blank and Groselj, 2013; Dutton and Blank, 2011; Dutton, Helsper and Gerber, 2009; Dutton and Helsper, 2007; Dutton, Gennaro and Hargrave, 2005) then a model could be constructed to fill in the gaps.

3.3.5 *Scales of analysis – linking traces to macro-level phenomena*

A proposition which informs most of the rest of this thesis is that, in the local government context at least, inferences can be drawn between organisations' websites and offline organisation. As there has long been a central government requirement to make all local government services "e-enabled" (ODPM, 2003), the online presence should reflect at least the most significant local government services:

"[...] resulting in a complete network that is open to structural exploration. The task here, therefore, is not one of inference or testing for random effects but rather of interpretive description based on structural features. Such an exploratory approach is valued and extensively promoted in network studies, not only for learning about networks in general (De Nooy, Mrvar, and Batagelj 2011) but also when studying fields as diverse as international relations (Maoz 2011) and "dark" (or illicit) networks (Everton 2012)." (Georgiou, 2014, p.158)

There is evidence supporting local government being well-linked along service lines. As early as 1999, Socitm were saying that councils' sites "should be well connected to other sites, both in the local community and to appropriate

⁷ Possibly an optimistic assumption – one would expect more keen and competent web teams to be more interested in the kind of benchmarking that the wrs offers – but not an outrageous one.

regional and national sites, to take advantage of the Web's ability to make distributed information link together in one place" (Socitm, 1999, p.5).

It is furthermore postulated that the relative size and degree of development of different areas of local government websites reflect something important about the services involved, that links from local government websites to other bodies reflect something about a functional relationship between the link donor and donee, and that similarities and differences in the structure of local authority websites should reflect real-life differences in the organisations involved. These propositions are tested in Chapter 4.

This potentially allows web link and content analysis to inform an understanding of the homogeneity and heterogeneity questions which have long been an issue in local government public administration. As Whalen (2011, p.5) puts it, web links "may shed light on how seemingly unrelated organizations are indeed "close" in an organizational structural context".

If it is theoretically reasonable to project the government web on to the social and informational networks at the heart of government, then it is possible to measure nodality fairly directly with the crawl data set collected here. Social network analysts have developed the measure of "betweenness centrality", to measure the extent to which a particular actor is at the centre of a given network. It is calculated as a function of the connections in the network that need to flow through that person to link two other nodes. By analysing the data at an organisational level, it should be possible to measure betweenness centrality for at least publicly-facing functions of government and thus get a reasonable (if indirect) measure of expressed nodality.

3.4 WEB CRAWLING: METHODOLOGICAL ISSUES AND WAYS FORWARD

3.4.1 *Definition of terms*

The terms used in analyses of web crawls vary between webometricians who have inherited from information science, social network analysts who have inherited from sociology, and computer scientists who have inherited from computer networking. None of the sets of terms are particularly meaningful to political scientists, for whom this whole approach is fairly new.

This study uses the terms selected by Petricek et al., who did some of the early work in political web crawl analysis and provide a helpful and admirably clear definition in their paper:

“Networks have been studied in a variety of different fields, including computer science and social studies. The diversity of these disciplines has led to a diversity of vocabulary so a brief definition of terms seems useful.

A network consists of a set of *nodes* and a set of *directed links* that connect pairs of nodes. In our case, the nodes are documents retrieved from the Internet and the links are hyperlinks that can be used to navigate between these documents. Every link has a *link source* that is the node from which it originates, and a *link target*, which is the node to which it points. A node can be described in terms of its links: every node has an *indegree* which represents the number of links for which the node is a link target, and an *outdegree* which represents the number of links for which the node is the link source. The sum of the *indegree* and *outdegree* is the *degree* of a node. A node with non-zero indegree is receiving links or has *inlinks*, while a node with non-zero outdegree has *outlinks* or is pointing to another node.

We distinguish different entities on the Web that are ranked in a hierarchy. The smallest entity is a document on the Web, identified by a *Uniform Resource Locator (URL)*. A set of documents constitutes a *website*. A website is primarily conceptual and depends very much on the perception of the user or the author of the site.” (Petricek et al., 2006, p.671, highlighting in original)

Like Petricek et al., this study automatically identifies websites by extracting the hostname from the fetched crawl's Uniform Resource Locators (URLs)⁸. Because of the larger size of the present crawl and the hugely variable use of multiple hosts for different services between local authorities, for most purposes the unit of analysis is the *domain*, taken here to be the third-level domain from the DNS root: *darlington.gov.uk* or *judiciary.gov.uk*. These reflect the basic level of registration in *.gov.uk*. For some purposes, the domain *owner* is more important, which is the real-world government entity that controls a particular set of domains.

A process for matching domains to owners has therefore been developed, which is described in Section 4.2.3. This allows either the domain, or the real-world entity responsible for the domain, to be used as the unit of analysis as is appropriate.

3.4.2 *Web crawling concepts*

Web crawling is the process of using a computer program to follow links around the web, collecting and storing the content as it goes. The aim is to follow each of the links within each page fetched and thereby download a complete copy of the portion of the web which is of interest. Such crawls are carried out routinely by search engines to build indices of the web, by web archives to collect snapshots of the web at different points in time, and (less commonly) by researchers collecting webometric data.

The data collected are split. The page content is used by search engine algorithms to know what the page is about⁹, but is relatively rarely used by academics. A link graph can be constructed treating each page as a node and each hyperlink as an edge in a logical network; this creates a relational dataset about the links between different items in the network and can be

⁸ e.g. www.darlington.gov.uk from <http://www.darlington.gov.uk/your-council/public-alerts/>

⁹ Using a variety of tools from the information retrieval field

analysed using social network analysis and webometrics tools to look at the individual connection patterns of particular nodes or the structural properties of the network as a whole.

Web crawling is not a neutral process of measuring a concrete object; web crawl data is the result of a complex series of researcher judgements, each one of which affects the results. The decision by Escher et al. (2006) to collect a restricted scope of data on the governmental subject of their study allowed the answering of certain types of questions and precluded the answering of others. Even the very process of controlling the order in which a crawl operates will change the resulting dataset, as the web changes while the crawl is ongoing. Decisions about what data is valuable and not valuable, taken at the beginning of the research process, have possibly long-term consequences for the utility of the dataset¹⁰.

“A fundamental point that needs to be understood is that the result of crawling a web site, even though it is by a computer program, is dependant on many human judgements.” (Thelwall, 2003, p.1)

A necessary precursor to studying the web using its primary source material is to index, via crawl, just what is present, where, and what links there are between different objects of interest. Although commercial search engines have already undertaken much of this indexing work, and have vastly larger databases and time-series than a researcher can hope for, their data is not generally available in a form which is useful for academic work. Searches can be made using the web interface, but these are designed to prioritise efficiency rather than accuracy, do not allow a full set of results to be fetched for any given query, and do not give effective access to the link graph they collect

10 There are more practical examples too: are images, Cascading Style Sheet (css) files, and javascript to be crawled? If so, the cost, scope and time of the crawl is extended. If not, it is very difficult to use the resulting archives in the Wayback Machine (discussed in Section 3.6.2), thereby precluding some of these new methods of analysis.

(Holmberg, 2009). As such, it is generally necessary for researchers to collect their own research data.

Once collected, a web crawl is an unusual kind of research dataset. With the amount of dynamic content on the web, the particular collection made for the research is likely to be unique – one of many possible sets that could have been served to clients at the time of the crawl. It is thus a particular representation of the web at the time of collection, not the web itself (Brügger, 2013). Completeness is a particular issue:

“On one hand the archive is incomplete – something is missing – but on the other it may also be too complete – there are simply too many versions of ‘the same’, some of which may be duplicates.”
Brügger (2013, p.314)

This raises a number of issues for researchers conducting their own crawls (how to maximise the once-only coverage of the parts of the web of interest), and a different set for those using crawl data gathered by others (how to handle the inevitable duplication and gaps)¹¹. Both issues are relevant in this research.

For the former, a broad crawl can be undertaken. Various parameters can be set during web crawls and a decision can be made to collect widely to preserve the opportunity to subset and filter at a later stage¹².

Handling duplication is in many ways a harder problem. One approach is to systematically identify duplicative crawling behaviour and websites set up with infinitely-looping links, and prevent such links being followed or retained. Other strategies are possible, perhaps limiting the crawl by elapsed time, number of links to follow from the original “seed” site, or an automated

11 Common Crawl <http://commoncrawl.org>, for example, have a large-scale crawl dataset freely available for research purposes. As this was not collected specifically for any given research, it is unlikely to be able to answer all possible research questions. Its coverage of .gov.uk is partial, for example, giving a different set of constraints than those faced by this research project fetching data for itself. The lack of clarity about exactly what is contained in the Common Crawl database is a major weakness for academic re-use.

12 Robust and well-developed crawlers, like the Internet Archive’s Heritrix (Internet Archive, 2013) can not only follow explicit links given in Hypertext Markup Language (HTML) anchor tags, but can also infer and extract links from javascript navigation elements, Hypertext Transfer Protocol (HTTP) referral links and page headers.

duplication detection system. The need for this kind of work should, in principle, be essentially eliminated by publishers' provision of properly configured robots.txt files (robotstxt.org, 2007), but this is a fantasy on the real web, which is full of misconfigured sites.

Likewise, some under-collection is made unavoidable by *following* the robots.txt directives, simply because a few sites exclude all crawlers from the whole site and others deliberately restrict content we would choose to collect. As the robots.txt rules are advisory, and not technically enforced¹³, this kind of constraint is a matter of ethics, not of technology and is discussed further in Section 3.4.4.

3.4.3 *Strengths and weaknesses for statistical analysis*

The great merit of a 'full-population' web crawl which is as complete as possible is that the coverage is broad and the data are finely subsettable. Even if it is not possible to guarantee coverage of every page, for the reasons identified by Thelwall (2003) above, the dataset will have much more even coverage than one created by attempting some kind of sampling technique. As the data are wide in coverage, it is straightforward to extract any category of interest without the prospect of needing to re-conduct a crawl to collect top-up samples (which would, in any case, introduce an unhelpful extra time dimension into the data).

From a pure point of view, however, even the most meticulously-collected web crawl data are in a very real sense not full-population as coverage can never be complete. Statistically, this leaves the researcher caught slightly between two uncomfortable stools. On the one hand, the main thrust of conventional social science survey statistics, which are dedicated to quantifying the extent of sampling error and its consequences for the testing the research hypotheses,

¹³ A real world analogy would be a series of "no entry" signs placed on particular doors.

simply do not apply: there is no sample involved, and the measurements taken on the resulting dataset do not suffer from the kind of systematic (yet statistically tractable) sampling error that occurs when randomly selecting from a larger population. On the other hand, the data are clearly not error-free in the wider sense: the effects of robots.txt-driven non-collection, for example, looks very similar to the analogous problem of (unquantifiable) non-response bias in survey research.

In a strange sense, therefore, the web crawl dataset is almost akin to the kind of data collected using non-random sampling methods, in how it needs to be dealt with; conventional statistical measures of sampling error do not apply, yet there will be subtle biases introduced by the very nature of the human judgements used in its collection. So as researchers, we need to eschew ‘statistical significance’ as an analysis framework and p-values as a metric.

The valuable measurements to be made are, like the underlying reality in social science, to do with effect size: are the patterns and regularities found within the data consistent with sizeable effects, or trivial ones? As we cannot control the circumstances of data collection sufficiently to be sure that small effects are not artefactual, we must to some extent adopt a self-denying ordinance for ignoring what could be chance findings or the effects of collection choices in the data.

Of course, none of these problems are exclusive to the web crawl. In fact, they occur for all non-perfect data collection methods. In this case, the non-existence of (substantial, and easily statistically tractable) sampling error simply highlights the sources of error that remain in any careful quantitative analysis.

3.4.4 *Web crawling ethics*

The ability of web crawlers to fetch huge amounts of data from the public web raises ethical issues. In particular, researchers must consider questions of resource use, privacy, and copyright.

A web crawler running on a fast computer is quite capable of overwhelming a slow server with enough requests for pages that its ability to serve its intended audience is compromised. Various strategies are possible to prevent this from happening, either from the researcher's end (normally variations on an enforced delay between page fetches) or from the site owners (a Crawl-delay parameter set by the owner, an email to the crawl operator, or a block of the crawler's Internet Protocol (IP) address). It has become common practice to put contact information for the crawl operator in the headers of each page request to make it straightforward for site owners to get in touch in the event of unexpected problems (Internet Archive, 2011)¹⁴.

Privacy is also an ethical issue faced by research web crawl operators. All the pages fetched are by definition published on the public web and therefore accessible to anyone with a web browser. Nevertheless, a good deal of personal information is contained within these pages – the names, work telephone numbers and email addresses of council employees are a common example in the present research. It is possible to insist that the lack of express informed consent for the use of this public data in research should preclude these approaches. Even if we reject this extreme claim, as most in the field do (Thelwall and Stuart, 2006), assembling transient web pages into a permanent research dataset for reuse does raise some questions about subjects' rights to correct inaccuracies or ask for sensitive personal data to be removed.

14 That was done for the crawl described in Section 4.2.2 and enabled a handful of contacts from site operators raising questions about crawl scope. Interestingly, none of the emails were in any way angry: two of them were explicit that they thought the research was a great idea and they'd like to help. The major issue raised was a side-effect of the web crawler Heritrix's handling of badly-coded websites using javascript as their principal means of making links. Part of its way of obtaining completeness in a crawl is in some circumstances to essentially guess the URLs of pages that might be relevant by assembling various bits of page content into a URL string (Internet Archive, 2014). Unsurprisingly, it sometimes assembles nonsense, leading to HTTP 404 (not found) error codes. Even more regrettably, for some misconfigured web servers, 200 (success) codes are returned, encouraging further guesses and a lot of chaff in the returned crawl. This feature of Heritrix (privileging completeness over accuracy) is somewhat problematic in social science research and perhaps should be disabled in future work. In one case, the crawler so consistently generated erroneous URLs that it set off a council's firewall intrusion alarms, which believed that a buffer overflow attack was in progress. This was easy to halt, once the crawl operator was warned.

A third issue is copyright. Although there is probably an implied licence in most jurisdictions to make copies of the content for the purposes of reading it in a browser, it's less clear that this implied licence extends to saving offline copies, much less systematically re-assembling a whole site's worth of web content for publication as a research dataset¹⁵. Many national libraries and archives, hoping to archive the nation's activities on the web, have felt the need to push for specific exceptions to be carved out of national copyright laws (e.g. *The Legal Deposit Libraries (Non-Print Works) Regulations 2013*) to explicitly legitimise these activities. Search engines and caches, on the other hand, tend to work on the principle of storing the content and removing later on receipt of a specific demand from a copyright holder. That such sites have not been shut down, provides some evidence that this is a legal approach, as well as a substantively reasonable one (Thelwall and Stuart, 2006).

Many of these issues can be handled in conjunction with the operators of the target web servers, using the Robots Exclusion Standard (robotstxt.org, 2007). This specifies a standard file called robots.txt which allows site operators to specify whether web crawlers have permission to crawl at all, if so whether there are any areas from which they should be excluded. Unofficial extensions to the robots.txt standard also allow site owners to specify a maximum crawl rate for crawl operators. Unfortunately, this is not sufficient to resolve the ethical issues entirely; not all site owners specify a robots.txt file, and some of those which exist are non-compliant with the standard, which makes them difficult for crawling software to use automatically¹⁶. An issue for the researcher is that some sites prohibit web crawlers entirely. To obey this instruction inevitably results in an incomplete dataset; to ignore it clearly raises ethical

15 In particular, the European Union's database right restricts unauthorised extraction and utilisation of a substantial part of even unoriginal databases (Lipton, 2003).

16 A particular issue encountered during this research has been the number of robots.txt files encountered which feature blank lines. Technically anything after the blank line should be ignored, but the intent behind many of these files is clearly that the latter parts should be followed.

questions. For the present research, a “polite” crawl has been conducted, with a non-aggressive default crawling speed and all valid robots.txt rules respected.

This is in keeping with the ethical standards of webometrics. It extends previous work in political science by focusing particularly on the opportunities to use these techniques to analyse full-population web crawl data, and develops new tools and approaches to facilitate this.

These ethical questions will be discussed further, to review the choices taken and give recommendations for future research, in Section 7.5.3.

3.5 TEXTUAL METHODS IN POLITICAL SCIENCE

This section outlines a new approach to analysing web content, using machine learning tools, and argues that it offers many advantages for political scientists.

The web is a large and complex place; the types of content contained within it range from raw data intended for machine consumption (such as that published as part of open data work by governments) to sophisticated games and web applications which use the browser as a platform to dynamically deliver content. To use the web as a source of research data it is necessary to carry out two related tasks: firstly, to *filter* the content to focus only on that which is germane to the research questions being asked; secondly, to *classify* content in order to carry out analysis.

Filtering can in large measure be done at the collection stage by careful selection of collection criteria (perhaps by including certain websites of interest and excluding those out of scope of the study). Nevertheless, this is not always straightforward: in the present study, for example, the service delivery uses of the Internet by local authorities are of key concern, with democratic uses less so. Both are, however, published on the same websites. While it is possible to differentiate some manually by URL, this varies by site, and does not scale.

When analysing large datasets, it is often necessary to undertake some form of classification: in scope / out of scope, positive / negative in sentiment,

belonging to group A/belonging to group B. The traditional approach in political science is to hire graduate students to hand-classify items (e.g. National Audit Office, 1999, 2002, 2007). If there are too many for this, only a sample is classified.

An alternative approach, based upon human / computer hybrid supervised learning methods, is gaining increased prominence in political science, being increasingly used as a substitute for, or supplement to, hand-coding techniques (Grimmer and B. M. Stewart, 2013). When effectively implemented, the costs of coding can be reduced and the reliability increased (Quinn et al., 2010, p.210). This is despite the statistical classification error that arises as a result of the statistical models used being unable to identify categories with complete accuracy. In particular, the use of Support Vector Machine (svm) classifiers to improve efficiency in classification offers opportunities for improving the effectiveness of different kinds of text-based research (D’Orazio et al., 2014).

Related unsupervised statistical topic modelling techniques have also been developed for automatic clustering of content based upon the extraction of statistical similarities rather than any pre-conceived coding scheme. These approaches have been used for extracting groups of topics from congressional press releases, for example (Grimmer, 2011). The methodological difficulties are transferred here from matching the accuracy of hand-coding to the less-structured problem of demonstrating that the clusters extracted by the analysis are substantively meaningful in some sense – itself a theoretically challenging and time-consuming process. Nevertheless, such approaches are clearly useful for exploratory research and have even been used for cross-national and cross-language comparative research in political science, using machine translation to produce comparable corpora (Lucas et al., 2015)¹⁷.

17 The comparability of texts produced using machine learning is an extremely tricky problem, which Lucas et al. treat quite lightly, though they do develop an analysis method which incorporates a “language translated from” parameter to deal with some of the consequences of translation variability.

Some of the uses to which machine learning tools have been put include automatically spatially coding political text, using a word frequency analysis (Laver, Benoit and Garry, 2003), and applying classifications to interstate disputes, using an svm (D’Orazio et al., 2014).

This section outlines how an application of this approach, based upon machine learning tools, can be applied to web crawl data to provide meaningful attribute data to allow richer and deeper analysis of government web provision.

3.5.1 *Automated classifiers*

The use of automated classifiers built on statistical models is well established within corpus linguistics and computer science. Common uses are for categorising email as spam or not spam, assessing the positive or negative sentiment of arbitrary texts, or for deciding the topic of a news article from fixed categories such as “sports”, “technology” and “politics” (Bird, Klein and Loper, 2009, Ch.6, §1). In principle, this allows classification of all pages in a web crawl on any scale with a modest number of different categories, enabling large-scale analysis of government web content. In contrast to small-scale sampling-based approaches, this provides a substantial volume of data as each unit is analysed.

Various forms of classifier are available, including fully automated approaches. The most promising from a political science point of view are the supervised classifiers, which take as input a series of manually categorised items, and a set of document features to analyse¹⁸. A statistical model can then be built with the document category as a dependent variable, and any arbitrary document can then be classified using the model. Common approaches are those based on so-called Naïve Bayes models, Maximum Entropy classifiers, and various categories of svm. It is common in the machine learning and information retrieval literatures to select between them on pragmatic grounds

¹⁸ A typical example is to identify a set of important words in the dataset and then use the presence or otherwise of each word in each document as a feature.

(Manning, Raghavan and Schütze, 2008); their success rates depend on the data being classified and as it is normal to fully cross-validate the final model on virgin data, this causes no empirical problems with false reporting of error rates.

Classifiers have significant advantages as social scientific tools for handling big datasets, as well as disadvantages. By classifying the whole population in place of a small sample, a classifier-based approach eliminates sampling error, at the cost of introducing classification error. Classification error rates (as well as the detailed precision¹⁹ and recall²⁰ rates for particular categories) can be accurately estimated by validating against a hand-coded validation sample which has not been used for training.

There is no consensus about how effective a classifier needs to be before its results are treated as reasonable for further analysis. Computer scientists' concerns are often to produce the most reliable classifier possible for a given standard dataset used in model development, but given the inherently fallible nature of the process accuracy will never be 100%.

This problem is not unique to training computers to classify content; the variation in human coders classifying the same material using the same coding frame has been a feature of hand-coded work for many years.

In political science, in fact, a popular approach to the problem of inter-coder reliability has been to ignore it, even in fairly prestigious projects. Laver, Benoit and Garry (2003, p.329) report that "We have seen no other published tests of intercoder reliability in relation to political science content analysis, but we know informally from our own experiences with this that it is a major unspoken problem." Benoit and Laver defend the use of automated textual analysis by noting that the "gold standard" being compared against is itself a

19 Precision is a measure of how well the classifier has selected only true members of each category, rather than false positives.

20 Recall is a measure of how well the classifier has selected all the true members of each category, rather than false negatives.

contestable classification made by a single researcher (Benoit and Laver, 2007, p.131).

Without the rigorously hand-coded standard datasets used by computer scientists to verify their new models, there is no way of establishing the ground truth of a particular automated classification. Nevertheless, the classification model is, in principle, simply another coder. It should be possible to conduct inter-coder reliability trials against human coders using conventional metrics such as Krippendorff's alpha (Krippendorff, 2004); this is at least as much assurance as is obtained in traditional computer science or political science classification processes.

In general, a supervised classification project can be approached in the following way:

1. Hand-classify a sample of items;
2. Run an inter-coder reliability test using a small sub-set of items multiply hand-classified;
3. Separate the classified items into training / testing and validation sub-sets;
4. Set up the classifier, selecting feature sets and optimising learning algorithms;
5. Test the performance of the classifier by automatically classifying the (hand-classified) test set;
6. Repeat steps 4 and 5 to maximise the performance of the classifier on the test data;
7. Validate the performance of the classifier by testing the finished classifier on the reserved validation sub-set; then
8. Classify the main dataset²¹.

21 It is possible to obtain confidence estimates for the classification for each page, as well as a best-guess classification. In this way, doubtful decisions can be followed-up by hand, or discounted if desired.

3.5.2 *Classifying web content*

One attractive quality of automated content classifiers are that they can be applied across essentially any hand-codable dimension: they can be applied to any problem with identifiable features that a human is capable of distinguishing. In public administration research, one key application would be to add attribute data to the individual items in big datasets. In the present research, for example, estimates of website size in a crawled dataset are complicated by the inclusion of various generated databases in the crawl archives.

One possible application for this approach, therefore, would be the identification of web pages which are entirely constructed from large-scale databases, rather than hand-written content. The scale of online databases such as land registry records, plus the ability with most database systems to present the data in a large number of duplicative ways (sorting by column, for example, or taking arbitrary subsets by tag) can give a wholly misleading view of the “true” size of the government web – with ephemeral database pages vastly outweighing the smaller amount of content manually added. For an example, see Table 2, which shows a selection of URLs for accessing a single database-stored document, with various subsets of its metadata and altered into various different formats. Some databases found during this research offered hundreds of views of each document, rather than only a few. A similar issue can arise with filtering systems auto-generating URLs for combinations of tags²².

There is, of course, a continuum from purely hand-coded HTML to pages which are a thin veneer over a pre-existing large-scale database. The choice of what to include or exclude would depend on the research question presented. For this study, for example, exhaustive coverage of online library catalogues is

22 <http://www.avonandsomerset-pcc.gov.uk/Your-PCC/PCC-Blog/Tags/dunktanks/motorcycles/mopeds/alcoholawarenessweek/businesscrimeforum/>, for example, is just one of several thousand URLs generated with various combinations of tags on this particular Police and Crime Commissioner’s blog.

URLs
http://wisenet.wiltshire.gov.uk/documents/dsweb/ServicesLib/Document-11086/History
http://wisenet.wiltshire.gov.uk/documents/dsweb/ServicesLib/Document-11086/Object.keywords
http://wisenet.wiltshire.gov.uk/documents/dsweb/ServicesLib/Document-11086/Object.title,Object.summary,Object.description,Object.keywords
http://wisenet.wiltshire.gov.uk/documents/dsweb/ServicesLib/Document-11086/Object.summary,Object.description,Object.keywords,content
http://wisenet.wiltshire.gov.uk/documents/dsweb/ServicesLib/Document-11086/Permissions
http://wisenet.wiltshire.gov.uk/documents/dsweb/ServicesLib/Document-11086/View
http://wisenet.wiltshire.gov.uk/documents/dsweb/ServicesLib/Document-11086/application/msword,application/vnd.ms-powerpoint,application/vnd.ms-excel
http://wisenet.wiltshire.gov.uk/documents/dsweb/ServicesLib/Document-11086/image/gif,image/jpeg,image/png,image/tiff

Table 2: Multiple auto-generated representations of a single document, from a live web crawl

not important, and these would be excluded²³, whereas the elected member contact and biographical information contained within democratic services modules on council websites is probably of interest. Training a classifier to detect pages of no interest would allow decisions on scope for a size analysis to be made independently of the original decision around the proper scope of the crawl.

There are, nevertheless, some challenges in the automated classification approach. The most problematic is where the variation in words between pages in different categories is not high, which makes classification more difficult. The error rate of completed classifiers will vary depending on the nature of the source material, as the algorithms are better at differentiating categories with widely different source material than those with very similar ones. The amount of information available in each individual text is also a feature; classifying Hansard debates with thousands of words of textual variation per item is much more reliable than trying to do the same with a dataset of 140 character tweets.

When classifying local government content, the task is made harder by the difficulty for even trained specialists of separating out a continuum of services:

“It is not straightforward to classify services for a number of reasons:

- i Almost all services can be organised in several different ways across local authorities sharing similar characteristics.
- ii Different terms can be used to describe essentially the same service (e.g. trading standards or consumer protection).
- iii In two-tier parts of England the same service (e.g. planning) actually covers different activities.
- iv Some services regularly straddle departmental boundaries (e.g. tourism and leisure).
- v Some items of information might be classified in more than one service category (e.g. information about transport might be part of transportation, economic development or leisure and tourism).

23 In principle, the site owner should probably have excluded these using robots.txt, in any case.

vi What constitutes a major service in one authority may be less important in another (e.g. economic development).”
(Socitm, 1999)

For some problems, an unsupervised clustering approach using topic modelling may be suitable. This uses no researcher input beyond the selection of features, and extracts “natural” clusters from the data, which then need to be interpreted. If not successful, other approaches can then be taken such as hand-classification of a sample.

This approach is, however, computationally expensive. Although the complexity of classifying each page is not high²⁴, the millions of items in a typical web dataset is sufficient to make the aggregate time of classification comparatively high, which limits the total number of classification dimensions possible within a reasonable time period. Nevertheless, the generality of the classification process does offers opportunities to add several imputed categorical variables to datasets of this size, if necessary, at modest cost in both researcher and computer time, and with validated levels of accuracy.

The final advantage of adding a population-level automated classification is that it allows the researcher to look at any scale within the dataset without change in the (constant) classification error rate. The alternative approach of hand-coding a sample of items can provide acceptable summary statistics for the overall dataset, but when looking at slices of data or sub-samples of the original, the small number of classified items in the sub-sets widens the error bars dramatically.

A sizeable hand-classified sample used to train a supervised classifier nevertheless offers two major advantages. Firstly, the accuracy of the final model generally increases logarithmically with the number of training items. Secondly, it allows two approaches to be taken for analysis. Where quantitative data are needed which are available in reasonable quantity in the random hand-coded training sample (perhaps the number of transactional and informational pages)

24 Of the order of centiseconds per page

then a traditional statistical approach using the sample is possible. Where data is sparse, perhaps looking at a small service area across government bodies, then the classifier allows reasonable estimates to be made despite maybe only a few matching pages being in the original sample. Overall, the classifier allows analysis of the complete set of textual information arising from the web crawl, rather than only that of a small sample, allowing finer-scaled comparison of different areas.

If the performance of a classifier is too poor while trying to select a small sub-population, it would be possible to use a classifier as the first-round of a hybrid automated-manual process to select a number of pages for further qualitative exploration. By selecting classifier parameters biased towards recall, rather than precision, the results will contain many of the high-category pages (and also many which are not). A human coder can then go through the subset and select the pages which truly belong in the small category.

3.6 NEW METHODS DEVELOPED

To deal with some of these issues in the context of a government web crawl this research has developed some new approaches. Partly these are conceptual (in applying diverse tools from webometrics and machine learning to the study of public administration, as seen in Chapters 4 and 5). In addition, though, software tools have been developed and released as part of this research project which advance the state of the political science art and will be of value to future projects using the web as a source of political data. This section describes these tools: Warctika and the handclassifier framework using OpenWayback.

3.6.1 *The Warctika framework*

To properly answer the research sub-questions posed, two sets of data are needed from the web crawl: links between pages to be compiled into a di-

rected link graph, and the textual content of pages in order to run automated classifiers. As outlined in Section 3.4.2, the web crawler Heritrix (Internet Archive, 2013) is a good choice for a large crawl. Extraction of links is then straightforward, as Heritrix saves its data in the standard Web Archive (WARC) format (ISO, 2008), and provides tools for automatically extracting links to a metadata-containing Web Archive Transformation (WAT) file from each WARC. Extracting text from the HTML and plain text formats collected is also easy, as HTML parsing libraries are readily available.

There is, however, a large part of the data collected which is not in easily parseable formats. In particular, Portable Document Format (PDF) files, Microsoft Word, Excel and Powerpoint documents, Rich Text Format (RTF) files, and files in the various OpenDocument formats are not so easily accessible. This is particularly problematic, as such documents represent a huge proportion by size of the local government web. This raises the additional problem that storage budgets for holding crawls (already large because of the full-population nature of the material collected) are being further increased to hold information which is not accessible or useful to researchers.

To deal with these problems, a new automated framework has been developed for this research using Apache Tika (Apache Software Foundation, 2013) to convert WARC records in proprietary non-text formats to plain text equivalents before final storage. Tika is a general-purpose text extraction library, which accepts as input a large number of different proprietary and other structured formats and outputs structured or unstructured plain text.

The new framework, Warctika, reads in a collection of WARC files and applies a configurable set of regular expression rules to decide which content to send through Tika. For this project, the HTTP Content-Types associated with file documents formats were selected. As the Internet is full of servers which send erroneous Content-Type headers, the selection criteria are fairly complex and were based upon a pilot study of the data returned by servers in

the real world²⁵. Once the text-based output from Tika has been generated, Warctika writes it into the WARC file and optionally deletes the original record.

On the dataset collected for Chapter 4, this yielded storage budget savings of approximately 60%, as well as providing the text in a more convenient format for future research. In some cases this results in little text being recorded (perhaps because the PDF in question consisted mainly of drawings instead of text) or having content or formatting errors as a result of the unsupervised conversion process. This is, nevertheless, preferable to ignoring the contents of all PDF files.

On the live web, problematic PDF and Word files tend to be clustered, because an author or set of tools which produces at least one impossible-to-parse file is more likely to produce another. In principle, this leaves non-random gaps in the text coverage, which is awkward for subsequent analysis. In practice, such bias would exist in a worse form if as much text as possible was not extracted as government bodies also put data in closed word processing documents instead of HTML at different rates. In this case, much more data would be effectively missing for larger chunks of the analysis.

Warctika development has highlighted some archive-related issues which should be noted. The principal issue is that the WARC format is principally used by archives for permanent storage of crawls, not the intermediate processing of archive data. As a result, the toolset for handling processing of content is somewhat limited. There is a defined format for “conversion” records within the WARC standards document (ISO, 2008) but it is presently unimplemented in many tools, as it was designed for the future up-conversion of archive records for long-term accessibility, not for transformations shortly after collection²⁶.

25 The detailed set of filtering expressions used are published with the full source code (Nicholls, 2014).

26 This gap has also recently been identified by the Common Crawl Foundation, whose WARC Encapsulated Text (WET) format is very similar to the approach adopted here, and is currently being standardised as part of a revision of the WARC ISO standard.

Nevertheless, the Warctika tools have been carefully coded to correctly write standards-compliant conversion records.

Source code for Warctika, including the precise regular expressions used to identify target files from their HTTP Content-Type headers, is available (Nicholls, 2014).

3.6.2 *Wayback Machine as classification platform*

When dealing with research data, it is often useful to be able to view the data in a fairly raw form. This has benefits both for the researcher's understanding of the data and its context, for the opportunity to undertake qualitative and semi-qualitative detailed work on small parts of the dataset, and for the opportunity to hand-code samples of content for subsequent supervised automated content classification (see Section 3.5).

This kind of access is difficult with web crawl datasets. The data are saved either in WARC files intended for archival storage or as a collection of free-standing HTML files. The nature of the modern web is that proper viewing of many websites relies on access to server-side resources which are not available in offline viewing, or in the careful processing of fetched web pages to reconstruct in an archival situation a simulation of the online environment in which their embedded javascript code was intended to run. This research therefore developed the use of software tools to allow more effective and straightforward researcher access to the crawl data in a view as close as possible to the original form in which it would have been seen on the web.

One advantage of using Heritrix as the data collection platform is that the WARC files generated with the crawled pages are suitable for feeding into the Wayback Machine, the Internet Archive's platform for interactive browsing through archived websites, including automatic handling of images and links. It allows users to browse the archived web largely as if it were live, at least

within the limits of the sites crawled²⁷. The latest version, OpenWayback (International Internet Preservation Consortium, 2014), is now an independent open-source project which is suitable for local installation.

OpenWayback has been used as the base for a newly-developed streamlined manual content classifier for this project, an extension of the handclassifier system developed for a related research project. This streamlined hand classifier (Nicholls and Bright, 2016) speeds up the process of hand-classifying items as input and validation items for the more automated supervised classifiers by presenting archived items in-browser, along with a convenient button interface for recording classifications. By reducing the amount of manual processing required for the construction of training and validation sets, this not only increases the number of items which can be comfortably classified, but also reduces the likelihood of transcription errors in the resulting datasets.

The library can be usefully transferred to other political science projects using large-scale web data. Furthermore, the use of an OpenWayback installation allows difficult data cleaning decisions (‘do these pages contain real data or are they crawl errors?’, ‘is this part of the site a duplicate of this other part?’, or ‘is this page in the scope of the study or not?’) to be easily made by looking at a copy of the data actually under study, rather than looking at the live web (which may have changed since the data were collected). At its simplest, this simply involves accessing a given page by the URL `http://<OpenWayback server>/<URL to view>`, though flexible searching and timestamping options are also available.

The heart of the system is to set up a small window to one side of a web browser, forced to the top of the desktop window stack and with buttons relating to the categories of interest. As a category for each page is selected, the results are recorded in a Comma Separated Values (csv) file and a call is made to the local OpenWayback server to produce the next page from the

²⁷ The main site is at <http://web.archive.org>, and allows users to surf through the whole Internet Archive collection, at any point in time.

sample for classification. This significantly reduces the period of time required to generate the manual classifications sets which are necessary for training supervised machine learning algorithms, but could also be used for speeding up manual classification if that was considered desirable for a particular project. The use of the OpenWayback index makes it trivial to scale this classification up to web scale²⁸.

As a bonus, the crawl data can be browsed in a natural way. Rather than relying simply on descriptive webometric statistics for exploration of the data, a much richer picture is available for basic understanding and deeper qualitative work. In principle, access to the local copy of OpenWayback can be shared over the Internet, allowing other researchers access to the dataset without difficulty – making data reuse more straightforward.

3.7 CONCLUSIONS

The opportunities offered by new data sources in political science come with a number of challenges. One set of these are challenges of interpretation; what can we learn from, for example, a directed link graph from a web crawl, and what are reasonable real-world interpretations for the trace data we can now collect? The second set are practical, methodological and technical; what is the best way of obtaining the data, how do we manage the trade-offs that are inherent in different crawl strategies, and what tools can we use to analyse the new data collected?

For the future, it would be of benefit to begin to converge on standard methods for web crawling in political science, if only so the particular distinctive features of each research project can be highlighted rather than exhaustively reporting the nuts-and-bolts decisions made on every occasion in every paper.

²⁸ OpenWayback doesn't yet handle WARC conversion records, so handling datasets which have been transformed to text using Warctika is currently less satisfying than working on untransformed data. The handclassifier framework handles this by extracting the transformed text directly from its storage database where necessary.

In order to achieve this, it would be helpful for those researchers using web crawlers as a data collection method to publish their crawler settings and crawl choices online, to build up a reasonable set of defaults for newcomers to the method to start working with.

This chapter has outlined the beginnings of answers for these questions, at least for this research. There are inevitably a lot of issues outstanding. Political scientists would do well to deliberately seek out the answers which web scientists, computer scientists and others have been developing, some of which transfer almost entirely to this new research domain. The approaches to supervised content classifiers and tools for handling web crawl data developed in this chapter are one attempt to develop tools which can be used by other political scientists, and to demonstrate that they are generally within the reach of quantitatively trained politics researchers who are interested in exploring them.

These issues are discussed, in the context of the study findings, in Chapter 7.

THE STRUCTURE OF GOVERNMENT IN THE DIGITAL AGE

What is the structure, shape, and connectedness of the English local government Internet presence and what are the relationships between different bodies? What can this tell us about the structure of government in the digital age?

4.1 INTRODUCTION: A WEBOMETRIC ANALYSIS OF ENGLISH LOCAL GOVERNMENTS

This chapter answers the first research sub-question, by analysing the structure of UK government web presence. It draws on a new full-population web crawl of .gov.uk, which is uniquely comprehensive and provides a picture for the first time of all levels of UK government online and how they interact.

The structure of the chapter follows the recommendations for a Web Network Thick Description (WNTD) set out by Thelwall (2013, pp.52-53). This gives a holistic account of the structure and interconnectedness of a web network by investigating its shape in both quantitative and qualitative ways. This approach is particularly suitable for answering questions like those tackled here, as it allows for both exploratory and explanatory research to be undertaken as part of a single structured process.

Local government in England follows a complicated but well-established pattern of “two-tier” (small district plus larger county) or “unitary” (all-purpose districts of a variety of types) organisation. Despite this regularity of organisation, there is a wide diversity in council size (Rutland, the smallest unitary, had only 38,000 residents in 2014; Birmingham, the largest, 1.1 million) and attitudes. The ongoing debates about local government’s role mentioned in Section 2.4.3 are mirrored by large differences in the places they serve. As far as the web goes, however, there is much greater uniformity. All principal authorities have websites, almost exclusively at some variation of `www.<councilname>.gov.uk`.

4.1.1 *Why structure is important*

The structure of government is important for two reasons. Firstly, if we accept the theoretical claim that organisations’ public presence reflects their structures and attitudes (or, more weakly, how they want their attitudes to be perceived; or, more weakly again, that links to other parts of government reflect that staff in one area are aware of and consider valuable for their customers this other area), we should find differences in the structure and content of council web sites which reflect their differences.

Secondly, and more specific to the electronic public administration context, some of the features of Digital Era Governance (DEG) should be visible in network metrics. A council demonstrating a needs-based holistic approach to the delivery of public services, for example, should be theoretically expected to present a more joined-up set of web services, with links between services delivered by different agencies meeting the needs of customers in similar situations. Similarly, thoroughgoing digitalisation of services is directly assessable by measuring the extent and quality of web provision (especially, in light of the rhetoric surrounding e-government, that which is transactional).

We should be able to see the variation between authorities played out in their websites both structurally (in terms of link relationships between services and connections with similar authorities and central government) and in terms of textual content. By contrast, if local government has now homogenised to the extent that it is no longer an independent decision-maker for a local area but rather a locally-distributed service delivery agency, we should expect the local government web to be more-or-less uniform.

In an analysis of web links within and between authorities, this study follows the approach of Hale, Yasseri et al. (2014), especially in the analysis of the effect of local geography on linking patterns. In this way, we hope to demonstrate that the linkages between neighbouring authorities are no stronger than they are with more distant authorities, undermining the idea of local linkages and distinctiveness. The pattern of linkages, if it follows the pattern of service delivery relationships, will reinforce the DEG claims of more joined-up electronic government. If not, it will serve to undermine this proposition for these authorities.

“The number of outgoing links reflects the ‘extroversion’ of a website. One might consider that this also reflects the ‘extroversion’ of the associated government entity, and reflective of such things as service delivery partners, related policy and service areas, key stakeholders, or important external information sources related to the content of this website-Government entity” (Henman, Ackland and Graham, 2014b, p.130)

Henman, Ackland and Graham use Kleinberg’s HITS algorithm to calculate Authority and Hub scores to estimate the value and centrality of each page in the crawl. This study manually coded the policy domain of the websites studied, and by tier of government. This approach is less appropriate in this research, and for local governments in which the policy domain varies for each part of the site.

“government web sites (including local government) seem to be tightly managed but allow linking to acknowledge organisational structures, such as to parent or sibling organisations. In addition,

there may be many public service links to useful and trusted sources of information (e.g., on line health sites, bus timetables) or services (e.g., pest control, local attractions) (Holmberg & Thelwall, 2009)” (Thelwall, 2013, p.38)

We can also begin to assess the extent to which the changes suggested made by DEG are happening in practice by assessing the actual government web against the recommendations of needs-based holism, digitalisation and service re-integration, assessing across the local authorities in the web crawl the extent to which these vary between councils.

4.1.2 *Network-centric measures of digital governance*

The distribution of different kinds of web links allows the calculation of various attributes at the domain level. Whether each unit is a source of links (by directing its visitors to other nodes), a sink of links (by receiving visitors via link from other nodes), both, or neither, gives us an insight into the degree of both perceived utility (inlinks), and the degree to which the node is aware of and willing to direct its visitors to other agencies (outlinks). Various other measures to reflect the perception of and centrality of different service areas are also available, such as structure-centric models of network centrality. As Hanneman and Riddle note:

“The number and kinds of ties that actors have are a basis for similarity or dissimilarity to other actors – and hence to possible differentiation and stratification. The number and kinds of ties that actors have are keys to determining how much their embeddedness in the network constrains their behavior, and the range of opportunities, influence, and power that they have.” (Hanneman and Riddle, 2005, p.95)

Network-centric measures are also available. For example, the distribution of indegree between domains allows study of the status and visible homogeneity of local governments as reflected by links from outside their organisation.

Other Social Network Analysis (SNA) concepts and measures are less valuable here. The concept of betweenness centrality, which in some senses measures the

influence of particular nodes as information intermediaries in social networks, provides little useful information in our context. The key reason for this is that links on the web between government agencies are only secondarily signposts between organisations. Although nodality has been held to be a central tool of government in the online world (Hood and Margetts, 2007), we have no direct measure of the authority of different parts of government online. Insofar as the government exists in a public informational space, though, and the government's presence in that space is a reflection of how it wishes to present itself to the world, webometric measures of centrality can be used to assess the nodality of different parts of the government both within government itself, as presented to the public, and also its nodality within the wider Internet.

Conceivably, an attempt could be made by running some kind of Pagerank-type analysis on the government web (Page et al., 1999). It is complicated in the present case by a lack of collected inlinks from outside the government itself. A more conceptual problem is caused by the local nature of local government bodies: each may well have local authority and nodality, but there is no obvious way to detect central government websites deferring to "local government in general" on any particular topic, because of the difficulty of linking to several hundred separate websites for each query. The lack of a single co-ordinated local.gov.uk could in this way be substantially impeding the ability of local authorities to receive traffic from central government bodies even in areas for which local government formally holds responsibility.

Although not central to the DEG thesis, another interesting measure on this dataset is reciprocity: the extent to which the (directed) web links flow from local government to central government, central to local, or in other patterns.

4.1.3 *Hypotheses*

These hypotheses are established to test different aspects of the underlying problem and analyse the primary evidence for answering research sub-question 1.

Hypothesis H1 tests the evidence for local government homogeneity: that the importance of each local authority to the network is largely a factor of their type and population. This would be expected if organisations' online presences were equivalent in quality and scope.

Hypothesis H1: Local authorities' network importance, measured by indegree centrality, is substantially explained by council type and population.

Hypothesis H2 tests the proposition that in government "the organisational socio-technical system is becoming increasingly manifest on the web" (Dunleavy and Carrera, 2013, p.228) and therefore that it is feasible to look at the online presence of government agencies as a reflection of their structures or a proxy for their offline activities.

Hypothesis H2: Organisations on the network will be logically clustered by the structure of government, not by shared function or common citizen needs.

Hypothesis H3 tests the proposition that the government web is at least minimally joined up, with councils and other government bodies linked into a single link network. This speaks to DEG notions of holism.

Hypothesis H3: The government web represents a single link network; there will be no isolates.

Hypothesis H4 tests the importance of the specifics of local geography in online local government, hypothesising that the heterogeneous localities in which local governments operate are responsible for cross-council linking. This implies that geography is more important to local governments' online presence than the local.gov.uk proposals give credit for.

Hypothesis H4: Web linking between local authorities is substantially explained by geographic proximity.

Hypothesis H5 hypothesises that web linking is principally structural. The DEG proposition of holism, by contrast, suggests links should reflect a desire by services to meet the wider needs of their citizens by directing them towards appropriate services from other agencies. Two separate approaches are used for testing this hypothesis: first by expert assessment of the purpose of links by hand classification, and second by looking at the propensity of links to be to an organisation's front page (suggesting structural links) or directly to an in-site service or information page (suggesting signposting). As the second approach is novel, two separate methods are proposed for triangulation.

Hypothesis H5: Web links between government bodies primarily reflect structural ties rather than service signposting, as reflected by hand classifications and the proportion of front page to deep links.

These hypotheses are framed as what might be described as a sceptical view of local authority website importance and function. Hypothesis H1 implies homogeneity between local authorities' network importance, with no impact from the (known to be variable) extent of web development or of any other reason. Hypothesis H4 likewise implies that there is no linking of citizens to useful informational content hosted by other local authorities, rather that each authority is a stand-alone island in the web, giving links to other councils only to reflect the intensely local ties that come from shared physical proximity. Hypothesis H5 likewise implies that links are made not for the benefits they may give to service users, but to reflect structural properties of the organisations that made them.

4.2 PROCESS

The first stages of the WNTD process require clarity about the precise scope of the data to be collected, the units of analysis to be used, and the processes used

to collect it. This section outlines these choices, and gives a detailed account of the dataset created to underpin the analysis.

4.2.1 *Data choices*

Webometric studies analyse the network created by a group of ‘nodes’, joined by ‘edges’.

The nodes for this analysis are groups of websites collected at the third-level domain level (that is, everything under a given x.gov.uk). The substantive reason for this choice is that third-level domains map fairly well onto organisations, at least at the local government level. For a public administration focused approach, the key questions do not arise at the level of the individual web page, but in understanding the choices made by organisations looking as a whole. For similar reasons, conducting analysis web server by web server (by treating <http://applications.huntingdonshire.gov.uk/> and <http://www.huntingdonshire.gov.uk/> as separate objects) adds a great deal of variation as a side effect of the way different local authorities break up their web services over multiple domain names.

As a further consideration, treating the network at third-level domain scale significantly reduces the computational and interpretive difficulties that would arise from treating 11 million web pages as the primary units of interest. This is particularly beneficial when using computationally expensive measures such as betweenness centrality¹.

The size and link density of the .gov.uk space is potentially an issue, as the analysis of web networks is made more difficult by large scope. To get a conceptual handle on the underlying relationships between different parts of the government web, a possible method is the use of a clustering algorithm to

¹ The standard Brandes algorithm for calculating betweenness centrality has a time complexity of $O(mn + n^2 \log n)$ (Nasre, Pontecorvi and Ramachandran, 2013), orders of magnitude slower with c.11,000,000 pages as nodes than c.2,300 domains as nodes.

group nodes into communities of interest; Henman, Ackland and Graham (2014a) used Infomap to cluster a similar dataset with interesting results².

For studies specifically of the web, the natural choice of edge definition is the hyperlink: a purposive link made between two sites. As a hyperlink is directional (it is made by a ‘source’ site to a ‘target’ site, who may or may not know that the link has been made) then the resulting graph is directed³.

To improve the robustness of the dataset and to reduce the complexity of analysis, it is common to apply the Alternative Document Model (ADM) to raw web crawl data. This dimensionally-reduces the data by collapsing all nodes in a group to a single node (e.g. one node per third-level domain) and binarising the edges (so instead of a count of connections between two collapsed nodes the metric is a simple ‘connected’ or ‘not connected’). The ADM can be applied as a method of cleaning data for more reliable link counting (see Holmberg, 2009, pp.69-71). Sometimes the key need is “to reduce the connections in the network to a level that would make it possible to extract patterns” (Thelwall, 2013, p.61). ADM can be applied either at the level of the page (in which case each page has a maximum of one link to each target), or at the more general directory or site levels. The latter have the attraction of reducing the network size as well as clarifying the structure.

A key choice is then the degree of data reduction to be applied to the crawl data. The units of analysis in the research as a whole are local governments, which suggests aggregating data to site or domain level in order to get an

2 The algorithm was selected by the authors on the pragmatic grounds that it handles graphs which are both directed and weighted – that is, different links can be assigned to be of different “strength” by use of edge weights (Henman, Ackland and Graham, 2014a, p.136). If the ADM is also used (see below) then the latter is not a constraint.

3 In the language of network analysis, a “directed” graph is one where links between nodes have an inherent direction: they are *from* one node *to* another. If the link is symmetrical, a pair of directed links is made pointing in both directions. The web is a classic directed graph. An “undirected” graph is one where ties are intrinsically bidirectional. One example of this is a network formed with links based on geographical proximity: that A is physically proximate to B implies that B is also physically proximate to A. The geographical analysis in Section 4.3.3 combines the directed government web link graph with the undirected local government proximity network. Some of the tools available for network analysis make better use of the extra information held in a directed graph than others.

overall comparative picture. Nevertheless, for some parts of the analysis it will be sufficient (and more convenient) to look at domain level rather than aggregating to site level. This is particularly important when looking at links between local government and non-local government bodies, for which a comprehensive mapping is not possible⁴. For this analysis, ADM will be applied at third-level domain level, dimensionally reducing the network under study to 2,275 .gov.uk domain nodes and binarised links between them.

4.2.2 *Data set selection and collection*

The second-level domain .gov.uk is selected as a reasonable proxy for the UK government web: the boundaries of government are ill-defined, especially in areas such as local partnership work. Although there are some quasi-governmental sites outside .gov.uk, that domain contains essentially all major local government web presence. In some sense, this is accepting the definition of what should be seen as a government web site made by the government themselves, who control registrations in .gov.uk. A full list of domains registered in .gov.uk is conveniently available as the result of a third party's Freedom of Information Act request; this study starts the crawler using the October 2013 list as a set of seeds (Cabinet Office, 2013).

For analysis, English local authority websites have been identified and separated (the method is fully described in Section 4.2.3).

There is no published standard for the collection and processing of web crawl data in political science. Consequently, it was necessary to develop a framework for this research drawing on the general webometric literature.

Social scientists doing web crawling in the past have generally done so using small-scale crawlers and scrapers which are optimised for collecting

⁴ In addition to the greater work required to replicate the Section 4.2.3 analysis for all bodies in the government, the boundaries between the various different government departments, executive agencies, and Non-Departmental Public Bodies are much murkier than they are for local government.

high-quality content from a relatively small selection of sites such as SocSciBot (Thelwall, 2009). Previous local government studies have tended to rely on either a non-random convenience sample or on the ability to manually examine the whole site for a small number of authorities (Baldersheim and Øgård, 2008; Holmberg, 2009). Clearly this is only effective where web sites are much smaller than the English context. For a full-population crawl of this size, it becomes necessary to take a step up in scale; it requires a heavyweight multi-threaded crawler which can cope with the standards-breaking pages found on the live web. The Internet Archive’s web crawler Heritrix (Internet Archive, 2013) was selected. This is highly configurable for a variety of purposes and allows a flexible choice of what to be crawled.

A high-level view of the web crawling process is set out in Figure 2. In simple terms, Heritrix was set to crawl the whole of .gov.uk, with exclusions specified below, and was monitored to ensure that various misconfigured web servers did not cause runaway crawling. The data were processed through the new Warctika library to extract plain textual content (see Section 3.6.1), and the content then stored in structured Web Archive (WARC) files (ISO, 2008)⁵ for long-term storage. Data were collected continuously between 14 April and 29 October, 2014.

The following major configuration settings were selected:

- Only images, files with a textual Content-Type (such as plain text, Hypertext Markup Language (HTML), Cascading Style Sheet (CSS), and Rich Text Format (RTF) files), and those in office document formats, were fetched⁶. In particular, video and audio files were not fetched or were discarded. Care was taken to work around web servers which returned non-standard Content-Type headers.

5 Although this document is a working draft and not the final published standard, this appears to be the version actually implemented by all WARC tools including Heritrix – possibly because the draft is readily available online and the final standard is expensive.

6 Images, javascript and CSS files were obtained to support use of the OpenWayback system to view and classify the data (see Section 3.6.2). They were excluded from other analyses as they are presentational and contain no substantive textual content.

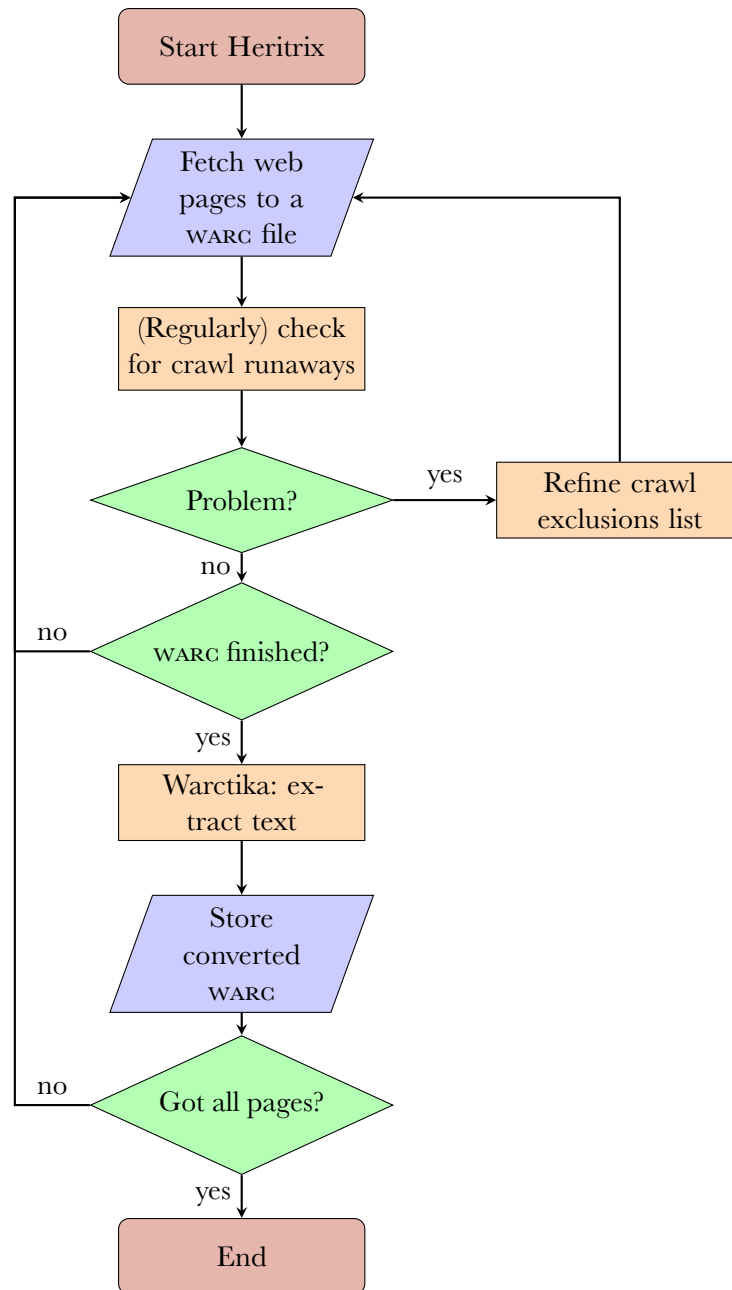


Figure 2: Web crawling process

- The site owners' robots.txt files were respected (see Section 3.4.4). These files are part of a standard which allows site owners to indicate that certain material may not be crawled, generally either to prevent the infinite crawling of algorithmically generated Uniform Resource Locators (URLs) or to prevent certain content from being indexed by search engines (for more details, see robotstxt.org, 2007).
- Documents greater than 100MB in size (generally videos with incorrect Content-Type headers) were rejected.
- The default restrictions were applied to prevent infinite crawl regress. In addition, during the crawl various additional issues were identified with sites leading to infinite regress or massive duplication of collection. In these cases, hand-coded rules were carefully added to restrict the crawl scope to stop the regress. The erroneously collected pages were later erased from the archive by re-running the final set of exclusion rules on the final set of WARC files. The most common cause of this problem was sites offering badly-constructed robots.txt files which failed to prevent access to unsuitable database-driven parts of the site.

There are a number of sites in the crawl that are unadvertised duplicates. That is to say, rather than pages on <http://www.site.gov.uk/> redirecting to <http://site.gov.uk> or vice-versa, both URLs happily serve identical pages without advertising that the other site exists. One example of this in our dataset is Transport for London, which runs both <http://www.tfl.gov.uk/> and <http://www.transportforlondon.gov.uk/> as separate but mirrored entities. Once such duplicates have been identified, one of the pairs has been excluded from scope.

The extraction of web links for analysis from the .gov.uk web crawl has been made easier by the use of the Internet Archive's own WARC processing tools. In particular, a pipeline has been established by which WARC files are run through Warctika to extract text from the Portable Document Formats (PDFs) and Office files, then the Internet Archive's own intermediate file format the

Web Archive Transformation (WAT)⁷ is produced from the WARC files. Two separate sets of data were then extracted from the WATs as standard Comma Separated Values (CSV) files: firstly the link graph, treating each page as a node and each link as a (directed) edge⁸; secondly the nodelist, treating each page as a node and extracting companion attribute data including the success or otherwise of the fetch (HTTP response code), the HTTP Content-Type of the object fetched. Both sets of data were then subjected to ADM and re-reduced to produce domain-level encapsulations of the websites overall.

Once crawling was complete, a data cleaning process was carried out (shown in Figure 3). The raw node lists were checked by hand to get a feeling for the data and to identify latent crawl problems. Specific checks were made for domains serving duplicate content either by two protocols (http/https:), by duplicating the bare domain and www (www.domain.gov.uk and domain.gov.uk silently serving duplicate content), or by multiple sub-domains hosting copies of the same site (democracy.domain.gov.uk and moderngov.domain.gov.uk could be duplicates of the committee management system, for example). All instances of at least 100 pages being hosted in duplicate in this way were systematically identified. Further exclusions included failed fetches⁹ and most parameterised URLs¹⁰. This part of the process was hugely time-consuming; this is discussed in more detail, and the consequences explored, in Section 7.5.4.

Once the data had been fully viewed and further exclusion rules were written, the data were re-extracted from the raw WARCS to produce a clean dataset. A summary of the nodes and edges contained within the web network is in Table 3.

⁷ WAT files are similar to WARC files, except that in place of the usual WARC payload of Hypertext Transfer Protocol (HTTP) requests and responses, a Javascript Object Notation (JSON) object is carried which contains the resource's outlinks and metadata in an easily machine-readable format.

⁸ Links were validated and those not syntactically valid were removed.

⁹ HTTP response codes not in the 200 series

¹⁰ URLs with '?' parameters overwhelmingly represent database queries and form submissions, neither of which are valuable for this crawl. Exceptions were made for the various web frameworks that publish ordinary pages through parameterised URLs, generally at variations of http://www.domain.gov.uk/?page_id=100.

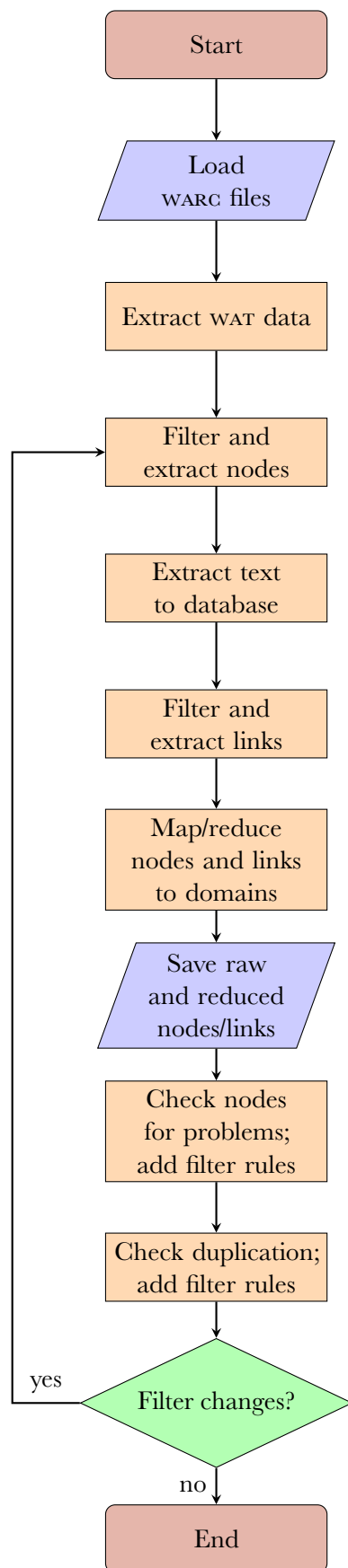


Figure 3: Data cleaning process

Content	Description
Nodes – URLs	Each third-level domain in .gov.uk
Edges – hyperlinks	Every detected link between pairs of nodes as defined above. Also every outlink from nodes in our dataset to sites outside .gov.uk which were not fetched. We do not have inlinks from outside .gov.uk to sites in the sample

Table 3: Web crawl dataset: final content

As a final stage, sites with fewer than 10 pages were excluded from the majority of the analysis. The majority of these are “site moved” notices for defunct domains, with others thin brochure sites for local partnerships and parish councils.

The node and edge data is available for wider research use or for replication of the results of this study, together with a third-level domain ADM network including ownership attribute data and calculated web metrics, the pages’ textual content, and the raw WARC files from which the text was derived. This dataset is too large, and too copyright encumbered, for UK Data Archive archiving; contact the researcher for access. This, as a barrier to replication and future use of crawl data, is discussed in Section 7.5.4.

4.2.3 *Mapping from websites to organisations*

This section uses partial free-text matching of registrant information from the WHOIS database to identify common ownership of multiple domains by the same local governments. This is based upon a new approach developed by the researcher for work in the Domain Name System (DNS) policy field (Nicholls, 2013).

The WHOIS system is a global network of free-text databases, containing information about domains registered in the DNS and the people and organisations who have registered them. Details are available in its standards document, RFC 3912 (Daigle, 2004). For structured information retrieval, WHOIS suffers from two serious limitations: data is variably available for different domains, and

the free-text format prevents easy parsing of registrant data. Fortunately, the whois database for .gov.uk is provided uniformly by the academic networking organisation JANET rather than being devolved to multiple providers as in .com. This means that coverage is essentially complete, the format is consistent, and therefore parsing the responses is much easier than found in the previous study.

For each .gov.uk domain in the dataset, the whois database was queried for registrant details¹¹. The “Registered For” field was parsed from the response and compared using a Levenshtein fuzzy-matching algorithm to a standardised Office for National Statistics (ONS) list of local authority names for the period under study. Finally, each domain was checked by hand. The steps followed are presented graphically in Figure 4.

In common with Henman, Ackland and Graham (2014b), domains were hand-coded by type of organisation and then further sorted into classes according to Table 4. The distribution is shown in Figure 5.

4.3 FINDINGS

4.3.1 *Node centrality and importance*

Unsurprisingly, the domains with the highest indegree centrality here are key central government bodies, with all local government bodies well down the table. Table 5 gives the top forty domains, along with their size. The sole non-central government bodies in the list are the main domains for the Scottish and Welsh devolved governments, and two joint organisations for local government. Interestingly, both of these (the Trading Standards Institute and the Local Government Association) are membership organisations for trading standards officers and councils respectively and provide few services

11 An exception was made for www.gov.uk, which is hosted off the main domain delegated directly from Nominet to the Cabinet Office and for which the whois data is in a different format.

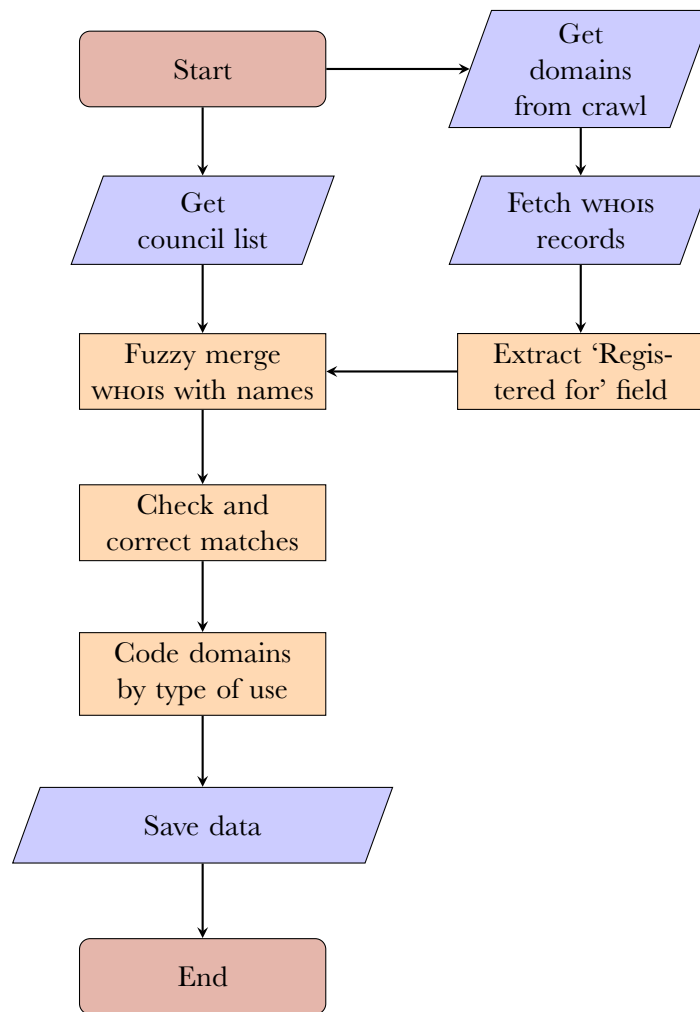


Figure 4: Domain ownership coding process

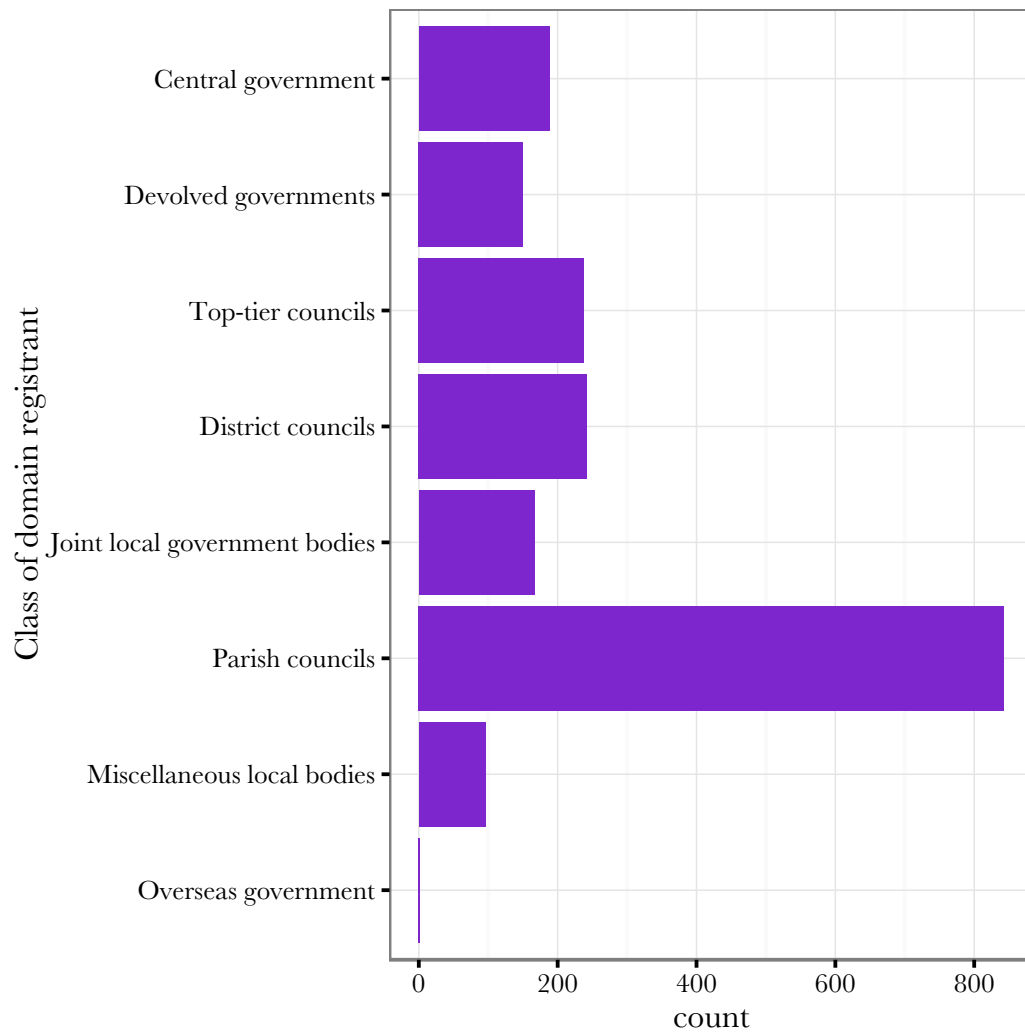


Figure 5: Domain classifications

Class	Type of body	n	total
Central government	Central government bodies including quangos, judicial bodies, and independent enquiries	185	185
Devolved governments	Scottish Government	72	150
	Welsh Government	18	
	Northern Irish Executive	54	
Top-tier councils	County councils	41	171
	English unitary councils (including Metropolitan Boroughs, London Boroughs and the City of London)	130	
District councils	English shire district councils	212	212
Parish councils	English parish (including those styled ‘town’ and ‘city’) and Welsh community councils	828	828
Non-English councils	Welsh unitary councils	32	94
	Scottish unitary councils	34	
	Northern Irish district councils	28	
Miscellaneous local bodies	National Park Authorities	14	95
	Police and Crime Commissioners	26	
	Fire Authorities	32	
	Inshore Fisheries and Conservation Authorities	9	
	Waste Disposal Authorities	6	
	Internal Drainage Boards	8	
	The Council of the Isles of Scilly (<i>sui generis</i>)	1	
Joint local government bodies	Joint boards, joint committees, and shared services bodies belonging to multiple local authorities	165	165
Overseas governments	Anomalously, both the London office of Gibraltar and the Latvian embassy have .gov.uk domains	2	2

Table 4: Domain classifications

to the public. Both have risen on the back of widely-distributed links from local government websites despite their relative lack of public utility. The differing motivations of linking and what it might mean for government website visibility are analysed in Section 4.3.4 below.

Table 6 shows the data just for principal councils in England, including the top and bottom 20 domains.

The bottom positions are dominated by secondary sites: that is, domains which are either specialist sub-sites on separate domains (e.g. *norfolk-edunet*.

	domain	indegree	size	type
1	www.gov.uk	723	262,300	Central government
2	direct.gov.uk	696	72,782	Central government
3	legislation.gov.uk	588	2,505	Central government
4	environment-agency.gov.uk	532	2,769	Central government
5	defra.gov.uk	512	99,788	Central government
6	nationalarchives.gov.uk	503	95,092	Central government
7	hse.gov.uk	478	38,554	Central government
8	communities.gov.uk	457	70,395	Central government
9	metoffice.gov.uk	451	61,440	Central government
10	homeoffice.gov.uk	447	2,609	Central government
11	hmrc.gov.uk	432	87,824	Central government
12	planningportal.gov.uk	414	33,436	Central government
13	food.gov.uk	406	60,419	Central government
14	businesslink.gov.uk	404	407	Central government
15	dft.gov.uk	373	23,481	Central government
16	dwp.gov.uk	372	3,342	Central government
17	justice.gov.uk	370	15,343	Central government
18	voa.gov.uk	351	5,314	Central government
19	audit-commission.gov.uk	336	41,227	Central government
20	forestry.gov.uk	318	44,732	Central government
21	bis.gov.uk	305	25,953	Central government
22	statistics.gov.uk	302	506	Central government
23	cabinetoffice.gov.uk	293	12,220	Central government
24	dh.gov.uk	293	5,112	Central government
25	culture.gov.uk	292	12,741	Central government
26	highways.gov.uk	291	6,361	Central government
27	ons.gov.uk	279	29,592	Central government
28	tradingstandards.gov.uk	278	3,778	Joint local government bodies
29	education.gov.uk	260	45,270	Central government
30	ofsted.gov.uk	259	107,828	Central government
31	data.gov.uk	249	38,112	Central government
32	gamblingcommission.gov.uk	228	2,742	Central government
33	landregistry.gov.uk	226	2,313	Central government
34	decc.gov.uk	219	28,058	Central government
35	consumerdirect.gov.uk	214	111	Central government
36	local.gov.uk	213	65,522	Joint local government bodies
37	oft.gov.uk	206	12,629	Central government
38	scotland.gov.uk	202	324,389	Scottish government
39	wales.gov.uk	183	27,573	Welsh government
40	berr.gov.uk	177	1,051	Central government

Table 5: Node importance: top 40 domains by indegree

gov.uk, and southribbletourism.gov.uk – these are suitable for linking to by a far smaller number of external bodies), or preserved domains from defunct subsidiaries (sedgefield.gov.uk and teesdale.gov.uk were both registered by district councils merged into Durham County Council in 2009), or alternative URLs to access the same council content (e.g. Ashfield District Council have ashfield.gov.uk but use ashfield-dc.gov.uk as their main domain).

A breakdown of the differing indegree centrality of different types of domain is given in Table 7 and graphed in Figure 6 (noting the logarithmic y scale).

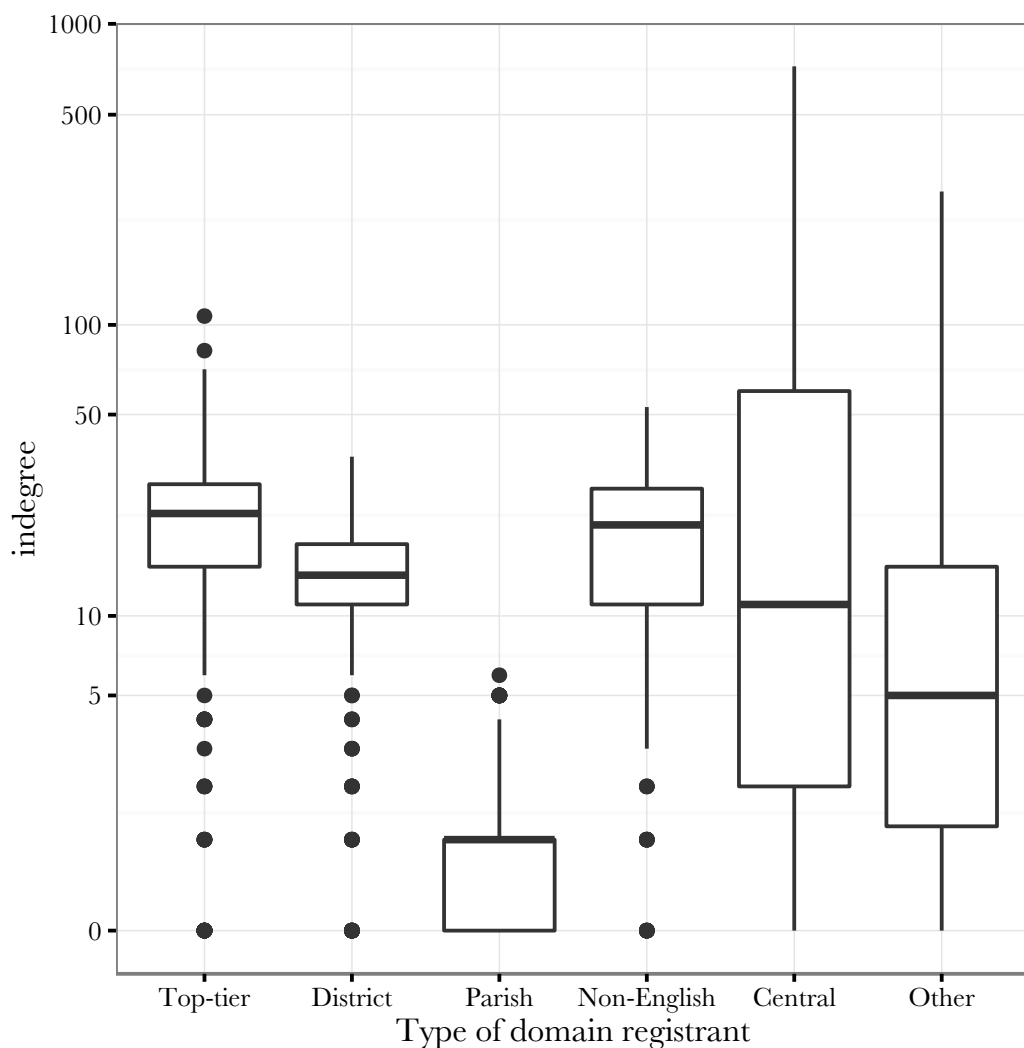


Figure 6: Indegree centrality by domain type

To test Hypothesis H1, an Ordinary Least Squares (OLS) regression model was fitted, modelling the effect of council type and population on indegree cen-

rank	domain	indegree	size	type
1	hants.gov.uk	107	55049	County
2	cityoflondon.gov.uk	82	32914	Unitary
3	devon.gov.uk	71	33917	County
4	warwickshire.gov.uk	67	79569	County
5	kent.gov.uk	65	18207	County
6	surreycc.gov.uk	65	64212	County
7	derbyshire.gov.uk	62	10935	County
8	cornwall.gov.uk	61	19321	County
9	westsussex.gov.uk	61	21177	County
10	birmingham.gov.uk	58	16425	Unitary
11	bucksc.gov.uk	57	16612	County
12	staffordshire.gov.uk	56	17367	County
13	essex.gov.uk	55	9881	County
14	lancashire.gov.uk	52	23991	County
15	wiltshire.gov.uk	51	69633	County
16	oxfordshire.gov.uk	50	57390	County
17	norfolk.gov.uk	48	6126	County
18	somerset.gov.uk	47	33734	County
19	northyorks.gov.uk	46	41328	County
20	gloucestershire.gov.uk	45	18895	County
...				
367	wyreforest.gov.uk	3	4175	District
368	ashfield.gov.uk	2	1781	District
369	askderbyshire.gov.uk	2	469	County
370	dorsetcc.gov.uk	2	20	County
371	eastdorset.gov.uk	2	9625	District
372	mansfield-dc.gov.uk	2	2435	District
373	myoxfordshire.gov.uk	2	103	County
374	sedgefield.gov.uk	2	299	Unitary
375	visitsouthribble.gov.uk	2	16	District
376	carlisle-city.gov.uk	1	2256	District
377	norfolkedunet.gov.uk	1	39	County
378	walthamforestclass.gov.uk	1	199	Unitary
379	westdorset-dc.gov.uk	1	32	District
380	wirral-mbc.gov.uk	1	2671	Unitary
381	essexlegalservices.gov.uk	0	312	County
382	newforestdistrict.gov.uk	0	11889	District
383	southribbletourism.gov.uk	0	17	District
384	teesdale.gov.uk	0	338	Unitary
385	westwey.gov.uk	0	19	District
386	worcestershirehub.gov.uk	0	376	County

Table 6: Node importance: local authorities in England

class	indegree	sd
Top-tier	24	16.13
Non-English council	22	12.66
Central government	22	149.42
District	15	5.71
Other	7	29.05
Parish	1	1.04
Overall	3	55.69

Table 7: Median indegree by domain type

trality of that council’s websites. Indegree was transformed to $\log(\text{indegree} + 1)$, as indegree is logarithmically distributed (see Figure 6) and some domains have indegree 0 so fitting a direct log is not possible. This model, Model 1, is shown in Table 8.

	$\log(\text{indegree})$	
	(1)	(2)
County council	0.576*** (0.203)	0.371** (0.162)
Unitary council	0.415*** (0.086)	0.185*** (0.070)
Population	0.000 (0.000)	0.000 (0.000)
$\log(\text{Number of web pages})$		0.291*** (0.019)
Constant	2.634*** (0.054)	0.122 (0.173)
N	386	386
R^2	0.114	0.443
Adjusted R^2	0.107	0.437
Residual Std. Error	0.678	0.538
F Statistic	16.341***	75.640***

* $p < .1$; ** $p < .05$; *** $p < .01$

Base category is “District council”.

Table 8: Regression models: indegree by English council type, population, and number of pages

This model is not a good fit for the data. Only around 11% of the variation in indegree is explained by population and council type. Although county and unitary councils do have a higher indegree than the district councils which are the base category, as hypothesised, this effect is small. The effect of population is so small to be negligible. This model does not support Hypothesis H1; other explanations should be sought.

Another key measure of site importance is size. All other things being equal, better connected and better trafficked sites from government bodies with more importance should be larger. In this dataset, the distributions of both indegree and site size are decidedly non-normal: indegree is consistent with a power law distribution¹². This is consistent with previous findings in the study of web structure, which consistently find power law distributions where network effects affect patterns of connections. In this case, nodes with high indegree are more visible, which leads to greater linking and ends in the kind of hugely disproportionate connection patterns found here.

The association between indegree and size for English principal authorities was plotted on Figure 7 for English councils and comparatively on Figure 8 for other government organisations. At first glance, a weak association appears to be present for councils, but no obvious association for non-council bodies.

The association for English local authorities was tested using Spearman's rank correlation; $\rho = 0.65$ [$p = 0$], confirming the apparent association shown on Figure 8¹³ on the non-council plot. Nevertheless, the preponderance of central government domains in the upper left quadrant of that graph shows that their disproportionate showing amongst the most indegree central domains is not simply a function of size.

¹² Albeit one with a small secondary peak around $\log_{10}$ indegree = 1.3.

¹³ Note the logarithmic scaling on both axes: if plotted linearly the data are mainly a dense blob in the bottom left hand corner. The points have also been jittered to make visible the huge number of points with $0 \leq \text{indegree} \leq 2$

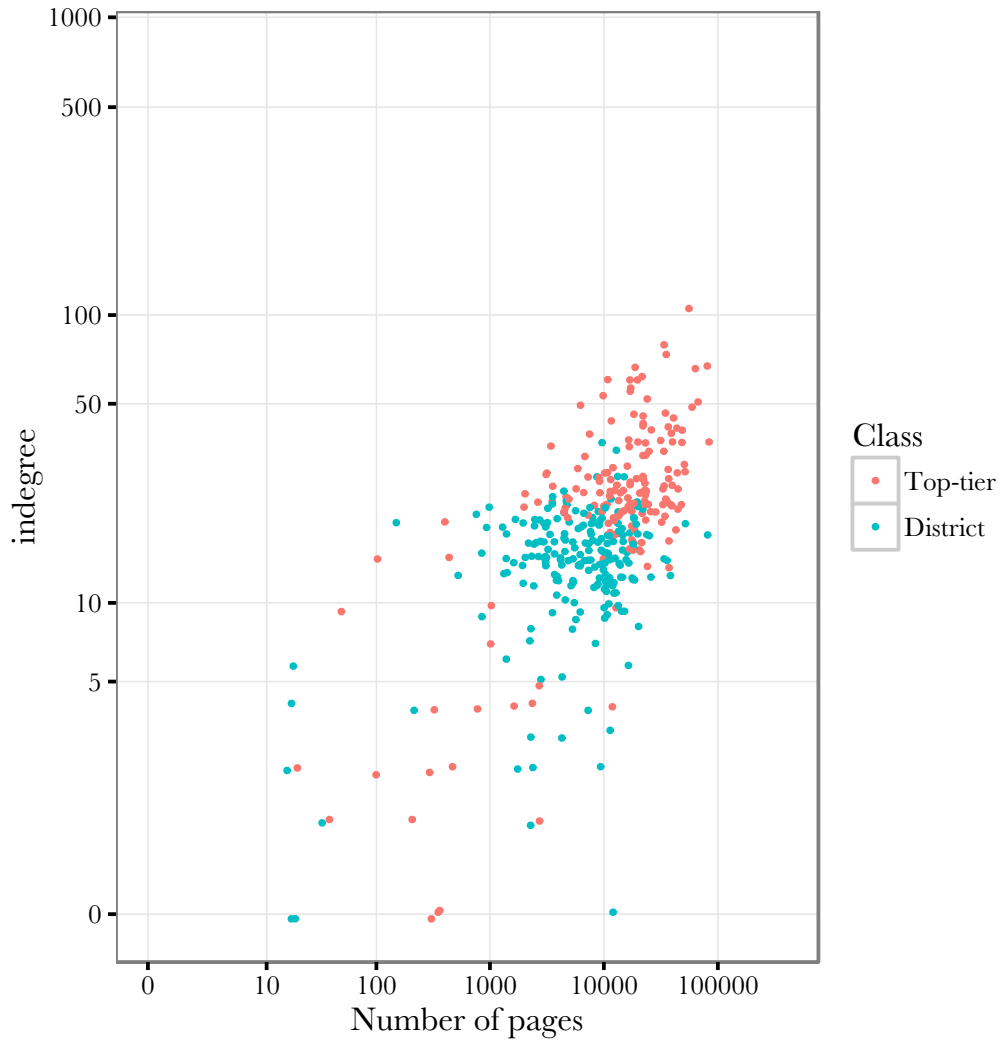


Figure 7: Indegree centrality by size: English councils

To see the effect of size on indegree compared with population and council type, a second OLS regression model (Model 2, Table 8) was fitted, adding $\log(\text{number of pages})$ to Model 1.

The results of this model are markedly better than before. Adjusted R^2 has increased to 0.44, showing that four times as much of the variation is explained by the model including site size as was explained by population and council type alone. The effect size of the population term is still very small, while the coefficients of the council type dummy variables have markedly shrunk.

This confirms the rejection of Hypothesis H1 suggested above. The size of each site, plus a significant amount of unexplained variation, explain more

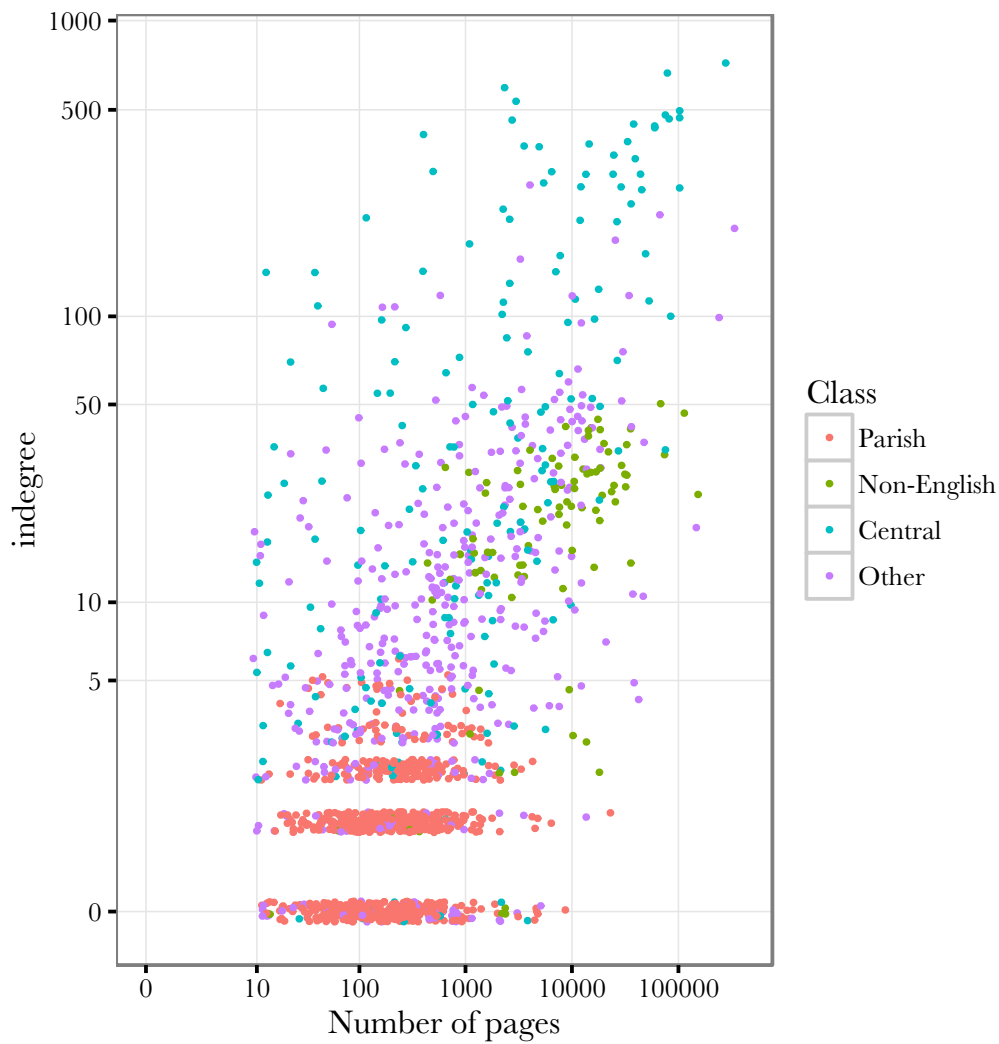


Figure 8: Indegree centrality by size: other domains

of the variation in indegree than population or council type. This suggests that the variation in site importance is down to variations in Internet content and placement, rather than simple demographic factors, and therefore that the relative importance of councils' presence online is related to council heterogeneity.

One caveat with this analysis is that it is being conducted at domain level, rather than taking the whole set of each authority's websites as a whole. This is because the key dependent variable, indegree, makes most sense when treated at domain level rather than being aggregated. Nevertheless, this means that the impact of small supplementary domains will have some impact on the

results. As the effect sizes found are large, however, this is unlikely to result in different interpretations being justified.

4.3.2 *Network communities: Infomap link clustering*

Clusters from the .gov.uk network were detected using Infomap (Rosvall, Axelsson and Bergstrom, 2009), which is considered one of the most effective approaches of its kind (Lancichinetti and Fortunato, 2009). It is particularly appropriate for work on web networks because it constructs hierarchical clustering making use of the directed nature of web graphs. Furthermore, its information theoretic mathematical base models prospective flows of information on the network; competing modularity-maximising approaches are based on optimising a parameter which is more loosely based on real-world network features of interest. The top level clusters are presented in Table 9.

The clustering for this study produced a single-level clustering with very loose structure, showing that there is little information in the patterns of connectivity. What structure there is, however, is strongly geographical: the clusters that arose are based around groups of governmental entities in a given area. The most prominent of these are the large clusters of Scottish, Welsh, and Northern Irish nodes; these clusters show a relatively dense interconnectedness between the devolved governments, their specialised agencies, and their local authorities.

Less pronounced but still observable clustering effects are also visible in England, where there are several clusters containing a county council, the various district councils in the counties, and also parish councils from the area. These include clusters for Devon, Cornwall, Warwickshire, and Somerset. There are also smaller clusters of just a district and its parishes, such as Huntingdonshire and East Hants. A further number contain groups of parishes which strongly interlink but without the parent authority, such as the Kent and Northamptonshire parish clusters. The one minor functional cluster

which can be seen from the Infomap output is the ‘Inshore Fisheries’ cluster, containing all the nine Inshore Fisheries and Conservation Authorities in the study¹⁴.

There are two large groups for which this geographic clustering is not apparent, though. The biggest, the cluster labelled ‘Core’, contains 1043 domains (more than half the total) including the large majority of UK central government sites, council domains, and also joint and miscellaneous bodies. That there is such a large undifferentiated block of nodes suggests that there is little pattern in the link structure. The second group is the 145 domains which are clustered as singletons or in trivial clusters of two or three nodes. Typically, when given a network with 1900 nodes, Infomap will recursively detect smaller sub-clusters which are characterised by tighter interconnectivity with each other than with other nodes in the cluster. That this pattern has not emerged in this study suggests that there is, in fact, little information for the researcher in the link topology of the government web.

Figure 9 gives a graphical representation of the .gov.uk third-level domain network with node areas proportional to the number of pages in the domain. For this diagram, all domains with fewer than 1,000 pages have not been plotted, with the remaining non-trivial Infomap clusters represented by colour. It is apparent that the network has a single centre rather than a hierarchical clustering or some other polycentric structure.

There is no form of hypothesis test that can give a quantitative answer to Hypothesis H2. Nevertheless, the outcome is striking: clustering in the .gov.uk webspace is so weak that neither of the alternative explanations given are valid. Not only are structural linking patterns too weak to be discernable, but there is little evidence of clustering by function or by citizen needs either. Rather, the evidence supports a view of .gov.uk domains as weakly clustered by geography but otherwise an unstructured group.

¹⁴ Including, intriguingly, the Isles of Scilly Council, which is only an IFCA incidentally to its main duties as a *sui generis* local authority.

	Domains	Largest domain	Descriptive label
1	1,093	www.gov.uk	Core
2	130	scotland.gov.uk	Scotland
3	108	carmarthenshire.gov.uk	Wales
4	82	niassembly.gov.uk	Northern Ireland
5	56	devon.gov.uk	Devon
6	49	essexcc.gov.uk	Essex
7	33	newforest.gov.uk	New Forest
8	29	hertsmere.gov.uk	Hertfordshire
9	29	cornwall.gov.uk	Cornwall
10	26	somerset.gov.uk	Somerset
11	26	wiltshire.gov.uk	Wiltshire
12	17	stroud.gov.uk	Stroud
13	15	kentparishes.gov.uk	Kent parishes
14	14	earlsbarton.gov.uk	Northants parishes
15	10	easthants.gov.uk	East Hants
16	8	nifca.gov.uk	Inshore Fisheries
17	8	huntingdonshire.gov.uk	Huntingdonshire
18	8	southbucks.gov.uk	South Bucks
19	6	wigan.gov.uk	Wigan regional
20	6	bridlington.gov.uk	East Yorks parishes
21	5	tyneandwearltp.gov.uk	Tyne & Wear regional
22	4	yaptonpc.gov.uk	Arun
23	4	eastdorset.gov.uk	East Dorset
24	4	ferryhill.gov.uk	Sedgefield
25	4	warp.gov.uk	ICT bodies
26	152		Singletons and cluster size ≤ 3

Table 9: Infomap network clusters: .gov.uk

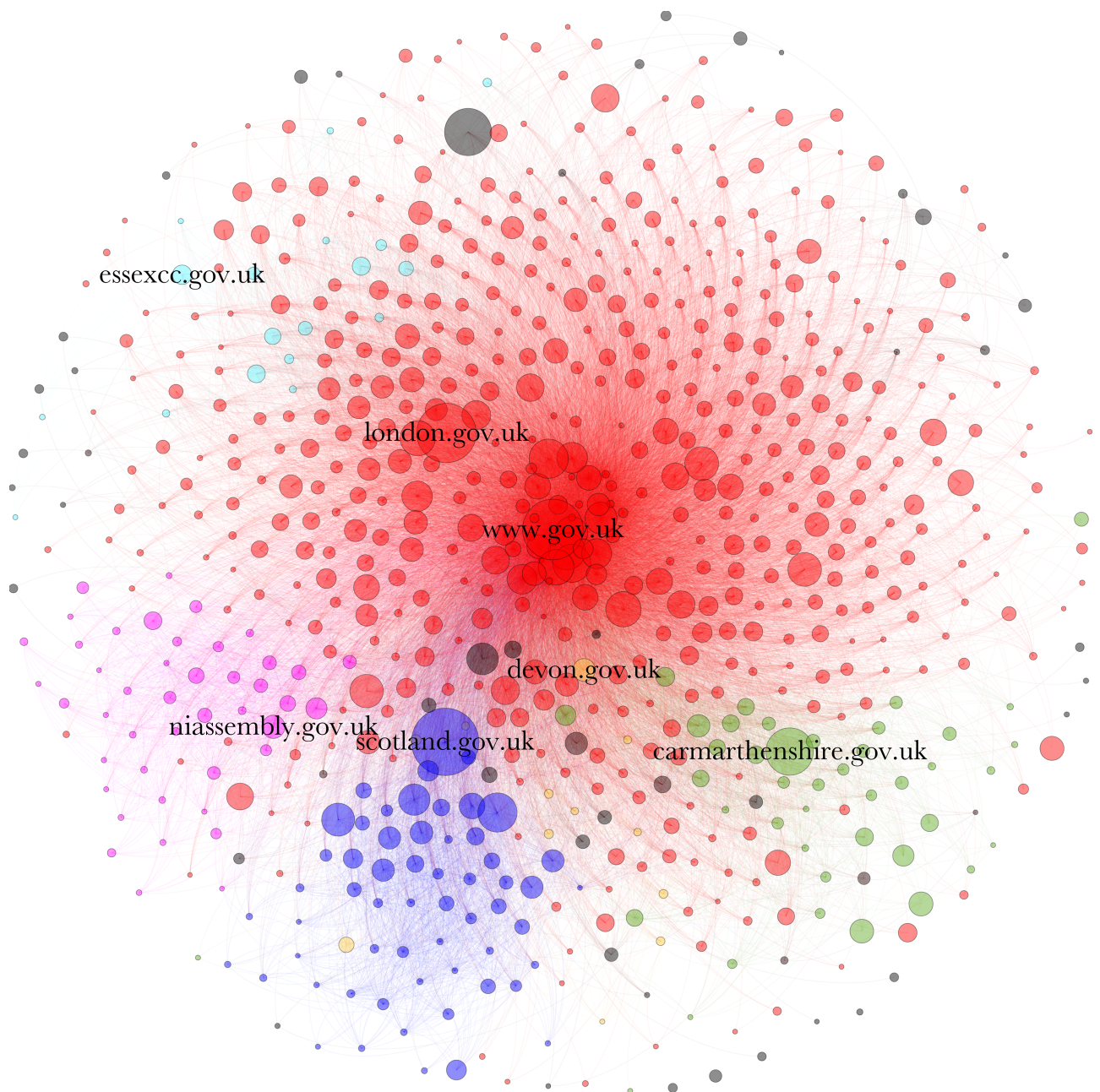


Figure 9: Third-level domain diagram, coloured by cluster

This is a surprising finding. It is unlikely that the result is artefactual, as the dataset coverage of links between government bodies is very good. It is plausible that applying a less rigorous data reduction process (perhaps by applying ADM at page level and analysing the resulting multi-million node network) would reveal slightly more, though it's more likely that it would just reveal that pages cluster tightly at site level.

Perhaps the principal difficulty is one of interpretation. With neither of the expected clusterings being revealed in fact, the question of what the unicentric and non-hierarchical distribution of sites means in practical terms is slightly troubling. Clearly sites are unlikely to be linking to each other at random, which is the naïve interpretation: there is no plausible reason for the Forestry Commission to link to the Department of Justice, for example, and the same is true of innumerable other pairs of government bodies. In addition, the hugely distributional distribution of inlinks seen in Figures 7 and 8 means that there are certainly nodes that are more important (and more central) to the network than others. Rather, it is in a lack of *local* centrality – particular groups of nodes which are tightly clustered together – that the Infomap results found in this analysis are created.

The conclusion, therefore, must be that sites are not linking in accordance with any kind of sense of local value or particular user need. Otherwise, we would expect dense interconnectivity in groups of government agencies dealing with similar issues and hence similar customers. If there is a linking principle, it must be in linking by a global sense of importance: as in other studies of the web, prominent sites (like the large new gov.uk) get links, less prominent sites do not.

To get a fuller picture, it will be necessary to do more in-depth analysis. The next section deals with one aspect of that, assessing the importance of geographical proximity on clustering. For now, it is perhaps just necessary to note that Hypothesis H2 is unsupported by these data. That also suggests that attempts by political scientists to infer government structure from web crawl data may be problematic.

4.3.3 *Network communities: geography and linking patterns*

Geographic variations in link patterns potentially offer a shape of the interactional geography of local government. Having a dataset with implicit physical

geography (such as local authorities) makes this analysis straightforward. This section tests Hypothesis H4, predicting effects of geographic proximity on web linking behaviour.

To allow an analysis of the extent to which online linking behaviour reflects underlying organisational and geographical links, an adjacency matrix was constructed. Each authority pair was coded for the presence or absence of geographical adjacency, and similarly for overlap (reflecting the relationships between County and District councils in two-tier structures). Boundary data were extracted from the definitive Boundary-Line dataset current at the time of the crawl (Ordnance Survey, 2014) using the GRASS geographical information system (GRASS Development Team, 2015).

In order to compare the geographical network formed by council proximity and the web link network formed by councils' linking behaviour, it is necessary to import both sets of relations into comparable network structures.

An adjacency matrix is a two-dimensional matrix, coding the presence (1) or the absence (0) of adjacency between each pair of councils.

A pair of adjacency matrices was constructed using reflecting each possible council-to-council link in England, one with the geographical data and one with the ADM link data matched to the councils that control each website (see Section 4.2.3 for details of the ownership matching process). It was then possible to use standard statistical tools to look for associations between the two matrices, reflecting a greater or lesser propensity for websites to be linked based on geographical proximity¹⁵.

The data are shown in Figure 10, with the logical links from the web network plotted on top of a map showing geographical proximity¹⁶. It is clear,

15 Although the association being measured does not, in principle, show the direction of any causal relationship, it's clear in this case that changes in website configuration cannot affect the UK's geography.

16 The map is based on Ordnance Survey Boundary-Line and Ordnance Survey Northern Ireland Open Data Largescale Boundaries (Ordnance Survey, 2014; Ordnance Survey Northern Ireland, 2016, © Crown copyright and database right 2014–2016), using the boundaries current at the time of the web crawl. These data are available under the Open Government Licence.

from looking at the map, that a substantial proportion of the online links are geographically proximate.

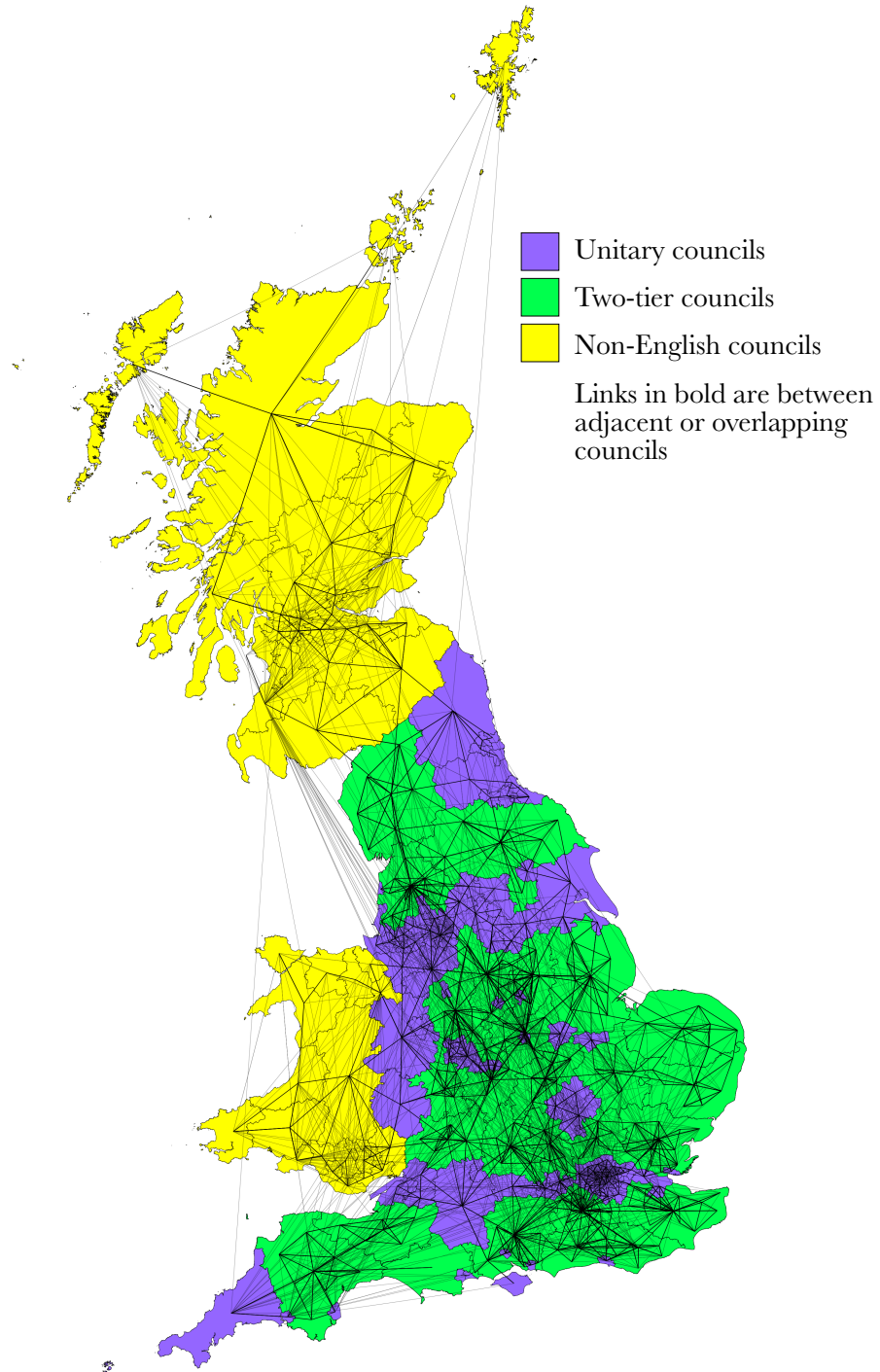


Figure 10: Mapping British councils and their web links

To conduct the analysis, a logistic regression model (Model 3, Table 10) was built, with hyperlink existence between authority pairs as the dependent

variable and geographical proximity (either adjacency or two-tier overlap) as the single independent variable. It is standard practice when conducting analyses of network statistics to also to conduct a diagnostic test of the result against constructed null graphs. This controls for the discovered structural properties of the network under study by permuting the labels and re-running the analysis on structurally-similar graphs but without the specific associations being tested for. In this case, the logit model was tested using the Quadratic Assignment Procedure (QAP). This is a Monte Carlo method comparing the test statistic¹⁷ with equivalent statistics calculated after the graph's labels are randomly permuted (Krackardt, 1987). This allows the estimation of the probability of finding a result at least as extreme as the found, assuming the null hypothesis. The results of the analysis are in Table 11.

Geographic proximity is associated with an increase in the log odds of being linked on the web of 4.77: neighbouring authorities are 117 times more likely to link to each other than non-neighbouring authorities. This single variable built into the model is enough to correctly guess 55% of the linking behaviour, which is impressive given the sparsity of links (less than 1% of the possible authority link-pairs were made).

The diagnostic test using QAP strongly supports this finding not being artefactual; the strongest result found fitting 1000 diagnostic null models was a log odds increase of only 0.43, compared to the 4.77 found in the true model, so we can be confident that the effect is not outweighed by measurement error due to the structural properties of the graph [$p = 0$].

This is very strong evidence for Hypothesis H4. That very few authorities were linked at all is also suggestive of the independence of authorities from each other and their presumed status as a heterogeneous and geographically-specific group.

17 In this case the coefficient of the geography term in the logit model

Network logit model of web linking propensity (3)		
	Estimate	e ^{Estimate}
Geographic proximity	4.764***	117.2***
Constant	-4.577***	0.010***
* p < .1; ** p < .05; *** p < .01		
Pseudo- R^2 measures:		
$\frac{Dn-Dr}{Dn-Dr+df_n}$: 0.555		
$\frac{Dn-Dr}{Dn}$: 0.901		
Contingency Table:		
Predicted	Actual	
	Not linked	Linked
Not linked	159,695	1,643
Linked	1,402	1,690
Fraction Correct: 0.9814815		
Fraction Predicted Links Correct: 0.5465718		
Fraction Predicted Non-links Correct: 0.9898164		

Table 10: Logit regression model: geography as determinant of linking patterns

Test diagnostics: distribution summary		
	(intercept)	Geography
Min	-4.576	-0.692
1stQ	-4.564	-0.117
Median	-4.561	0.005
Mean	-4.561	-0.008
3rdQ	-4.557	0.100
Max	-4.542	0.437
Null hypothesis: QAP ; 1000 replications		

Table 11: Logit regression model: geography diagnostics

4.3.4 *Content analysis of connections*

“As with web impact evaluations, a content analysis of a random sample of the links is highly desirable in order to be able to interpret the link counts.” (Thelwall, 2013, p.30). Not all links between pages mean the same. In order to effectively interpret the reality underlying the web structure found, we need to be able to differentiate links created for different purposes.

The automated techniques used in the categories above treat links as equivalent (equal in value and presumably of similar purpose) while in practice there are known to be a variety of motivations for linking, including signposting valuable content, reflecting the structures and preferences of the linking organisation, and wholly functional links intended for the delivery of the site’s own content rather than reflecting any kind of extra-organisational link.

To come to a better understanding of link motivation, 297 links were randomly selected and hand classified. The sampling frame was every link in the dataset, excluding links within domains: that is, any link from a .gov.uk domain to a different domain (inside or outside .gov.uk)¹⁸. The classification structure was developed iteratively based upon a preliminary analysis on a separate sample, with the categories representing the most common linking reasons:

Internal links	Cross-domain variants of internal links. These are largely artefactual, as it comprises links that are actually internal to an organisation but which are between separate third-level domains
External web provider	Links to external organisations (in this data, exclusively private ones) for the purposes of the

¹⁸ As the aim is to comparatively assess the meaning of English local government web links, links which are neither from nor to English councils were included. The rationale behind excluding within-domain links was to capture the motivations for the linking of the overall network, rather than just the contents of domains’ own navigation bars.

website itself. A large proportion of these are to the source organisation's own content hosted on Flickr, YouTube, Twitter, and Facebook; smaller numbers are to external HTML validators, Adobe as hosts of Flash and Acrobat Reader, and to various automated accessibility and translation services.

Information signposting	Links which signpost visitors to information provided by another organisation. This was required to be specific: a link to a large organisation's front page from a sidebar is probably not signposting.
Service signposting	Similar to information signposting, but signposting visitors to services delivered (either online or offline) by another organisation. This was similarly required to be specific.
Gratuitous	This category is drawn from the webometrics literature. They reflect either a structural relationship or some kind of support for the target organisation's goals. These are normally sidebar, footer, or A–Z directory links to geographically close organisations, but without a particular service in mind ¹⁹ .

The breakdown of link types for this analysis, broken down by whether the source or target of the link was an English local authority, is given in Table 12.

The first finding of interest is that not one of the 297 links analysed was to an English local government body without also being from an English local

19 So Bolsover's Trading Standards sidebar link to Derbyshire County Council is some kind of service signposting, while Ryedale District Council's listing of all neighbouring authorities in its directory (when its citizens don't get services from any of them) is gratuitous.

government body. That is, the number of unreciprocated outlinks from local government (122) is much larger than the number of unreciprocated inlinks to it (0).

Approximately half the sample was links that did not involve English local government at all. The two largest blocks of these turn out to be internal links between different parts of the Northern Irish government (many sites are branded NIdirect and have a common cross-domain sidebar) and between parts of the Government Digital Service (GDS) Govstore (which is prominent mainly because it's so large and thus is sampled frequently).

The majority of the remaining links, both from local government and other domains, are external web provider links to external commercial organisations. Only 10% of the links in the analysis (31/297) are signposting, with 80% of these (25/31) being informational rather than service-focused links.

	Neither	To	From	Both	TOTAL
External web provider	63	0	76	0	139
Gratuitous	10	0	24	6	40
Information signposting	16	0	9	0	25
Internal cross-domain links	76	0	9	1	86
Service signposting	2	0	4	1	7
TOTAL	167	0	122	8	297

Table 12: Manual classifications of web links by link to/from English councils

A number of local authorities, for example, have links to gov.uk or its predecessor DirectGov in their sidebar or footer. On the one hand, this shows an admirable desire to signpost citizens in the direction of a major provider of online central government services. On the other, the links carry no indication of what gov.uk is, which services it provides, or why the citizen may want to visit. As a means of providing holistic online government service, it's useless: if people know what gov.uk is and why they want to visit, they can go there directly by typing six characters into their browser address bar.

The external web provider links have nothing to say about how the structure of government is reflected online, except perhaps to observe that both online

and offline government bodies have always extensively used external private organisations as contractors and as sources of goods.

The principal conclusion from this analysis, therefore, is that very few of the links to other government organisations are to specific service or information pages which have been selected as relevant to the visitor to a specific source page. The volume of links and therefore the indegree and clustering identified above is dominated by vague links at organisational level. This supports Hypothesis H5, that links primarily reflect structure rather than the signposting of services. This challenges the assumption that the DEG recommendation of needs-based holism is visible in practice on the local government web.

There are two features of the analysis which reflect on the generalisability of this finding.

Firstly, as the sample denominator is links, links from the navigation sidebars of sites links dominate. This is reasonable, given their much greater prominence, but means that the effects of service delivery links tightly tied to customer need (as might be suggested by DEG) are less prominent than they would be for an analysis featuring only links within the unique content of each page. The latter is difficult to collect (there is no automated way of stripping navigational content from arbitrary sites) but would be a useful adjunct if possible.

Secondly, the sample size for the analysis is fairly small ($n = 297$). Nevertheless, the large magnitude of the category variations in Table 12 strongly suggests that the effect size here is substantial rather than random.

4.3.5 *Outliers*

This section, also a recommendation for WNTD analysis, looks qualitatively at the outliers, clusters, and important nodes identified in the sections above. Its purpose is to explore in concrete terms the context of the earlier findings, and to identify reasons for outlier behaviour where possible.

The central node in the network in terms of indegree centrality (and one of the most important nodes in terms of size) is the centralised government website run by GDS, www.gov.uk. This centrality is not surprising; it has been steadily absorbing the service and public-focused information from other government websites for several years (and has recently been expanding into the provision of information from departmental sites intended for internal stakeholders²⁰).

Many of the most significant outliers in this analysis are an artefact of the nature of the DNS, rather than a consequence of the structure of government. Looking again at Figure 8, for example, the improbably well-linked but small domains in the top left corner of the graph. Those with indegree > 50 and page count ≤ 10 are charity-commission.gov.uk, hmso.gov.uk, opsi.gov.uk, legalservices.gov.uk, dfes.gov.uk, info4local.gov.uk, dca.gov.uk, hm-treasury.gov.uk, and thepensionservice.gov.uk. These are all alternate domains hosted by central government bodies which still have inlinks but for which the content is now hosted elsewhere. hmso.gov.uk and opsi.gov.uk, for example, used to hold the UK Statute Law Database, but this has now been moved to legislation.gov.uk.

Similarly, the site in the bottom-right corner of Figure 7 is a copy of a site at a different URL, rather than an independent host; brixhamtowncouncil.gov.uk appears to also contain the whole content of Torbay Borough Council's website without having been detected as such and excluded.

Much smaller issues arise from the clustering. As there are so few substantive clusters, the only unusual object is the "Core" clustering which consists of more than half the objects. This is certainly unusual for an Infomap clustering, but it's difficult to see why there is little linkage between the nodes sufficient to form subclusters if not for a case-by-case analysis of each link²¹

20 Intriguingly, another of the domains with high indegree centrality is direct.gov.uk, www.gov.uk's immediate predecessor, which was still linked to extensively from across the public sector even though it was no longer active at the time of the crawl.

21 Which is clearly impossible for a network this large

In network terms, the outliers in the analysis are those nodes which are isolates: those entirely disconnected from the rest of the network. These are parts of .gov.uk which are neither linked to by any other .gov.uk site nor link out to any other .gov.uk site. There should be very few isolates in this dataset, though the data collection was carried out in such a way as to pick up those that exist²².

This is principally because each government body oversees, is overseen by, works with, or collects money or authority from, at least one other government body. As the UK is a unitary state, all governmental power ultimately derives from Parliament or the Crown; if linkages in the government web reflect linkages in government, the network graph should be a smoothly connected whole.

Only 81 of the 1902 .gov.uk domains for which web pages were collected are isolate; they are listed in Table 30 in Appendix A. The large majority (75 of 81; 93%) of these are parish councils. None are English local government domains, so Hypothesis H3 is entirely confirmed in respect of local government and almost fully confirmed for all other significant government bodies.

4.4 DISCUSSION

4.4.1 *What explains the structure?*

What structure there is within the .gov.uk network is primarily geographic. The findings in Section 4.3.2 that the devolved governments of Scotland, Wales, and Northern Ireland had consistent Infomap clusters support this, as

22 One problem with sampling isolates on the live web is that, by definition, they are not linked to other nodes in such a way that they can be picked up in a web crawl. For this research, every registered .gov.uk domain was used as a crawl seed; this will pick up all the sites on the seed list (whether isolates or not) leaving only those domains registered after the list was released but before the crawl as being vulnerable to being excluded as a result of isolation.

do the weaker county-level clusters and the findings of the local government adjacency and overlap analysis in Section 4.3.3.

There is less evidence of extensive structural connections, much less of a pervasive attitude of needs-based holism in governmental online offerings, as is recommended by DEG. The results of Section 4.3.4 suggest that the motivations to the construction of the local government web are as complex as the motivations for any other aspect of government behaviour; there is not the sense, looking at the data, that agencies are being particularly helpful in meeting needs of visitors outside the narrow organisational functions that they have.

More broadly, the answer to “what explains the structure” is at least partially “there is little structure to explain”. The UK government web is not neatly partitioned into a number of tightly-integrated cross-organisational service delivery clusters, nor does it significantly reflect the offline indicia of power such as budget or staffing levels.

By contrast, the government web stands to be interpreted in a way parallel to that of the government it represents: it is *by* parts of the government, but does not *reflect* them in a structural way.

The most obvious way of examining this is to look within local authorities for one of the most significant, politically sensitive, and expensive services: adults’ and children’s social care. This takes up 25% of local government service spending²³, but the amount of content relating to it is comparatively very small.

This reflects the fundamentally offline, face-to-face, and labour-intensive nature of social services. Child protection is not primarily a large-volume transaction, to be delivered online to reduce staff contact costs. Nor is it primarily an information service, which can be effectively delivered by the construction of large volumes of online content. Rather it is bespoke, professionalised, in-

²³ £22.6bn of £91.2bn total in 2014-15 (see DCLG, 2016, Table 2)

tensely bound to the locality of the children being protected, and (perhaps most of all) strictly confidential. Likewise, adult social care is largely daily living support, delivered face-to-face by specialists and care workers. Even if the largely elderly recipients were great users of the Internet, very little of it is electronically deliverable.

On this basis, it is understandable that the structure of the government web tells us much less than we'd like about the structure of government: it does, by its very nature, only show those parts of government service which benefit from being put online. The social services pages of most top-tier authorities have a decent amount of guidance, plenty of opportunities to contact a professional, but essentially no service delivery opportunities.

Replicate this finding across all the other areas covered by local government which are rooted in place, including administration of schools, cleaning of streets, and provision of public toilets, and a picture emerges of an iceberg organisation online: although all parts of the organisation may be visible through the medium of the web, little of it is natively digital and the potential gains from informatisation look somewhat more subdued than the rosy picture presented in the e-government literature.

4.5 CONCLUSIONS

In order to answer research sub-question 1, a new full-population web crawl of British government online was conducted. This has provided a unique primary dataset to examine at first hand the revealed size, scope, and shape of online government. By analysing this data using the tools of webometrics, which are uncommon in political science and public administration, new insights have been made into local government's activities online and how that fits into the DEG quasi-paradigm.

A key finding is that the structure and shape of English local government on the Internet are not a close reflection of the structure and shape of councils

offline, with significant variation in online shape even between councils of a similar type and size.

The relationships between the different bodies online are remarkably lacking in expected structure. There is a weak geographical association visible both in cluster analysis and in the tendency of neighbouring authorities to link to each other more readily than can be explained by chance. Little evidence of topic-based linking is visible; evidence for the kind of needs-based holism predicted and advocated for by DEG is largely absent.

There are few parts of .gov.uk which are truly isolate, so the UK government web does indeed form a single web rather than a set of unconnected islands of connectivity. The poorly-connected network outliers that there are are identified are either very small and local (e.g. parish councils) or secondary domains of larger bodies which are either obsolete or poorly connected to much better-connected main websites.

Given that local government offline and online have different content, different strengths, and different service focuses, this chapter concludes that the measurement of government web content only gives us partial information about the overall structure of government in the digital age, and inferences from online to offline activities should be treated with caution.

4.5.1 *Limitations and further research*

There are a number of issues with the collected dataset which remain after all the work described above. A good number of unwanted pages will have ended up in the crawl sample, principally as a result of web server misconfiguration; a number of .gov.uk websites wrongly return 200 ‘success’ HTTP codes when in fact an error has occurred. A huge number of these have been removed as a result of the careful filtering of the dataset, but short of accessing every URL by hand, it’s impossible to guarantee that they have all disappeared.

Conceptually, this is not problematic. All quantitative political science has to deal with noisy data collection processes. The key question is whether the dataset collected is a good representation of the underlying subject of interest. The dataset collected for this research not only meets the collection standards set by the experts in webometrics, but is a good deal more complete than any previous public administration web research. Despite the inevitable gaps, coverage is broad and sufficient to answer the questions set.

The more fundamental issue that arises from this chapter is that the scope of government online does not appear to reflect, in large measure, the scope of government offline. This is not problematic for this research: indeed, the issue forms a key part of research sub-question 1. It does, however, raise questions about the types of questions for which web data will be useful in the future.

The theoretical and methodological conclusions arising from the new methods employed in this research, and the issues that have arisen from the nature of the dataset, are discussed in Section 7.5.1.

It would be ideal to conduct similar research on historical web crawls and then to link that into time-series measures which reflect the theoretical understandings of DEG. These could then be fed into the modelling process in Chapter 6 assessing the links between online activity and service performance. Unfortunately, historical data of suitable quality aren't available. There is a sizeable historical dataset held by the British Library covering .gov.uk, snapshots of particular government websites held by the National Archives²⁴, and a further collection collated by Common Crawl as part of their general web archive collection²⁵. Unfortunately, collection criteria for each of these sets have varied over the years and there are substantial coverage holes. Webometric analysis depends critically on getting at least a reasonable sample of the site under study, which is one reason for the care taken with collection.

24 Browsable through a Wayback Machine installation at <http://webarchive.nationalarchives.gov.uk>

25 Downloadable from <http://commoncrawl.org>

General purpose web archives, which have been collected on an ad-hoc basis to preserve content rather than to preserve webometric measures, do not lend themselves to this kind of retrospective analysis, and the .gov.uk collections identified above are not of sufficient coverage to allow confident analysis.

The webometric analysis conducted in this chapter only goes part of the way towards understanding government activities online. Just as important as analysing the structure of government as it is revealed online is to look at the content. Chapter 5 extends the findings of this chapter by looking at the content of government websites, rather than the linkages between them.

LOCAL GOVERNMENT USE OF THE INTERNET: CONTENT AND CONTEXT

What is the content of the English local government web? What can this tell us about local government and its approach to e-government?

5.1 INTRODUCTION

This chapter extends the analysis of Chapter 4 to answer the second part of research question 1 by focusing on the scope and nature of the web content published by local authorities. The dataset drawn upon is that described in Section 3.4.2, with the approaches developed in Section 3.5 used to get a grip on the large quantity of textual content data that the crawl provides.

The key understanding to be elicited from the data is the extent to which web content can illuminate local government and how it relates to the local government web. Is the local government web homogeneous, with web site sizes reflecting the size of the authority and the scope of service-related coverage stable between authorities and roughly proportional to the relative sizes of those services? Or do local government websites reflect the differing local priorities of each council by being unique and distinct, not at all suitable for merger into a central local.gov.uk domain (see Section 2.4.8)?

The approach taken in this chapter is strongly based upon the machine learning techniques discussed and developed during Chapter 3. These meth-

ods allow the analysis of large datasets at full-population level, which is an opportunity which political scientists have rarely had without huge difficulty and expense. It is a suitable approach here because the underlying question being answered (research sub-question 2, What is the content of the English local government web? What can this tell us about local government and its approach to e-government?) requires a broad look across the whole field of web provision by local governments.

The web itself provides a substantial new body of data. Automated classification tools provide a rich new set of methods to extract meaning from this data. In this way, the state of our understanding of local government online behaviour is not just taken forward by the development of new insights, but also by the testing and development of the tools needed for future work.

5.1.1 *Hypotheses*

The key contribution from analysing the content data is not an understanding of local government relationally with other organisations (which arises from the linkage analyses in Chapter 4), but of the uses within each authority. Critically, our understanding of local government and the possibilities for its future online development are informed by an understanding of the homogeneity or heterogeneity of the authorities and the online content they publish (see Section 2.4.8).

As a result, the hypotheses for this chapter focus on three different aspects of online heterogeneity, measured using two different approaches.

Hypothesis H6 mirrors Hypothesis H1 in Chapter 4. It assesses whether the online presence of each authority is a factor of its size

Hypothesis H6: Local authorities' extent of online activity, measured by the number of web pages, is substantially explained by council type and population.

Hypotheses H7 and H8 assess heterogeneity using two of the key yardsticks proposed by practitioners and academics – the extent to which the web is being used for the delivery of services, and the extent to which those services are transactional in nature – that is, delivering the service online rather than providing information about or supporting the delivery of an offline service. The suggestion is that, while transactional uses will be widespread in some areas, they will be little implemented in others. To that extent, local governments and their use of the web can be reasonably described as heterogeneous.

Hypothesis H7: Local authorities' online presence is heterogeneous, measured by the proportion of each council's pages classified as service-focused.

Hypothesis H8: Local authorities' online presence is heterogeneous, measured by the proportion of each council's pages classified as transactional.

Hypothesis H9 proposes heterogeneity at the level of content topic. Based upon a novel application of automated topic modelling to local government web activity (explained in Section 5.2.2), this hypothesis suggests that there will be a wide variation in the content published by each local authority.

Hypothesis H9: Local governments' online presence is heterogeneous, measured by the proportion of each council's pages in each topic.

5.2 PROCESS

The raw data for this chapter is the web crawl described in Section 3.4.2. Content in proprietary document formats was processed through Warctika (see Section 3.6.1) as part of that process and saved as plain text in the crawl

Web Archives (WARCS). The page content was subsequently extracted from the crawl WARCS and loaded into a document database for easy access. As part of the loading process, Hypertext Markup Language (HTML) format content was also converted to plain text using the Beautiful Soup HTML parsing library (Richardson, 2015). The full text of the large majority of pages was therefore available. The dataset was then available for further analysis.

5.2.1 *Classification of government web content: service and non-service content*

A sample of 449 randomly-selected pages were hand-classified by the researcher. Pages in Welsh were excluded¹. Hand-classifications were made according to the following categories:

A. Service focus:

- Service-focused: externally-focused documents, intended to inform service users or support service delivery ($n = 175$); or
- Non-service-focused: internally-focused documents released for transparency of internal decision-making or to document regulatory actions such as planning control and licencing ($n = 274$).

B. Informational/ transactional:

- Informational: documents principally giving information, including pointers to offline services ($n = 435$); or

¹ Pages in multiple languages make language-based classification techniques much harder. The pages in the sample were automatically classified according to language using the *langid* software (Lui and Baldwin, 2012). This uses similar machine learning techniques to those described in Section 3.5, but highly specialised by the careful selection of features to differentiate the primary language of given texts. In this case, the software was configured to only select between English and Welsh for the languages detected; focusing on only two languages improves computational speed and classification reliability. There is a small amount of Scottish Gaelic in the dataset, but far less than Welsh on account of the tiny number of native Gaelic speakers in Scotland compared to Welsh speakers in Wales. There are also tiny numbers of pages in other languages, including Polish, Urdu, and Farsi, which are largely either overseas-focused content from the Foreign Office, or pages on council websites directing non-English speakers to translation and interpretation facilities.

- Transactional: documents principally aiming to deliver a transactional service online ($n = 14$).

The service focus classification is important because it is relevant for measurement of volumes of information to track that which is not actually following the “services online” rhetoric.

The second dimension (Informational / transactional) relates to the deeper question of whether the services in question are set up to allow transactions to be completed online, or just to provide information to allow offline service delivery.

The scikit-learn library (Pedregosa et al., 2011) was used to extract statistical features from the selected documents and then to train a linear Support Vector Machine (svm) on the extracted features to differentiate service-focused and democratic-focused pages. The features were based on four sources: the full text of each selected page (using a unigram bag-of-words approach), the length of the page in words, the directory and filenames contained within the Uniform Resource Locator (url), and finally the suffix of the filename (such as .doc, .html, .pdf) if present. The full extraction and classification pipeline is shown in Figure 11.

Subsequently, a further sample of 201 randomly-selected non-Welsh pages was hand-classified by the researcher according to the same categories and used to validate the performance of the classifier against out-of-sample data. The raw matrix of validation results is in Table 13, with the resulting metrics in Table 14.

		Machine		
		Service	Non-service	TOTAL
Human	Service	63	16	79
	Non-service	15	107	122
TOTAL		78	123	201

Table 13: Automated service classification: validation confusion matrix

Metric	Score
Precision	0.85
Recall	0.85
F_1	0.85
Krippendorff's α	0.68

Table 14: Automated service classification: validation results

As can be seen from the first table, the large majority of classifications by machine and human were concordant (shown in the top-left and bottom-right cells of the confusion matrix). In the second table, the precision and recall statistics are standard measures of the effectiveness of the classifier at correctly classifying *only* the members of each class (precision) and *all* of the members of each class (recall). F_1 is the harmonic mean of the two, reflecting overall performance; as the precision and recall statistics for this validation are equal at 0.85 (a respectable performance on real-world data) F_1 is equal also.

A significant issue when dealing with machine learning approaches in disciplines such as political science where automated classification is uncommon is in assessing whether a particular classifier is “good enough” to infer real-world attributes in the data being classified. The machine learning researchers’ own measures (such as F_1) are intended, by contrast, for evaluating the performance of new techniques against standard datasets with known ground truths.

To deal with this issue, Krippendorff’s alpha (α) has been calculated. This is a standard measure of inter-coder reliability used during qualitative content analysis by political scientists, and measures the reliability of judgements by different coders. In this case, the human researcher has been treated as one coder and the machine as a second. The calculated α is 0.68, which is conventionally considered to show enough inter-classifier reliability to be able to draw tentative conclusions from the data² (Krippendorff, 2004).

² In this case, there is a tradeoff between completeness and reliability in classification, resolved in favour of having full-population coverage. See Section 3.5 for a full discussion.

Once the model was constructed, each remaining document in the corpus of web pages was classified by the model. Due to the choice of a linear svm, this was a computationally inexpensive operation: single-core runtime for the whole process (including database reading and updating) was only 50ms/page. At only a few days to classify all 11 million items in the dataset, and no special computational hardware required, this performance puts the technique well within reach of any political scientist with an interest in assigning attributes to a full-population dataset³.

5.2.2 *Topic modelling of the .gov.uk corpus*

In Section 3.6.1 a hierarchical clustering of .gov.uk was developed based upon the link structure embedded within the network. This used Infomap, which is an unsupervised algorithm: that is, it relies entirely on information built into the object being analysed, rather than on training data provided by the researcher. A similar approach can be taken with the textual content of pages. While the supervised svm classifier described in Section 5.2.1 above was based upon a body of training data (and thereby on a classification scheme developed by the researcher), this section describes an unsupervised approach known as topic modelling. This allows an analysis of the topics covered by different parts of the .gov.uk web without requiring either the researcher to have a comprehensive content schema in advance, or the huge amount of manually-classified content which would be required to train a supervised model for a large number of topics.

In a probabilistic topic model, textual documents are partitioned into multiple categories (notionally by topic). It is a corpus-based approach, relying on the availability of textual data. In contrast to the supervised techniques,

³ A linear svm is suitable for bag-of-words based textual approaches and saves computational power. Nevertheless, the use of parallel processing on multicore hardware would keep the runtime of even complex models for this kind of data to an order of weeks rather than months.

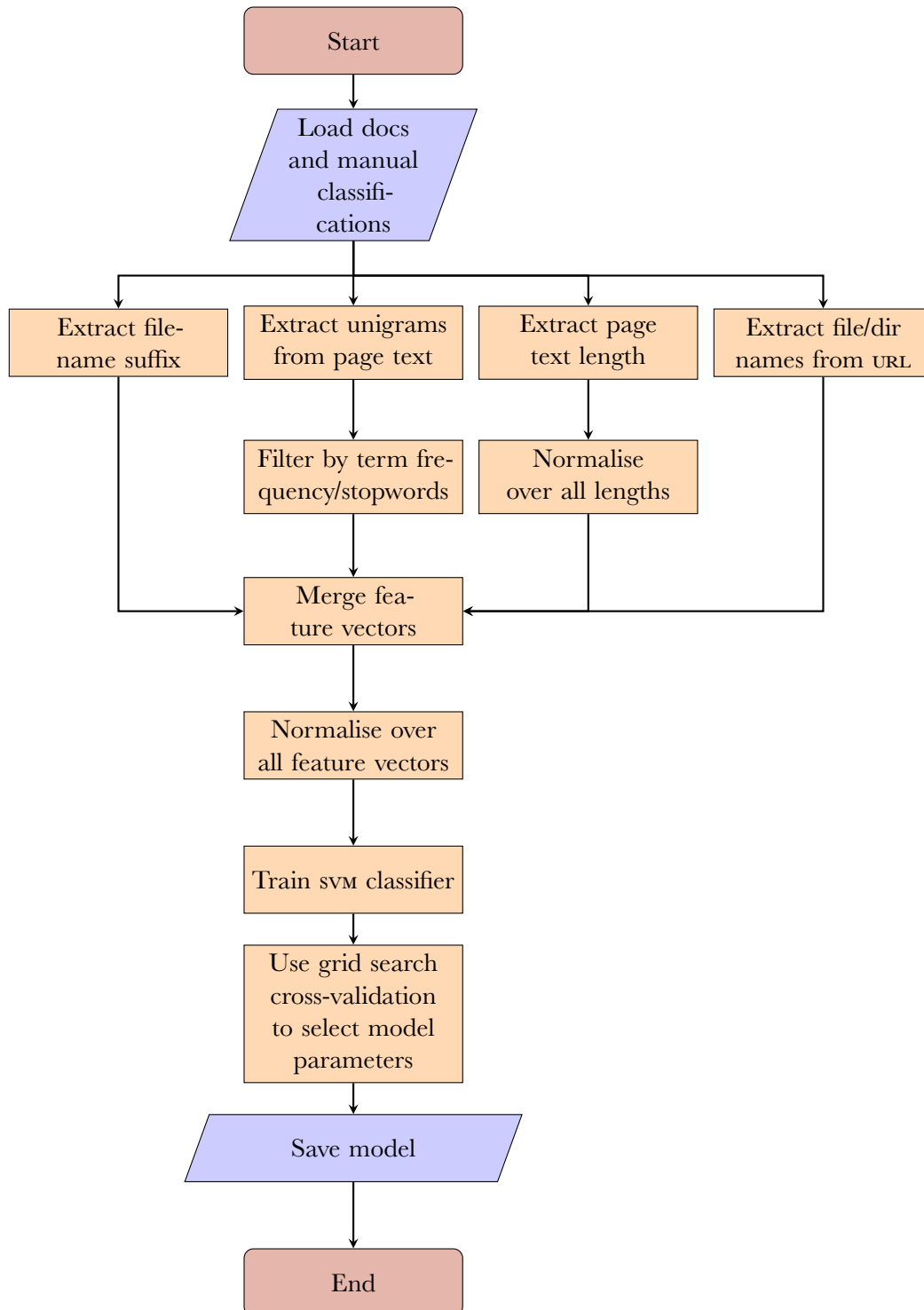


Figure 11: Supervised classification: feature extraction and training process

it extracts information only from the texts processed. The intention is to decompose a very large number of input features into a smaller number of structurally-linked output topics. This is an analogous process to Principal Components Analysis which is common in numerically-focused quantitative political science. The common Latent Dirichlet Allocation (LDA) approach builds a Bayesian model based upon feature presence (features typically being unigrams of the kind used as one of the inputs to the supervised process – see Figure 11) and then uses the posterior distribution to reveal latent semantic structure by iteratively adjusting the topic balance in each document (Hoffman, Bach and Blei, 2010).

An LDA topic model was fitted to unigram features extracted from the full text of one million randomly-selected web pages from the study using scikit-learn (Pedregosa et al., 2011), using the process described in Figure 12. This is a much simpler model specification than that used for the SVM models described above; the lower number of features extracted from each document is compensated for by the much larger volume of input data available (as the input data does not need to be manually coded)⁴.

One issue with the LDA is that it requires the prior selection of the number of topics to be modelled. There is no theoretical reason to prefer any one number of topics over any other for this dataset, as the scope of different areas of government is extremely wide and has fuzzy boundaries. 10 topics were selected, in the expectation that the resulting model would be sufficiently fine-grained to support nuanced analysis but small enough not to be overwhelming.

⁴ In fact, the much larger number of features possible using such a large volume of output data causes practical problems for the analysis: feature extraction generally requires all possible features to be held in memory for comparison with each new document. This analysis uses the so-called “hashing trick” (Weinberger et al., 2009), where the content of each feature is processed with a one-way hashing function before use. The fixed relationship between a feature and its hash value allows the features to be coded on their hash value, with the (longer) original features discarded to reduce memory overhead.

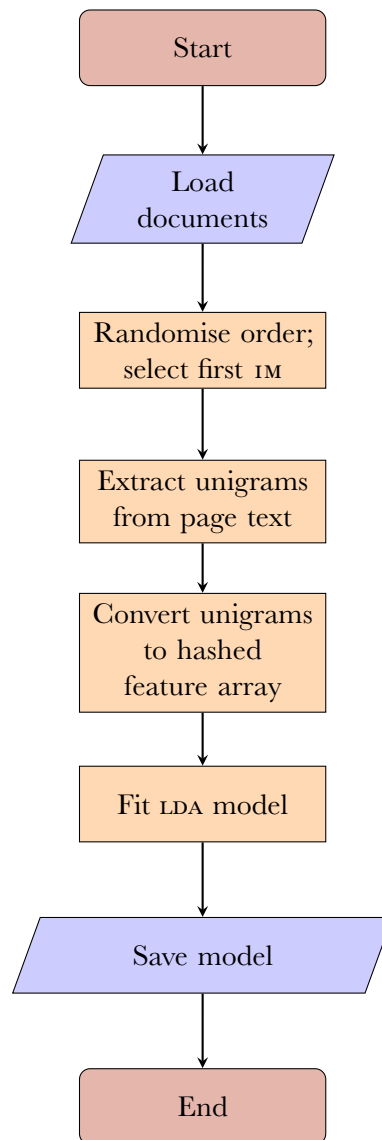


Figure 12: LDA: feature extraction and fitting process

In order to understand the nature of the topics, keywords were extracted by topic from a five hundred thousand page subset, using the tf-idf metric⁵ Figure 13 shows the process.

⁵ This simply calculates the relative frequency of words in a section of the corpus by counting the frequency of each term in the target document(s) and dividing the frequency score by the number of documents in the corpus in which the document appears.

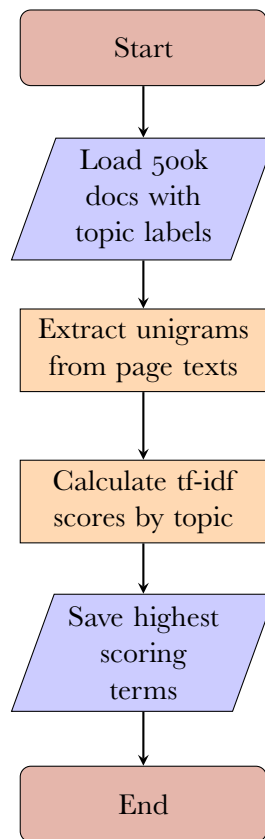


Figure 13: LDA: keyword extraction

5.3 FINDINGS

The government web is not now, if it has ever been, a simple structure of carefully-curated web pages. The scope and volume of content available for download is enough to intimidate the most intrepid of researchers.

The analyses in this section are grouped by local authority, so the unit of analysis is an individual English council, with each of the domains owned by each authority taken as a group.

5.3.1 *Number of pages*

This section tests Hypothesis H6, which suggests that the number of pages published by each council is explained by the population and type of each authority.

Local authority websites are large, by the standards of other organisations in .gov.uk, with top-tier councils particularly large. Figure 14 shows the distribution by domain, with the median for local authorities being well above the median for all other groups and only a handful of central government domains (such as gov.uk and scotland.gov.uk) being larger than the largest local authority sites.

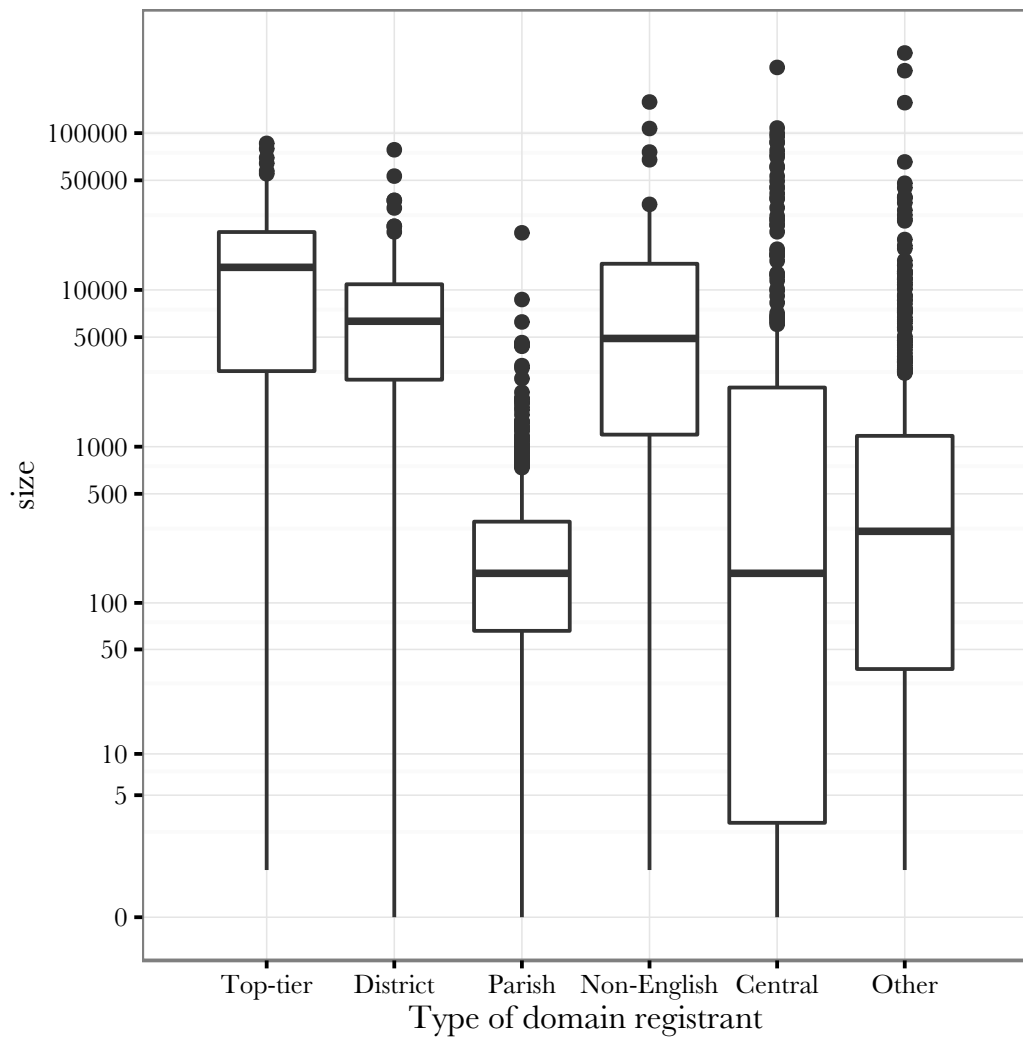


Figure 14: Website size by type of domain registrant

There is an association between the number of people a council serves and the combined size of all the council's websites. Figure 15 graphically shows

the distribution⁶. Although there is a positive association between size and population, it is not especially strong (Pearson's $\rho = 0.413$).

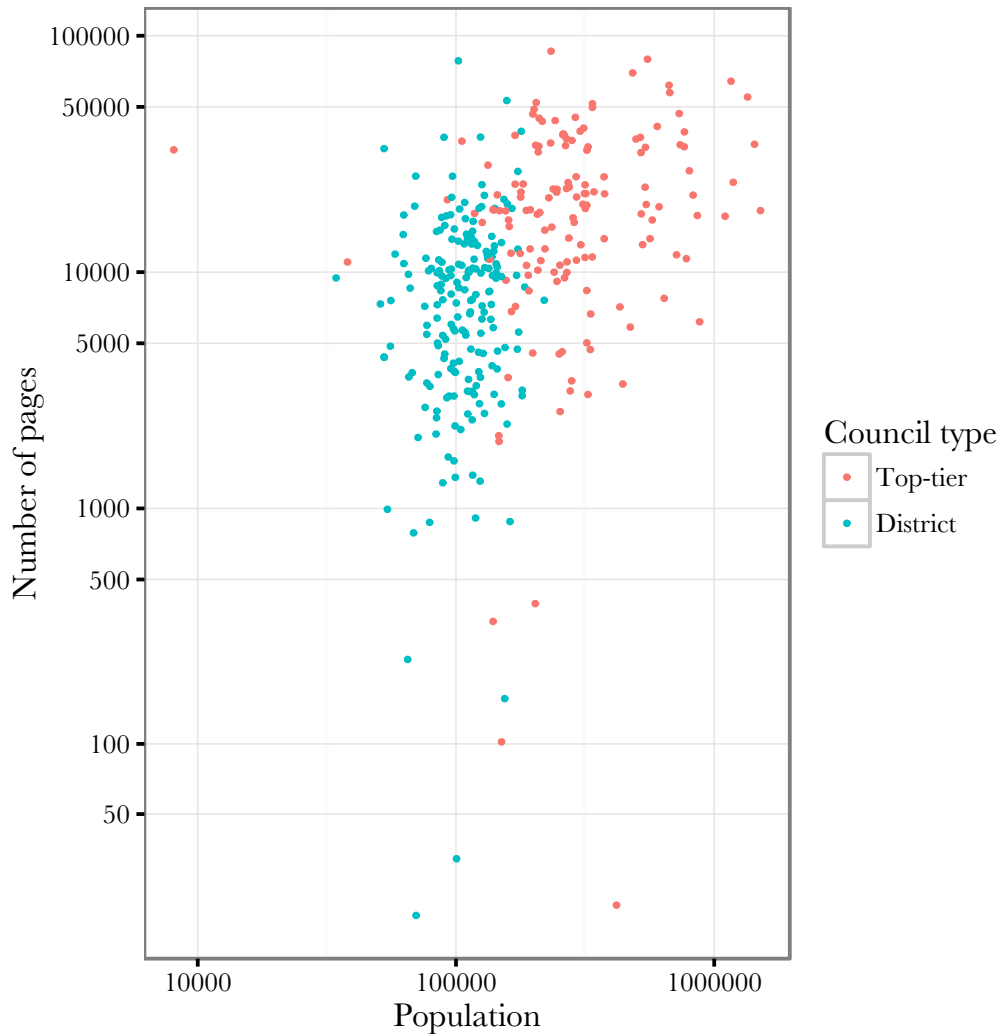


Figure 15: Total English council web site size by population

The number of pages hosted by a given local authority does not, contrary to the findings of (Holmberg, 2009), have a clean relationship with the population of the authority concerned. This is contrary to the assumption in Hypothesis H6.

⁶ Note that the outliers at the lower end of the y axis represent authorities where the main web domain has been not been fully captured (see below). The smallest regular council website group has at least 800 pages.

A formal test of the hypothesis was conducted by building an Ordinary Least Squares (OLS) regression model, with page numbers as the dependent variable and the population and type of each council as independent variables. This is Model 4, shown in Table 15.

	Number of web pages (4)
County council	17,133.370*** (4,255.287)
Unitary council	9,901.544*** (1,655.642)
Population	0.007 (0.006)
Constant	8,829.887*** (1,066.457)
N	346
R ²	0.250
Adjusted R ²	0.244
Residual Std. Error	12,340.230
F Statistic	38.043***

* $p < .1$; ** $p < .05$; *** $p < .01$

Base category is "District council".

Table 15: Regression model: number of web pages by council type and population

The association of size and type to the volume of web pages published is weak; the adjusted R^2 for the model shows that only 25% of the variation is explained by population and council type. The population term in particular is negligible, despite the visualisation in Figure 15: each additional resident is associated with only 0.007 extra pages. The dummy variables for the type of council have much larger effects, though, as there is a difference of over 17,000

pages between the base category (district councils) and the county councils which are the largest group in the model⁷.

On investigation, most of the additional variation seems to be explained by three non-population factors:

Firstly, some councils push much larger quantities of content online than others. This is particularly easy to spot in cases where a particular class of document is hosted on a separate subdomain. A large number of local authorities are using the software ModernGov as a committee management and democratic support system, with online access normally available to the archive of agendas, papers, and minutes via a web front end on a separate subdomain. The volume of these documents in terms of pages posted is vast: in most cases, the ModernGov subdomain is larger than the council's website with thousands to tens of thousands of documents available. However, this only weakly reflects structural differences between the authorities: the principal authorities have similar governance arrangements as a result of Labour's forced move to a cabinet system from 2000 (J. Stewart, 2003, p.61). Similar structures of cabinets coupled with scrutiny committees and separate regulatory committees (such as licensing and planning) apply everywhere. The main difference between the authorities is simply how long they've been running ModernGov and therefore how much of their archive is contained online. Similarly, some authorities run a directory containing local businesses and voluntary organisations (or even, in a few cases, an entry for every road in the council's area). This adds substantial page volume.

Secondly, decisions about how to structure the website make a disproportionate difference to the number of pages found. For example, some authorities have sites which are "responsive" – that is, they are designed to work on both mobile and desktop systems. Others do not. Hillingdon London Borough Council, for example, has a separate mobile site with more than 6,000 pages

⁷ The Council of the Isles of Scilly was excluded from this model, as it is a unique type of council with a uniquely small population.

(principally duplicating content from the main site and likely fed from a joint content management system). Similarly, the decision to split parts of a site into multiple small pages multiplies the page count without necessarily increasing the number of words; spending.hillingdon.gov.uk alone is several times larger than most local authority web presences by simply publishing every council invoice as a separate web page rather than as a single spreadsheet⁸. This variation is so large that it seems to outweigh the actual differences between authorities.

Thirdly, some domains are restricted by crawl limitations as discussed in Section 3.4.4. Norfolk County Council's main website at www.norfolk.gov.uk, for example, has the following rule embedded in its robots.txt file, which prohibits all non-whitelisted crawlers from visiting:

```
User-agent: *
Disallow: /
```

No pages are present in the dataset for Norfolk, therefore.

After being immersed in the dataset for an extended period of time, it's possible to give an impressionistic view of what the bulk of the web content is. The significant bulk appears to be semi-automatically generated database content of various kinds. Prominent blocks of pages are filled with the minutes of meetings and calendars of events (by day, by week, and by month)⁹. Other large content blocks include special purpose content databases: forestry.gov.uk, for example, contains information about each of the major forests run by the Forestry Commission in a more-or-less standard format. This isn't auto-

⁸ Local authorities have been obliged to publish details of every item of spending over £500 since January 2011 (Local Data Panel, 2010).

⁹ In fact, an essentially infinite supply of near-empty events calendars can be generated from many sites simply by repeatedly clicking the virtual "last week" button. This was prevented for this research by putting limits on the times for which the calendar would be collected. Refusing to collect it at all would have prevented a major source of outlinks from being gathered, but collecting pages from before the Internet existed (or, indeed, separately collecting the daily and weekly calendars) adds volume without adding content and gives a false impression of the contents of the web.

generated content (all of it will have been produced by staff) but it adds up to a great deal of information in aggregate.

Consequently, we must reject Hypothesis H6; variations in the volume of content put online by councils are not largely explained by population and council type. This extra heterogeneity seems to be significantly explained by non-structural factors to do with variations in online presentation, however, rather than reflecting underlying heterogeneity in council attitudes to the web or service provision.

5.3.2 *Page classifications*

This section aims to test Hypothesis H7, that the proportions of service-focused pages are heterogeneous between local authorities.

In analysing the data to answer this question, an unexpected finding becomes apparent: not only do the proportions of service-related content vary between authorities, but they are on average surprisingly low.

In fact, the majority of content in the local government web is not specifically customer-focused, but is ‘back office’ paperwork put up for internal, regulatory, or democratic purposes. It is difficult to measure the relative volumes exactly, as so much of the minutes and planning documentation in particular is tied up in huge databases which we have not attempted to crawl. Nevertheless, the breakdown of content classified as internally-focused from the .gov.uk crawl is given in Table 16. This can be considered a lower bound for the actual proportion of content.

Immediately this raises questions about the models of service transformation through online transactions presented in the literature. If the value of the Internet to local authorities is in replacing expensive face-to-face transactions with efficient and customer-friendly online ones, why is the large bulk of the content of local authority websites made up of resolutely non-interactive strategic plans, minutes, and building control documents? The revealed preference

class	Service	Non-service	Excluded	N
Overseas government	0.72	0.25	0.03	1,479
Joint local government bodies	0.67	0.31	0.02	481,375
Devolved governments	0.60	0.35	0.05	1,166,845
Central government	0.60	0.38	0.02	1,951,446
Miscellaneous local bodies	0.55	0.41	0.04	220,299
Top-tier councils	0.36	0.61	0.04	4,501,398
District councils	0.33	0.67	0.00	1,965,040
Parish councils	0.32	0.67	0.01	315,986

Table 16: Proportion of service and non-service pages by domain type

of the local authorities appears to be the presentation of *informational* content rather than transactional. This is in contrast to the central government and especially devolved government bodies, for whom over 50% is service-focused content.

In many ways, this makes sense. In addition to requiring services from local government, citizens also need information. If the information is not available online, they will need to contact the council to find it out, which is also an expensive contact for the authority. A good example here is finding out the opening times of the council refuse disposal site: citizens may only go a few times a year to dispose of bulky waste, but opening hours often vary by times of daylight. Consequently, if the opening hours weren't on the website, this would generate a large number of phone calls to the council enquiry line. Such enquiries are likely to be substantially more onerous for local government, with its multiplicity of distinct public facing services, than it is for central government's preponderance of policy departments and single-purpose delivery bodies.

A more nuanced picture comes from the graph of the results in Figure 16. The interquartile range shown on the boxplot for each category is very wide: for district councils, for example, rates of service content vary from less than 25% at the lower quartile, to 60% at the upper quartile. This speaks to a wide variation in focus, and scope of democratic content, *within* each class as well as *between* them.

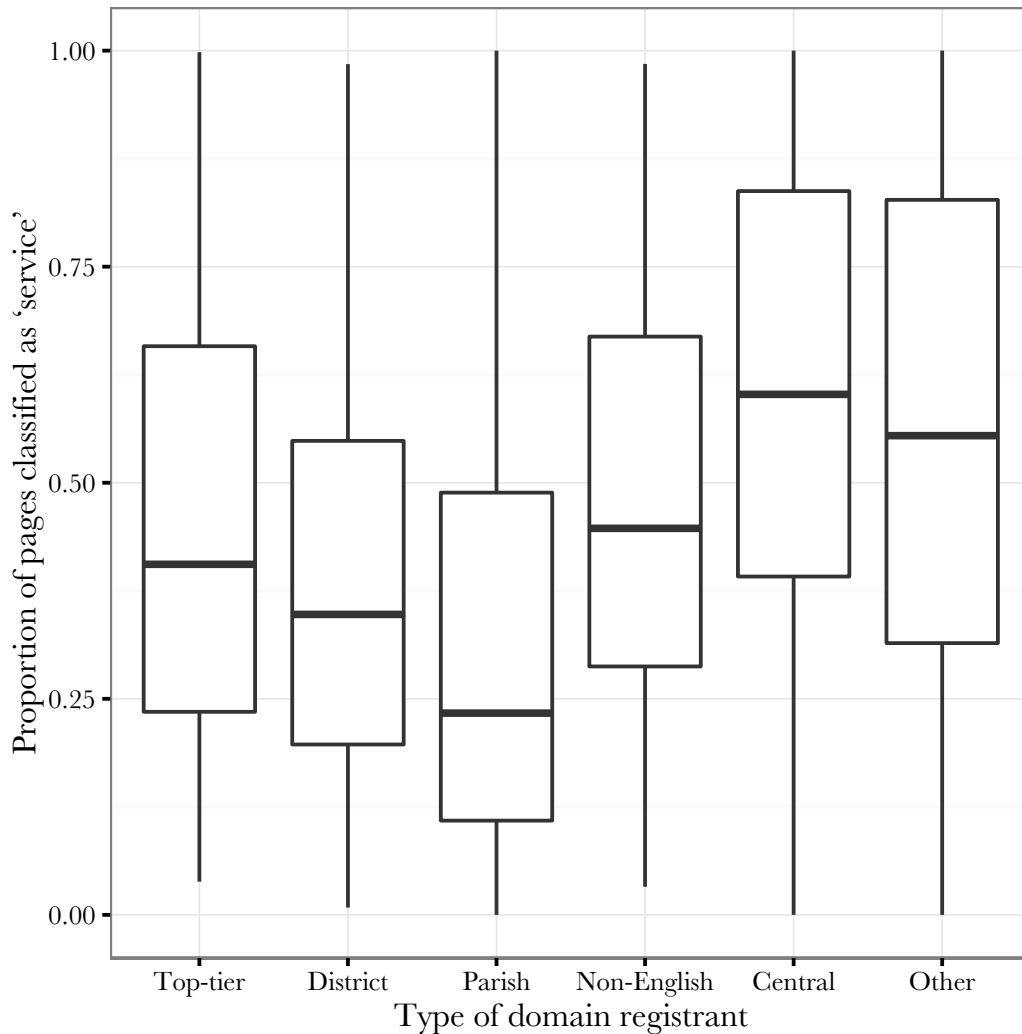


Figure 16: Proportion of service-focused pages, by domain type

This is the picture we would expect from a heterogeneous group of authorities with heterogeneous websites, which therefore confirms Hypothesis H7.

The common factor for much of the informational content on local government web sites appears to be that it is material which is already produced internally, and so the cost of publication online is relatively low. Looking through the data for subdomains of the councils in the sample, it's clear that the ModernGov archives of committee minutes and papers are often larger than the rest of the council's web presence put together. It's also likely that much of it is very rarely looked at by members of the public. However, the web availability is largely a side effect of the existence of a committee support

database run by the council for the benefit of its elected members. Despite its hugely disproportionate volume, the marginal cost of putting each paper online is very low¹⁰.

Analysis of content publishing patterns can inherently only tell us about the supply side of the equation, as it looks wholly at the actions of local governments and not those of their customers. To know more, it will be necessary to conduct a further study analysing the relative patterns of demand for the documents – that is, the relative frequency of access for the informational and transactional parts of the local government web. For reasons explored in Section 3.3.4 this is inherently difficult, but must be a key direction for future research.

5.3.3 *Classification of government web content: transactional and informational services*

This section aims to test Hypothesis H8, that transactional pages volumes are heterogeneous between local authorities.

As noted above, out of the sample of .gov.uk pages sampled and manually classified into ‘transactional’ and ‘informational’ categories, only 14 (3%) were transactional, with the vast majority of pages in the sample providing information or signposting to offline services rather than providing an opportunity to conduct an online transaction¹¹.

This sparsity of transactional content, coupled with limits in the textual differences between the categories, makes it harder to classify using standard supervised classifiers. Even during the model development stage, a similar

10 Of course, the cost of producing the papers in the first place is very high in terms of staff time, but that is an issue outside the scope of this study.

11 A short pilot study was undertaken before the main approach, with a three-level coding scheme with “interactive democracy” as a third label. This was based upon the interactivity predictions of some of the more utopian maturity model theorists (e.g. West, 2004). This was abandoned as none of the random sample of pages selected actually fit within this category; perhaps this is in line with Norris and Reddick (2013), who observed that such democratic transformation through technology essentially didn’t exist.

model to that successfully applied to the democratic categorisation achieved an F_1 of only 0.44 for detecting the transactional items¹²

Consequently, the random sample of pages hand-coded to train the classifier, together with the further hand-coded sample used to validate the classifier used in Section 5.3.2, were combined to produce a fully-hand-classified random sample on traditional statistical lines. The total size of the sample was 912 pages, of which 452 were pages from English council web domains. A breakdown of the results, separating English local authorities and the rest of the sample, is in Table 17.

	Not English council	English council	TOTAL
Informational	447	440	887
Transactional	13	12	25
TOTAL	460	452	912

Table 17: Manual classifications of pages as transactional

This use of a sample rather than the full-population is not ideal, as it limits the ability to break down the result by sub-group. Nevertheless, the number of transactional pages is so low, across both categories, that the results are stark even when accounting for the margin of error.

Using the broader sample summarised in Table 17 the proportion of transactional pages falls slightly further, to 2.65% for English councils [95% CI: 1.17%–4.14%]. This is nearly identical to the proportion for the sample as a whole.

This is a very interesting and surprising result, given the focus placed in both the academic and practitioner literature on the importance of online transactional services. It mirrors the finding in the previous section that rates of non-service-focused content are also unexpectedly high.

¹² Although extremely precise, with no informational items classified as transactional, it only detected 4 out of the 14 transactional items successfully (a recall of 0.29). Inevitably, its out-of-sample performance on a validation set not used for training would be somewhat worse.

Unfortunately, the use of a sample here makes it impossible to directly answer Hypothesis H8 with any degree of statistical confidence: the confidence intervals for the $n = 452$ sample of English local government pages in general are narrow, but there are simply too few samples to build a comparative analysis of variation between each authority individually as was possible in Section 5.3.2.

This highlights one value of the automated classification and full-population approaches pursued in this research: that it becomes possible to undertake comparisons at granular levels that are unreasonably difficult when good-sized samples need to be collected for each subject.

Local e-petitions sites offer an interesting example of the transactional uses which are found on the local government web. These have been established by all local authorities by government fiat (*Local Democracy, Economic Development and Construction Act 2009*, §2). E-petitions websites replicate the traditional petitioning function of government online, by allowing people to create, sign, and receive responses to petitions using the web rather than on paper. Local e-petitions sites were required in another modernising tendency after the success of central government's own petitions site¹³, with the requirement brought into force in December 2010.

Looking at the e-petitions sites in our dataset, though, the level of use is shockingly low. Boston Borough Council's petition site has 14 petitions. Harrogate Borough Council has only 2. Even the largest site, that for Bristol City Council, has only assembled 332 petitions.

This is a striking example of disconnect compared with the central government web, with 50,661 petitions published on epetitions.direct.gov.uk at the same time and with some petitions having hundreds of thousands of signatures. Note that this data is from three and a half years after the sites were made compulsory.

¹³ Initially at no10.gov.uk, later at epetitions.direct.gov.uk and currently at petition.parliament.uk

I theorise that this reflects two distinct differences between central and local government – firstly, the low level of democratic engagement with local government, which consistently has lower levels of voter turnout than for parliamentary elections; secondly, though, the lack of any route to critical mass from local e-petitions. We know from analyses of central government e-petitions that success is highly dependent on first-day signing rates and the punctuated equilibria typical of online activities (Hale, Margetts and Yasseri, 2013). Much of local political activity is face-to-face, though. Those with close online connections to local political activists are also not necessarily local; the means by which an e-petition could ‘catch fire’ and become widely spread are less kind to local activity.

More than that, however, it shows the limits of the kind of top-down drive for electronic government that characterised both the Blair administration’s pledge to put services online and the Coalition government’s e-petitions mandate. The sites may be made, but will they be any good, and will they be used? Local authorities are individually too distant from the ministers giving the instructions for close monitoring, and the relatively low priority given to e-government compared with the hugely expensive areas of health and education is likely to have limited long-term follow-up monitoring.

This also highlights the limitations of e-government research without a robust dataset giving the user’s perspective. For e-petitions use data is available, as uses (signatures) are by their very nature publicly visible. For essentially all other local government services, usage levels are not apparent to the researcher without further work to obtain usage figures from either local governments themselves or users directly (see Section 3.3.4).

5.3.4 *Topic modelling*

This section aims to examine Hypothesis H9 by analysing the results of the topic model fitted using the LDA process described in Section 5.2.2.

The volume of pages in each topic is quite uneven. The distribution of pages across topics for each English council has been tabulated in Table 31, in Appendix A. If the scope of provision for local authorities was homogenous, rather than heterogeneous as in Hypothesis H9, we would expect similar proportions of content in each topic for each council.

Figure 17 plots the proportions of pages for each council in each of the ten topics, with summary statistics in Table 18. More than half of the pages (2.7 million of a total 4.7 million) have been modelled as part of a single topic (topic 3).

	Min	1Q	Median	Mean	3Q	Max
Topic size	21,670	67,170	114,900	469,800	146,200	2,691,000

Table 18: Topic modelling: topic size summary statistics

This dataset allows us to test Hypothesis H9. That there is variation between councils is visually clear. A Chi-squared (χ^2) test of independence allows us to formally reject the null hypothesis and accept that H9 is true ($p = 0$). This is not, however, sufficient given the large number of categories in both dimensions of the table: practically any variation would result in a positive χ^2 test. The key question here, as in much of the rest of this research, is not “is there variation?” but “what is the size of the effect?”.

As both variables (topics and councils) are nominal and have no intrinsic ordering, many of the most effective measures of association are not available. Nevertheless, there are two measures which have Proportional Reduction in Error interpretations which therefore allow an assessment of the degree to which changes in the publishing council are associated with changes in distribution of topics:

- Goodman and Kruskal’s lambda (λ) is the most prominent of these measures. It measures the improvement in prediction of the modal dependent variable (topic) that can be made by taking into account information from the independent variable (council). It does, however,

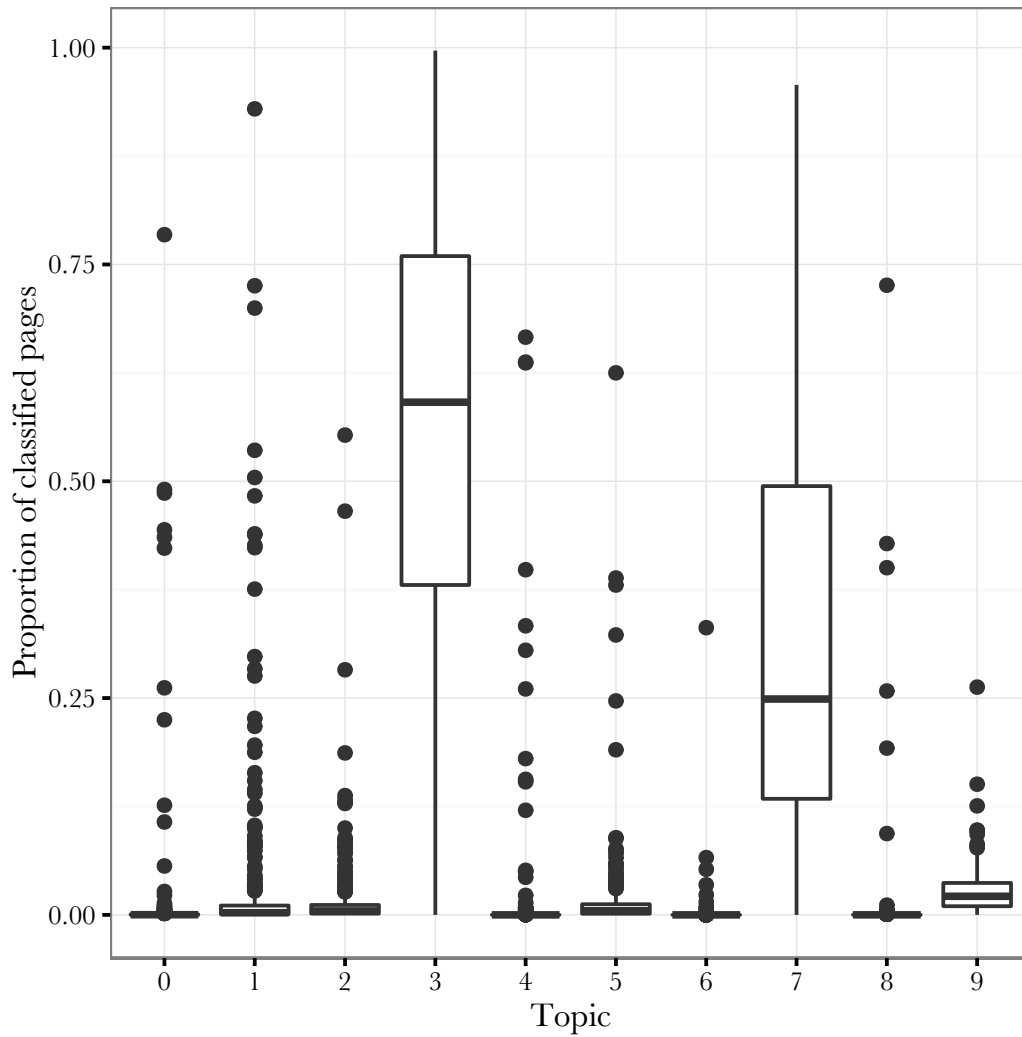


Figure 17: By council: proportions of pages in each topic

have the weakness that a dataset where data for each level of the independent variable has the same model category for the output variable reports (i.e. the variables are in accord) will have $\lambda = 0$, despite variations in the distribution.

- Goodman and Kruskal's tau (τ) is similar to λ , but bases its statistic on the probabilities of distribution of the dependent variable rather than the modal category. It is less used than λ , but for this dataset more useful as it is the extent of distribution variation rather than changes in the modal category which is of substantive interest.

The calculated statistics for this dataset are in Table 19.

Statistic	Value
Goodman and Kruskal's λ	0.184
Goodman and Kruskal's τ	0.192
Pearson's χ^2 [3,450 df, $p = 0$]	12,780,000

Table 19: Topics by council: measures of association

For this data, around 19% of the variation in topic distribution can be explained by systematic council-by-council variation, with the rest unexplained. This implies that the effect size of any observed heterogeneity is small.

A significant caveat with this topic model analysis is that we cannot directly measure the quality of the generated set of topics in terms of theoretically meaningful categories. Unlike the hierarchical Infomap process used in Section 4.3.2 no hierarchy of higher-level topics is extracted. In contrast to most modelling processes, the process by which these topics were selected from amongst the distribution of possible ones is opaque.

The total number of pages assigned to each topic, together with selections from the keywords extracted using the process in Figure 13 and an indicative label for the topic, are in Table 20. Variations in the degree of substantive coherence of the topics are clear: the 'planning diagrams' and 'accounting' topics are highly specific and the content is relatively consistent within the group. The shared feature of the 'contact forms' topic, on the other hand, is the presence of a common list of the world's countries, found in various drop-down lists on address forms, and two groups are largely defined by the contents of specific councils' websites. Many agendas and minutes, but not all, are to be found in the 'committee documents' topic, whereas the 'service delivery' group contains a wide variety of different pages mainly delivering specific local government services.

These reasons, together with the uneven size of the topics shown in 18, should make us cautious about reading too much into these results.

Num	Topic	Selected keywords	Pages
0	Lichfield	Wellingborough, Lichfield, search, browse	146,898
1	Committee documents	committee, scrutiny, minutes, agenda	144,030
2	Planning applications	storey, rear, extension, road, proposal	86,466
3	General	been, within, local, area, development	2,691,444
4	Street repairs	fixmystreet, reports, update, problem	87,225
5	Planning diagrams	lane, house, survey, licence, infringes	52,976
6	Contact forms	Republic, Islands, Guinea, Netherlands	21,674
7	Service delivery	services, contact, search, information, housing	1,264,364
8	Ely tourism	Ely, visit, food, ghosts, Cromwell	60,735
9	Accounting	capital, supplies, Ltd, total, 110	142,667

Table 20: LDA: topic labels and selected keywords

5.4 DISCUSSION

This chapter has set out to answer research sub-question 2: What is the content of the English local government web? What can this tell us about local government and its approach to e-government?

The first thing that is striking from looking at a random sample of local government web output is the sheer volume of documents published which are primarily for internal use. Committee paperwork is routinely published online, including agendas and papers for the multiplicity of groups which a modern local authority operates. Further to this are quasi-judicial processes and notices, such as the documents for individual planning applications, premises licence requests and so on. Finally, there is a large group of documents dedicated to public transparency: reports that may have originated with the service delivery side of the authority, but which are primarily for internal use and published by government decree. This includes responses to Freedom of Information Act requests, documents such as Corporate Plans, and specific information sources such as records of all council spending over £500¹⁴.

¹⁴ This was the result of an initiative from the coalition government under Eric Pickles

This cuts against the prevailing rhetoric discussed in Chapter 2, which focuses on the benefits of e-government being in the better delivery of services. The revealed preference of local authorities, however, seems to privilege at least in volume terms the transparency and regulatory aspects of information provision above the service delivery ones.

With many of these documents, it is likely that usage levels by the public are very low. They are, nevertheless, produced as process documents for other activities of the council, so the costs of putting them online are not likely to be very extensive beyond the costs of commissioning the web-linked databases they are kept in. ModernGov, for example, supply committee management systems to many English local authorities which have the capacity to automatically publish minutes and agendas to the web. For smaller authorities, such as parish councils and special purpose authorities, these are often the dominant items by volume on their website and have been uploaded by hand. This substantially differs from their private sector equivalents, where internal documents are not published, and “investor relations” segments are carefully curated. Central government has likewise been separating bureaucracy paperwork¹⁵ from service delivery functionality on .gov.uk.

The overwhelming majority of committee documents are published as Portable Document Formats (PDFs), with essentially no interactive features, links with related subjects or other scaffolding which would make them useful as online reference documents. A substantial fraction of them are, in fact, PDF formatted image files, made from scanned paper originals, which is not entirely in keeping with the spirit of “digital first” delivery.

Two kinds of directories also represent large volumes of local government web content. Firstly, electronic versions of the A–Z of council services are nearly universal and recommended by Socitm. More significantly, many councils are now hosting large online directories of local services and sometimes businesses

15 Originally headlined “inside government”, now the more mundane “departments and policy”

operated by other agencies. These latter potentially serve an important signposting role for local people, especially as the local government service delivery role is increasingly fragmented, but do not represent services delivered by the local authority directly, so their categorisation as local government pages is marginal.

Within local government sites, the boundary between internally-focused and externally-focused documents is not always clear. Freedom of Information Act requests, for example, are the release of internal documents specifically by public request. Likewise, many of the major public-facing council documents such as Local Plans go through an internal series of approvals before being formally published and will likewise show up on both halves of the classification.

Even more striking than the volume of non-service-focused content, however, is the paucity of transactional content revealed in Section 5.3.3. While the small-scale sample used for this analysis limits our ability to explore usage levels between different local authorities, the overall levels are can be obtained with good accuracy from the sample. With only 2.65% of the sampled content transactional in scope, it's clear that the realities of government provision online do not match our assumptions of service delivery.

5.4.1 *Homogeneity, heterogeneity, and prospects for a local.gov.uk*

There is little evidence, from this analysis, of any homogenising effect of online service delivery on local authorities. The size, content, and linking patterns of local governments online are hugely variable.

The effects of central government's consolidation at www.gov.uk are apparent within the crawl data, with key departments of state like the Foreign and Commonwealth Office having entirely retired their main www.fco.gov.uk site in favour of pages on the consolidated site. On the other hand, it's equally clear what a long tail of non-mainstream content is hosted by government

departments: the Foreign Office have active subdomains blogs.fco.gov.uk, arabic.fco.gov.uk, www.hrdreport.fco.gov.uk (which hosts a bespoke website with the results of the department's Human Rights and Development Report), www.visa4uk.fco.gov.uk, legacy.visa4uk.fco.gov.uk, treaties.fco.gov.uk and others.

The local government web, as revealed in this dataset, on the other hand, is a huge and multifaceted group of content both internally and externally focused. Little homogeneity is visible in the dataset, with different authorities presenting different pictures in scale and scope of content. It is inconceivable that this would not result in hundreds of similar anomalies even in a world where websites were nominally centralised under local.gov.uk.

If such a project were to be undertaken, it would certainly not be a simple tidying-up exercise resulting in savings. To an even greater extent than gov.uk, it would to be a major undertaking. While IT improvements certainly can be used as an opportunity for improvements in public administration, it's difficult to see how the consequent wholesale central control over local government services and information structures could be justified.

5.4.2 *The value and limits of web data for studying government*

The webometric and content analyses presented here and in the previous chapter are interesting and present both new facts and new insights into the behaviour of governments online. However, this project has also identified the limits of using web data to study government. This provides a partial answer to the second part of the research sub-question by reflecting on the linkages between online and offline views of government.

The first serious limitation is that the presentation of government online is only loosely related to the activities of government itself. The topic classification of web content discussed in Section 5.3.4 highlights the huge volume of local government publication in particular which is related to the democratic and

regulatory aspects of local governance. Using budget figures as a proxy for the activities of local government considered most valuable, however, we can see that the relative proportion of web publishing and activity related to social services in particular is tiny in proportion to the amount of attention and money spent by the councils themselves.

The second limitation is one of presence vs visibility. Put shortly, the majority of publications by public bodies are part of a long tail of content not regularly viewed by the public. The large majority of regulatory and democratic pages belong in this category: they are put up for transparency but not with a particular expectation of use. They certainly don't have high visibility to the casual web visitor, as they are generally held in a database on a separate "democracy" server and not directly linked from the front page.

As such, trying to extract insights even from a full content and links study of government would be a little like extracting insights on Amazon by looking only at its product catalogue: without knowledge of which pages people are visiting (or which products they're buying) we can see only the supply side of the equation and not the picture as seen by actual citizens.

Consequently, if similar studies are to be undertaken in the future by political scientists it is imperative that measures of use are also developed.

The other set of challenges, that the transactional content sampled was too sparse to build an effective model, may be solvable by technical means. Following on from the initial attempt, a general-purpose classifier optimised for F_1 which equally weights precision and recall, an experimental preliminary run was made with a classifier optimised for $F_{0.05}$: that is, for specificity of selection instead of completeness of recall. This was run on the whole dataset to extract probable transactional pages.

For further analysis a group of these, perhaps a supplementary sample of 150, could be checked, classified by hand, and added to the original random sample for further model training. By providing additional data for the 'transactional' side of the categorisation, and also collecting data to correct some

false-positives generated by the initial model, this could allow a much more effective re-training process to be carried out. A further alternative to this approach could be to use tools from the information retrieval field, which show promise for dealing with extremely sparse political datasets.

5.5 CONCLUSIONS

There is a very large volume of content in the local government web which has been published by councils but which has no service purpose. A substantial proportion of actual council web presence is dedicated to the hosting of democratic services databases and to providing access to planning and licensing decisions. Moreover, councils' web content appears to be substantially heterogeneous, with disparate volumes, page classifications, and topic distributions for different authorities.

There is no evidence from this dataset that the publicly-facing digital activities of local governments in particular are established in such a way as to support the transactional services which are theoretically the source of much of the predicted service improvements from e-government. It's not clear that the work that has been done to date will, therefore, support the pressure of service improvement and value for money aspirations placed upon them by public administration scholars.

This chapter has also highlighted a key point for future research, which is the importance of getting measurements of use. Without an understanding of which parts of the government web are valued and used by citizens, our understanding is necessarily only partial.

The present dataset does not allow a robust comparative assessment to be completed between quality of web provision and improvements in services, simply because the data here are for a single moment in time and do not reveal a direct measure of quality. Consequently, Chapter 6 tackles this question

directly, by building a supporting time-series dataset based upon government performance figures and expert assessment of local government websites.

COUNCIL PERFORMANCE AND USE OF THE INTERNET

What is the relationship between council administrative performance, cost-effectiveness and use of the Internet? What explains this relationship?

6.1 INTRODUCTION

As outlined in Section 2.4, the rhetoric surrounding the opportunities for service transformation from local government use of the Internet has been fulsomely positive from the early part of the period under study. Repeated government and practitioner reports outlined the necessity of bringing services online, and of adopting the kind of Internet-enabled dynamism exemplified by the .com entrepreneurs of the late 1990s (Cabinet Office, 1999b, 2005). Practitioner literature, though generally less utopian, saw the Internet and e-government as a substantial opportunity, as well as a challenge (Socitm and Improvement and Development Agency, 2002).

The amount of work done to justify this enthusiasm, by contrast, is much smaller. Again, a sampling of the available studies is discussed in Section 2.4. In principle, questions about whether organisations actually have been transformed using Internet-based approaches are empirically testable, unlike the more speculative dialogues about whether such changes might theoretically be possible. This work is the focus of this chapter.

Due to the unique features of English local government both council performance and council use of the Internet are easier to operationalise than they would be in other contexts. Because of the intense government pressure for accountability and data-driven performance improvement in the period under study, relatively good-quality proxy measures for both these variables of interest are available. Critically, they are available not just for a convenience-sampled subset of authorities, but systematically – and collected to a common standard across local authorities¹.

The rest of this chapter will outline the Comprehensive Performance Assessment (CPA) and Better Connected (BC) datasets, explore their utility as a means of answering the research sub-question above, then construct a novel panel data series to offer an empirical answer to the sub-question. Three separate hypotheses, relating Internet engagement to service quality, cost efficiency, and citizen satisfaction, are outlined and tested.

6.1.1 *The Comprehensive Performance Assessment*

The most attractive high-level council performance measurement is the CPA dataset produced by the Audit Commission as part of the oversight process for local authorities. CPA results were calculated between 2002 and 2008².

The CPA data corresponds to English top-tier authorities, which are those councils responsible for larger-scale or more strategic services such as education, social services and transport. In some areas of the country, these are provided by county councils with several smaller-scale shire district councils acting as a lower tier of local authority providing local services such as refuse collection, housing and environmental health regulation. In other areas, both

1 That standards were laid down by the government and the Audit Commission for the data to be collected does not mean that it was collected entirely uniformly across the country, of course (there were strong incentives for authorities to produce “good” results) but the accompanying audit procedures give more confidence than with most local performance indicator data.

2 For 2009, the series was superseded by the dissimilar Comprehensive Area Assessment measures, themselves cancelled after only one year of collection

sets of services are provided by the top-tier authority, one of the unitary councils of various kinds³. Shire districts are smaller than top-tier authorities and generally represent areas which are more rural, richer, older, more Conservative and more southern than non-County top-tier authorities (see Section 2.4.1 for more details).

Although CPA results were published on a single occasion for shire district councils (Audit Commission, 2005a), they were never subject to the same degree of scrutiny as the main CPA series and less data is available to measure their performance.

An original motivation for the system was to provide a simple score against which local residents could hold their councils to account, as well as a mechanism with which government could drive up performance (DTLR, 2001). The underlying approach is to combine ratings from across a number of service areas together with the results of a corporate assessment of capability to improve, into a single score.

The original service blocks selected were education, social care, environment, housing, libraries and leisure, benefits, and use of resources⁴ (Audit Commission, 2002b). Each service received a rating between zero and four based upon inspectors' judgements and performance indicator results, which were then combined to make a single service performance score⁵, which was

3 There are multiple technical distinctions between London boroughs, metropolitan boroughs, unitary districts, unitary counties, and the two *sui generis* authorities, the Common Council of the City of London and the Council of the Isles of Scilly. These variations in powers and duties exercised are minor, tedious and generally the result of historical accident rather than any overarching plan. To confuse matters further, district councils may be known for historical reasons as borough councils or city councils regardless of whether the council is a shire district or a unitary district. Worse still, there are a few parish councils (the very local tier below districts, not normally considered within the group of principal local authorities) which are called city councils, again for historical reasons. For those not intimately familiar with the complexities of English public administration, it is generally necessary to refer to an index to know the exact status of any given local authority.

4 Not a separate service as such, this block covered financial performance indicators, audit judgements and corporate planning and was nevertheless included in the service scores.

5 Some services, principally education and social care, were weighted more heavily in this process than the others. The precise method of achieving this weighting evolved over the CPA period from a simple score multiplier in 2002 (Audit Commission, 2002b, p.8) to a hugely complex rule-based formula in the 2008 edition, which was designed to avoid councils with poor individual

Coding	2002–2004	2005–2008
0	Poor	0*
1	Weak	1*
2	Fair	2*
3	Good	3*
4	Excellent	4*

Table 21: CPA coding and grade label changes

then further combined with the corporate assessment to produce a final CPA score.

The makeup of the indices used to construct scores varied somewhat over the years of data publication as the criteria were adjusted by the Audit Commission and by central government, including the addition of a fire and rescue service block for those county councils with fire authority responsibilities. There was a general intention to keep the scores comparable between years, though criteria were generally toughened in response to councils' increasing scores. Indeed, the refresh of the methodology for the 2005 exercise was explicitly subtitled "The Harder Test"⁶, which noted that the old Poor–Excellent and the new 0*–4* ratings were no longer directly comparable (Audit Commission, 2005a). Full details of the formulae for constructing the scores in different years can be found in the Audit Commission publications which accompanied them (Audit Commission, 2002a, 2003, 2004b, 2005b, 2006, 2007b, 2008); Table 21 shows the changing labels and how they have been coded.

Despite these changes in methodology, the trends for council scores are fairly clear: in general, councils did better as the assessments went on (see Table 22 and Figure 18). Although CPA ratings were expressly targeted by council senior management as performance goals, and therefore have all the usual issues created when measures become targets, the comparator BC scores

service blocks from achieving high scores, irrespective of the quality of the other parts of the authority (Audit Commission, 2008).

⁶ Intriguingly, the undoubted toughening of the methodology in 2005 did little to arrest the long-term trend of authorities achieving progressively higher scores – as can be seen in Figure 18.

	2002	2003	2004	2005	2006	2007	2008
0	13	10	1	1	0	0	0
1	22	18	15	9	5	2	4
2	39	40	33	31	23	22	26
3	54	56	60	70	71	67	57
4	22	26	41	39	50	58	62

Table 22: CPA ratings by year

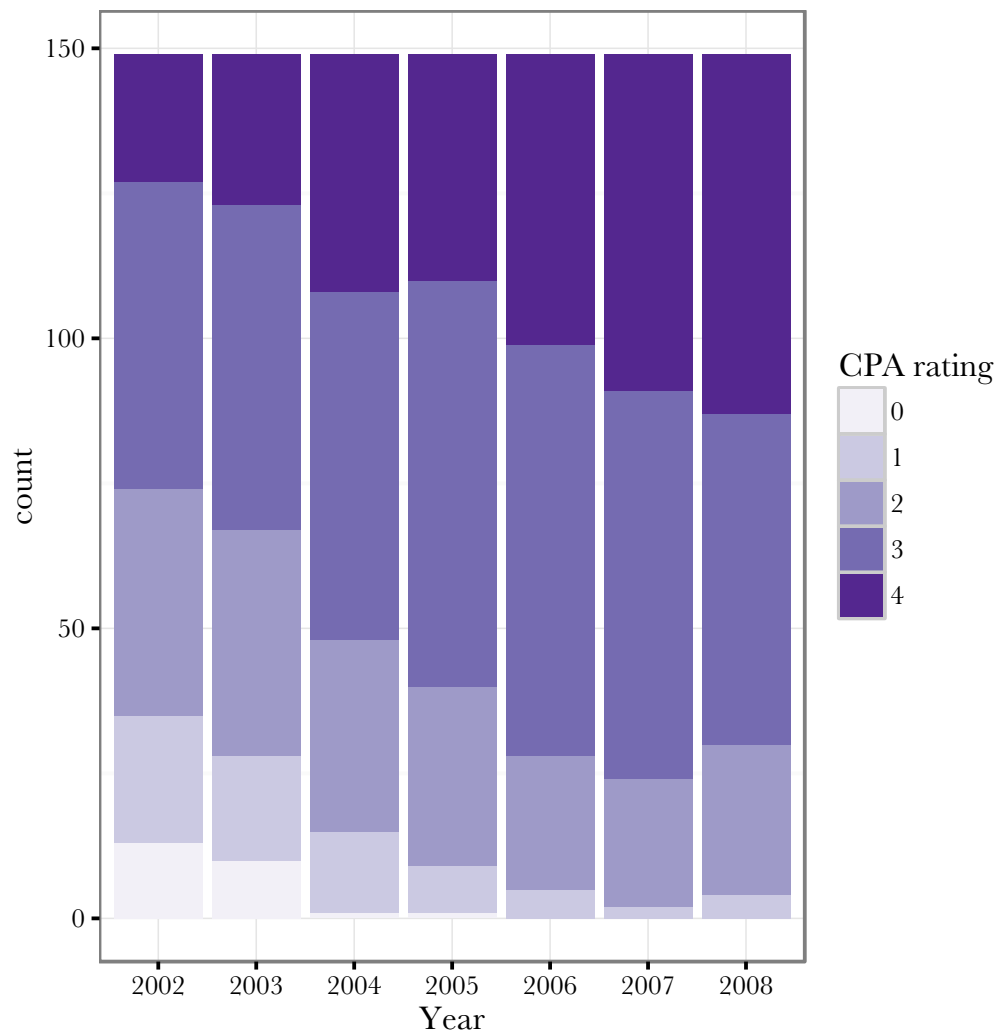


Figure 18: CPA ratings by year – top-tier authorities

are targeted in a similar way by council web teams, so some of the upward variability over time should be common to both of the key variables in the analysis.

This CPA dataset has some strengths and some weaknesses for our research purposes. Chief amongst the weaknesses is that the results are not contemporary⁷. This is mitigated by the data coinciding with a period when local government was markedly developing its web provision, which supports comparative evaluation. Further advantages of using CPA data as the main measure of council performance include complete coverage across our target population of local authorities, and a degree of credibility that is difficult to achieve with researcher-constructed composite measures. The CPA measures were embedded in local authority work for many years and have been used widely in academic studies which need a council performance measure (e.g. Andrews, Boyne and Walker, 2006; Walker and Boyne, 2006; James and John, 2007; Boyne et al., 2010) and as an object of academic study in their own right (e.g. Andrews, 2004; Andrews, Boyne, Law et al., 2005; Game, 2006; Lockwood and Porcelli, 2013)⁸. As Revelli (2010, p.187) notes, “CPA has the unique feature of combining information on public service level and quality with indices of costs of services, thereby approximating an ideal performance score that promises to be superior both to indices built on gross output only and to crude measures of technical efficiency.”⁹

For this research, the service performance scores have also been disaggregated from the overall ratings, in order to allow a closer focus on service outcomes rather than the Audit Commission’s judgements of corporate capacity (which made up an increasingly large component of the scores as time went on). This disaggregation has been made needlessly complex by the construction of the CPA framework, which not only changed every year but also

7 The last CPA set was produced in 2008.

8 Game (2006) is a particularly good critical introduction to CPA and its methodology.

9 Some of these studies couple CPA measurements, either using the overall judgement or just the service block scores, with other internal measures of performance. Given that the Audit Commission had the ability to specify whichever internal measures it felt would be useful in constructing the CPA, it’s not abundantly clear that adding further general measures would improve the quality of the measurement. Certainly any other data systematically collected within local authorities will have been used for performance management purposes and will therefore be subject to all the same perverse incentives as the CPA components.

provided a changing combination of averaging (of already averaged service indicators) and rule-based approaches to summarisation. Previous research using the CPA has also taken a similar approach, using bespoke combinations of service performance indicators (James and John, 2007; Lockwood and Porcelli, 2013) or a version of the Audit Commission's original arithmetic approach to construction of an overall services score (Boyne et al., 2010).

In order to provide better comparability between years, the government's service score rules have been replaced here by a simple weighted average of service scores, with those services identified by the Audit Commission as "Level 1" (more important) being double-weighted. For the regression analyses which follow, this score has been normalised across the whole population to produce an easily interpretable measure with a mean of 0 and a standard deviation of 1.

This measure not only separates out the service quality effects from the other aspects of the CPA scores, but by bypassing the final rounding also provides a finer-grained measure than the five-category ordinal official output. The flipside is that this alternative measure has less of an official imprimatur than the published final CPA grade, and does not take account of the implicit government judgements on the importance of particular services which are contained in the rule matrices.

A graph of the constructed measure is Figure 19. Interestingly, the "Harder Test" claims of CPA 2005 show up much more strongly as a discontinuity here than they do in the final CPA scores in Figure 18. This will be best handled in analysis by treating the year as a series of binary dummy variables rather than a single continuous one; in that way, there is no presumption of linearity in the effects of time.

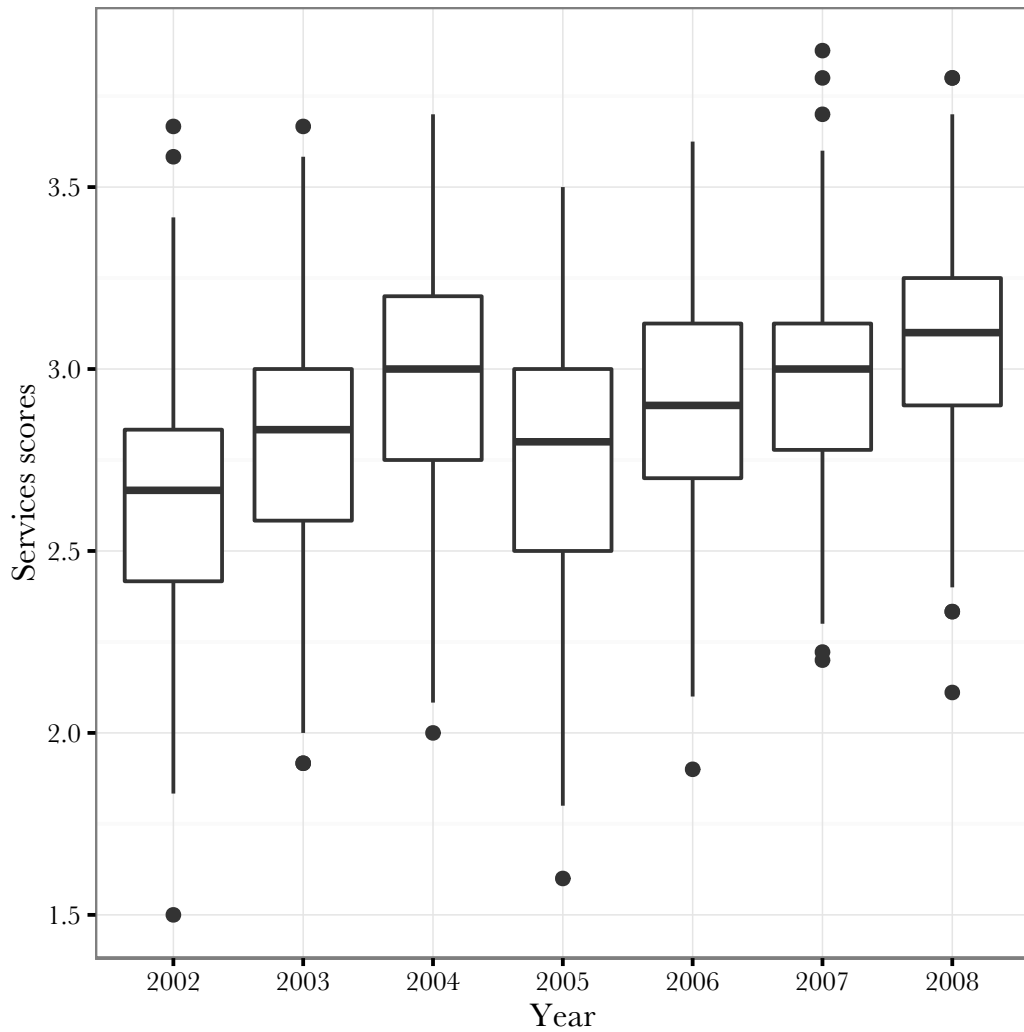


Figure 19: CPA-derived service scores by year – top-tier authorities

6.1.2 *Better Connected*

As outlined in Section 2.4.6, Socitm's BC data has been made available for this research project, to measure the extent to which councils are developing their web presence. The first data release in the series (Socitm, 1999) set the pattern for the years to follow: judgements were made on both a quantitative measurement of various features considered important and also a qualitative assessment of quality of design and navigation. Sites were graded "Unclassified", "Promotional", "Content", "Content+" or "Transactional" (though there were no sites graded transactional in the first year). The labels

are perhaps slightly misleading: they are applied based on cut-points from an underlying one-dimensional scale and do not directly represent particular strengths at information delivery or transactional services (though the scale is assembled such that “Transactional” is not reachable on the strength of information content alone).

This grading scheme continued until the 2007 exercise, with the number of unclassified sites (normally those with no site at all) and promotional sites falling year-on-year (see Table 23 and Figure 20). In 2007, an additional grade of “Excellent” was added to the rating scale, with the remaining “Content” and “Content+” sites being regraded “Standard”.

Although the methodology is intended to be as objective as possible, the authors note that “Website assessments are not an exact science, however much we try to be analytical and objective” (Socitm, 2013a, p.10).

Spreadsheets were obtained from Socitm containing data for each assessment for each council for each year from 2006 onwards, combined with the final reports for those years (Socitm, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013a).

For 1999–2005, the final reports were supplied (Socitm, 1999, 2000, 2001, 2002, 2003, 2004, 2005), the data from which was carefully input¹⁰. The handful of inconsistencies between the reports and the spreadsheets were resolved with reference to previous years’ results published in later data sources.

As the methodology is developed every year, there isn’t strict comparability between different years’ ratings, though it would be reasonable to take them as the review team’s best judgement as to which council websites were performing best in any given year’s assessment.

A key question with the BC dataset is the extent to which councils working to achieve the standards laid down by Socitm can be treated as a reasonable

¹⁰ The BC reports have restricted circulation and are only routinely available to members of Socitm’s *Insight* service. The data used in this chapter were directly input from Socitm’s original results spreadsheet (years 2006 onwards) and manually input from the tables printed in the published reports (1999–2005). Data were obtained with the kind permission of Socitm.

	2002	2003	2004	2005	2006	2007	2008
Promotional	13	3	1	1	0		
Content	63	56	33	23	30		
Standard						80	57
Content+	69	83	96	98	78		
Transactional	4	7	19	27	41	68	88
Excellent						1	4

Table 23: BC grades by year 2002–2008: top-tier authorities

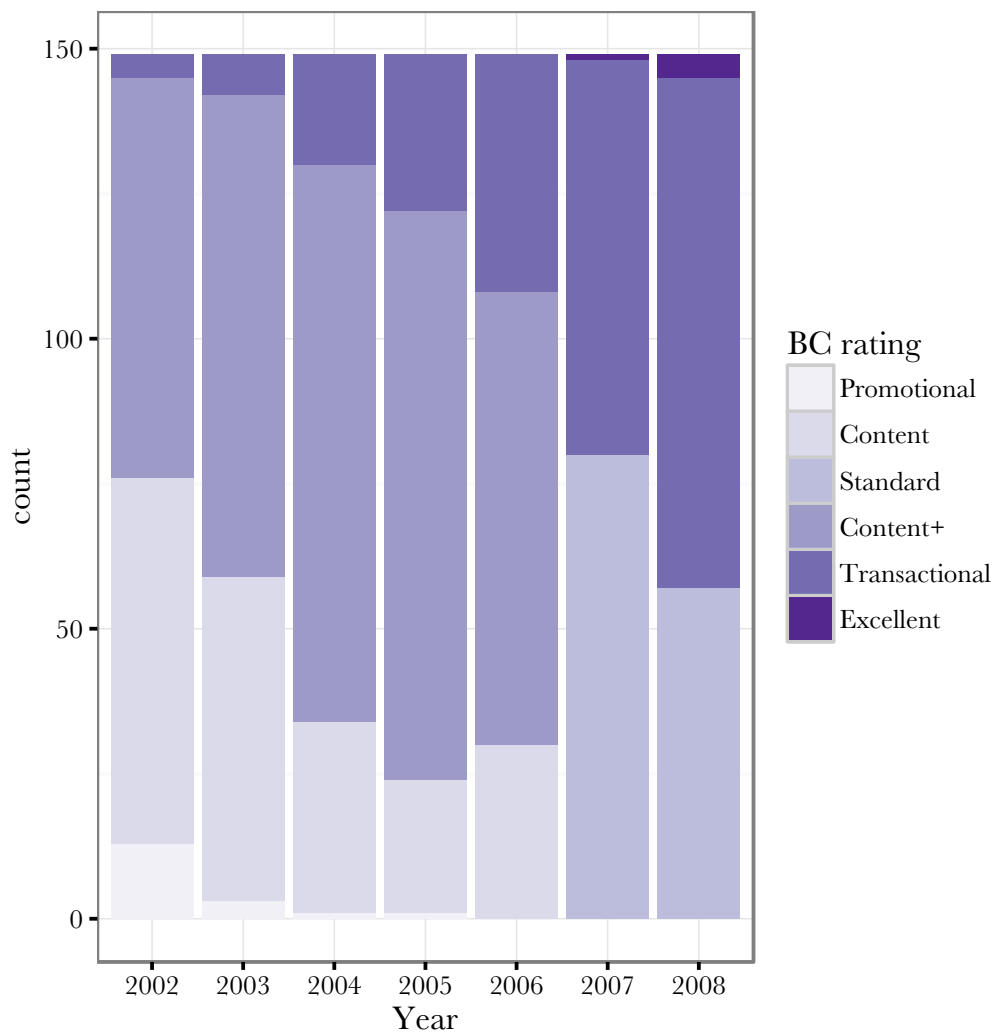


Figure 20: BC grades by year, 2002–2008 – top-tier authorities

proxy for their degree of web engagement. On the one hand, success in this dataset is obviously hard-earned. On the other, the majority of the focus is implicitly on customer-facing “top tasks” (selected on the basis of Socitm’s

Website Performance Service user survey data) rather than infrastructure building or breadth or depth of content.

In the context of the policy context discussed in Section 2.4.3, this is the appropriate set of measures. The rhetoric underlying local e-government activities, and therefore the standards against which they should be measured, has been in line with Socitm's approach. Indeed, the BC data series was largely created to generate a measure of the extent to which local governments were engaging with the web as they were being encouraged to by local and central government bodies.

6.2 INTERNET ENGAGEMENT AND SERVICE QUALITY

Hypothesis H10: Local authority engagement with the internet, as measured by BC performance, is positively associated with service performance, as measured by CPA scores.

Hypothesis H11: Local authority engagement with the internet, as measured by BC performance, is negatively associated with the cost of delivering services, as measured by spending in excess of the government-established standard when controlling for service performance.

As an exploratory assessment of the association between levels of internet engagement at an authority and that authority's service performance, a regression model (Model 5, Table 24) was constructed, using the CPA scores for 2008 (Audit Commission, 2008) as the dependent variable and the council's overall rating on Socitm's BC assessment of website performance (Socitm, 2008) as the key explanatory independent variable. The CPA data are, in principle, in ordered categorical form (the ratings consist of categories from 0 to 4, which have been variously labelled over the years of CPA as poor to excellent, or 0* to 4* – see Table 21).

	CPA rating (5)
BC Transactional or better	-0.093 (0.138)
Constant	4.246*** (0.108)
N	149
R ²	0.003
Adjusted R ²	-0.004
Residual Std. Error	0.818
F Statistic	0.459

* p < .1; ** p < .05; *** p < .01

Table 24: Pilot data analysis: CPA and BC, 2008 only

As this model showed no support for the idea that greater engagement with the Internet was associated with higher overall service performance (in fact, the estimated effect was very weakly negative – see Table 24), a more robust analysis was undertaken by constructing a full panel dataset of CPA data alongside BC results and control variables.

The results of CPA have been shown to be influenced by the demographic and economic characteristics of local authorities in addition to the intended measure of council competence (Andrews, Boyne, Law et al., 2005) despite the initial claims of the Audit Commission that there was no strong correlation in pilot studies and thus “no justification” for adjusting on the basis of deprivation (Audit Commission, 2002b, p.14)¹¹. Deprivation would, therefore, normally be added as a control measure in this analysis. As the particular circumstances of each authority, including deprivation, are being controlled

11 This is likely due to the large numbers of performance indicators which were used to calculate service scores, many of which are themselves not robust to local poverty.

by the panel regression method selected, no control variable is required for the time-invariant effects of deprivation¹².

By contrast, time-varying control variables may be useful. In particular, this approach can be used to separate out the effect of competent management (which is a component of the CPA rating) from high service quality; the “corporate assessment” measure of management is published alongside the overall outcome, so could also be controlled for by adding as an independent variable.

6.2.1 *Regression modelling of BC and CPA results*

The time frame analysed here covers 2002–2008, which is the full period for which CPA data are available to combine with the longer BC series. The data set contains only top-tier authorities and excludes shire districts (for which, as noted above, CPA results were only calculated once and on a very different basis). Furthermore, the Council of the Isles of Scilly has been excluded. It is a *sui generis* authority with an unusual scope of responsibility and for whom the standard CPA measure is less reliable. Excluding it from analysis is commonplace (Revelli, 2010). I differ from Revelli in including the Corporation of the City of London, whose functions are much closer to those of a standard unitary authority, despite its unusual status.

To test whether the BC and CPA measures have enough association between them to justify a more complete analysis, a simple measure of association was calculated between the two for the full set of authorities for the full set of years for which both datasets are available (2002–2008). The measure of association

12 If the hypothesis were directly about the effects of deprivation on CPA scores, a different analysis method such as a random-effects analysis would have been chosen. In principle, some measure of changes in deprivation over time could have been included by including the various Indices of Deprivation produced by Department for Communities and Local Government (DCLG) and its predecessors as a time-varying factor (DTLR, 2000; ODPM, 2004; DCLG, 2007, 2010a). As there are no compelling theoretical reasons to suggest that the impact of deprivation changes over time, this analysis has not been done.

Year	γ	SE
2002	0.31	0.10
2003	0.31	0.11
2004	0.21	0.12
2005	0.02	0.13
2006	0.05	0.11
2007	0.22	0.13
2008	-0.06	0.13

Table 25: BC / CPA association by year

chosen was Goodman and Kruskal's gamma (γ), which is appropriate for the comparison of two ordinal measures.

γ for the full dataset is 0.247 [CI: 0.178–0.315]. That suggests a positive relationship between these two variables, with a reasonable effect size. As the data are a population and not a sample, the confidence interval is not directly meaningful.

Data for CPA, BC and control variables are available for multi-year periods, allowing “one of the major advantages of panel data: test for the presence of omitted time-invariant factors that may bias our results” (Boyne et al., 2010, p.1270). Interestingly, when γ is calculated separately for each of the years of data, we find that the relationship becomes less strong over time (Table 25, graphed in Figure 21). This strongly suggests that a modelling method should be chosen which models the temporal component.

A fixed effects panel data model was fitted (Model 6 in Table 26), using the Least Squares Dummy Variable (LSDV) method (Wooldridge, 2010, §10.5.3). This provides unbiased estimates of the coefficients even in the presence of unobserved time-invariant heterogeneity (Bröderl, 2005, p.21).

The panel approach has previously been used to study the CPA (Boyne et al., 2010). In this method, a standard Ordinary Least Squares (OLS) regression is undertaken, but with dummy variables included for both time (in this case

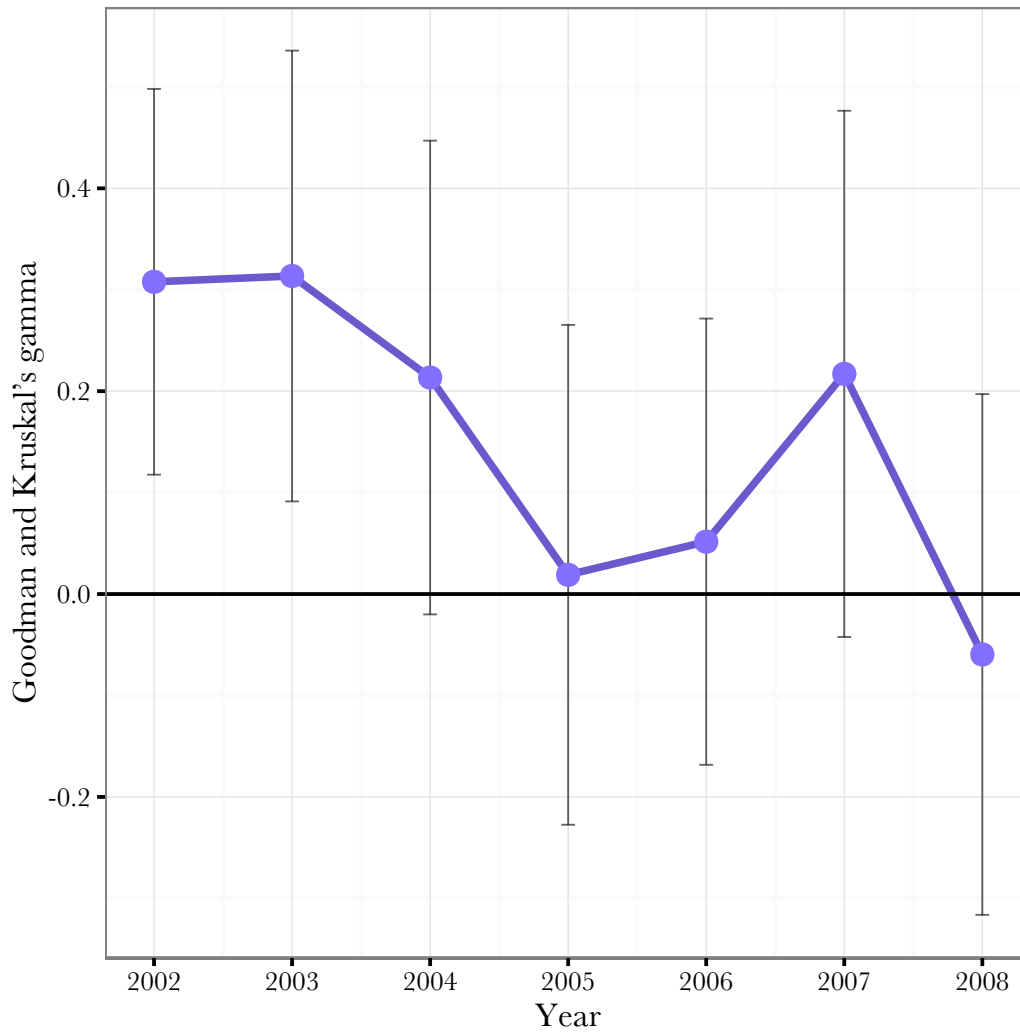


Figure 21: BC/CPA association by year

year) and individual (in this case council) to model variation for all authorities over the sampling period and between the authorities across all years¹³.

The fixed-effects choice is supported, in addition to its theoretical merits for this dataset, by the large number of units and small number of observations per unit, for which fixed effects approaches are more robust to error (Clark and Linzer, 2015).

As the individual council-level estimates are not the focus of this analysis, the approximately 150 coefficients calculated for them have not been printed in

¹³ In the straightforward linear case, LSDV is equivalent to the demeaning-based panel methods often used by econometricians.

	CPA services		CPA rating
	(6)	(7)	(8)
BC Transactional or better	-0.001 (0.064)		-0.022 (0.053)
BC Promotional		-0.272 (0.182)	
BC Content		-0.121 (0.098)	
BC Standard		0.019 (0.079)	
BC Content+		-0.026 (0.085)	
BC Excellent		0.319 (0.277)	
2003	0.513*** (0.066)	0.492*** (0.067)	0.123** (0.055)
2004	0.864*** (0.067)	0.824*** (0.069)	0.352*** (0.056)
2005	0.295*** (0.068)	0.248*** (0.071)	0.353*** (0.057)
2006	0.581*** (0.070)	0.535*** (0.074)	0.450*** (0.059)
2007	0.783*** (0.074)	0.686*** (0.111)	0.524*** (0.062)
2008	1.145*** (0.077)	1.043*** (0.108)	0.643*** (0.064)
CPA corporate assessment	0.390*** (0.055)	0.382*** (0.055)	0.922*** (0.046)
Constant	-2.131*** (0.251)	-2.085*** (0.258)	0.999*** (0.210)
N	1,042	1,042	1,043
R ²	0.725	0.726	0.804
Adjusted R ²	0.676	0.677	0.770
Residual Std. Error	0.565	0.564	0.473
F Statistic	14.932***	14.625***	23.311***

* p < .1; ** p < .05; *** p < .01

Council-specific dummies omitted. Base BC rating for Model 7 is "Transactional"; base year is 2002.

Table 26: Regression models: CPA and BC, 2002–2008

the regression tables presented here, though the between-authority variation they capture is included in the models and their respective R^2 values.

The dependent variable was taken as the constructed CPA services score discussed in Section 6.1.1, normalised to a mean of 0 and a standard deviation of 1. This reflects our interest in the service quality judgements of the CPA inspectors rather than their assessment of council management. The control variables included in the model are year dummies (both because the CPA assessment process was updated every year and because of the decreasing relationship between BC and CPA scores over time – see Figure 21), and the council’s corporate assessment rating. This latter presumptively has an association with both the BC rating (higher quality management teams are more likely to push web development) and with service scores (higher quality management teams are more likely to pursue *other* methods which will lead to improved service outcomes). Most of the other factors tested by Revelli (2010) are common to a particular council across the panel, and to which the fixed effects model is robust. A model was also tested with council expenditure and the presence of a local election as independent variables, but these added nothing substantive and were therefore removed from the final model.

Tools for building hierarchical models using ordinal independent variables are less developed than those for ordinal dependent variables; it is conventional to either treat the ordinal variable as linear in some sense and use it as a continuous variable, to binarise it at a suitable cutpoint, to code it with multiple dummy variables as a categorical variable, or to code a series of overlapping “greater than x ” dummies.

For the regression models which follow, the BC results have been binarised into a single dummy variable: whether the result was “Transactional” or better. This is a simplification of the full picture which could be obtained by handling the data in an ordinal fashion. This approach has been taken for two reasons, one theoretical and one practical.

Note that the models in this chapter have a varying number of cases, despite the intention to keep the analysis full-population. This is partly due to a single missing case for the CPA services data¹⁴ and a missing authority for the excess spending data¹⁵ and partly because of variations in the number of years included: CPA data covers 2002–2008, excess spending data 2002–2007, and satisfaction data was only collected in 2001, 2004, and 2007. In all cases the analysis has been conducted with the most complete dataset possible.

Theoretically, it is in the development of transactional web services that government effectiveness has been expected to improve (see the discussion in Sections 2.4.3 and 2.4.4). The “Transactional” cutpoint on the BC data series represents Socitm’s judgement about whether that ambition was being met. Practically, the scoring scheme discontinuity between 2006 and 2007 makes the choice of a base year for dummy variables difficult: the only common category is “Transactional”, but having the base category in the middle of an ordinal range makes interpretation of the regression coefficients difficult. Reduction to a binary variable makes interpretation much more straightforward.

In practice, the decision to use a binary independent variable turns out not to be problematic. To ensure that the results were not sensitive to this decision, versions of every model presented in this chapter were created with BC treated as a series of categorical dummy variables¹⁶. No substantive differences occur in either the model results or their interpretation, with the effects of the category dummies being small and somewhat noisy; an example is included as Model 7 in Table 26.

The data demonstrate serial correlation when Wooldridge’s test for serial correlation in fixed effect panels is used. This has no impact on the regression

14 Walsall Metropolitan Borough Council was awarded no service scores for 2002 (Audit Commission, 2002a, p.17).

15 The City of London was excluded from the original dataset.

16 It’s far from clear that the relationship between the BC grades is linear with the underlying variable of interest, so modelling them as a continuous linear variable would be inappropriate.

coefficients, but tends to result in over-optimistic standard errors; this is a further reason to eschew p-values in this case.

Finally, a companion model (Model 8, Table 26) was built using the raw CPA rating as a whole. The motivation for this is twofold: partly because the CPA measure is a better known but a less-close proxy for the underlying variable of interest and partly to provide some confidence that any effects being detected are not artifacts of the final CPA services aggregation process. In practice, raw CPA models have a slightly higher R^2 (likely due to the collinearity between the output variable and the CPA corporate assessment variable used as a control) but very similar coefficient estimates for our variables of interest.

The models provide extremely limited evidence that web engagement measured using the BC data series is associated with higher council performance. Although the lack of statistical significance is not relevant to this full-population dataset, the regression coefficients and therefore the detected effect size are effectively zero. In the main model, the coefficient of the BC Transactional or better term shows a trivial *reduction* in service quality is associated with improved web provision (a mere -0.001 standard deviations). By contrast, a one point increase in the corporate assessment rating is associated with a 0.39σ increase in the CPA services score.

Despite the lack of sampling error, an effect of this size is very likely to be statistical noise, caused by minor specification error or variations in the measurement process. The effects in this model are supported with those from the companion (where the effect is a trivially negative -0.02σ).

Overall, there is insufficient evidence here to support H10, and we must fail to reject the null hypothesis; engagement with the Internet is not associated with higher performance.

As the CPA and BC scores are both complex proxy measures for the underlying variables of interest (service quality and web development), a step back is needed. Do these results reasonably reflect what was actually going on in the local authorities concerned?

Notwithstanding the enthusiasm of the practitioner literature, this result is quite possible to square with the reality of local public administration. As with other putative service and management improvement techniques, there is nothing magic about the Internet which will result in automatic service improvement. If Internet-driven automation is applied to an existing disastrous service the outcome is likely to be the classic “disaster faster” (Margetts and Willcocks, 1993). Making the most of the opportunities that are afforded by use of the Internet requires skill and understanding, just as with other ways of improving local public services.

It is possible that there is a missing factor in the analysis, that of the *use* of services. It would be desirable to be able to control for which authorities in the model had good levels of citizen use of online services, for which there is a stronger theoretical argument that improvements would be seen. Such data are not readily available (see Section 3.3.4) which is disappointing as these effects might well be time-varying and therefore not adequately controlled by the fixed effects modelling.

Although there is no compelling reason to believe that results would be substantially different, this would be a valuable avenue for future research. The possibilities for scale-of-use measurement on a wide scale to supplement studies of e-government in public administration are discussed further in Chapter 7.

6.3 VALUE FOR MONEY

At first glance, efficiency gains generated by e-government activities are more likely to be found associated with higher performance than with overall council budget reductions. This is largely because local government in the period under study was a revenue- and spending-constrained sector, with government threats of council tax capping reducing budgets below those which would

have been freely set by the councils concerned¹⁷. It would be reasonable, therefore, to expect savings to be re-invested in service delivery. Having found no evidence of improved performance associated with Internet engagement in Section 6.2, however, it is important to test the secondary hypothesis: that improvements made were reflected by a reduction in the council's costs rather than an improvement in its quality.

6.3.1 *Cost effectiveness measures*

Measuring the cost-effectiveness of government organisations is difficult; unlike in a commercial organisation there is no single measure of success such as profit, and hence the output part of a cost-effectiveness measure is difficult to estimate. This is particularly true in local authorities. Some significant transactional parts of the modern state are structured along functional lines with a small number of commodified products: the Driver and Vehicle Licencing Agency, for example, handles large volumes of driving licence applications and vehicle registrations, both of which are readily measurable. Local government is not like this: there are several hundred discrete local government services (Wilson and Game, 2011, p.35) and the relative importance of them (necessary for creating a composite output measure) is politically determined and unclear.

There are some satisfactory approaches to this problem. Total Factor Productivity analyses build an index of outputs weighted by the cost of each service, which are then compared to the cost of inputs to create a measure (Dunleavy and Carrera, 2013, p.35). This effectively uses the organisation's own view about how it chooses to spend its money as a proxy for the importance of the different public services delivered.

Unfortunately, the cost of producing such an analysis for a local government would be high due to the diversity of the measures used for the outputs and

¹⁷ In any case, efficiency gains are difficult to remove as cash savings rather than increased output, as Willcocks (1992, p.248) notes.

the difficulties of disaggregating the costs of each service from the published council financial reports in order to get an accurate service weighting. To do this for each local authority would be a sizeable research project in itself. To repeat this annually to build a dataset suitable for comparison against the BC panel is quite beyond the scope of this research.

Fortunately, alternative approaches are available for local government researchers. Local government finance in England works on a multiple funding stream model: income for authorities at the time of the data is a combination of a local residential property tax (Council Tax, circa one third of revenue), an allocation of a share of a nationalised pool of business property tax (National Non-Domestic Rates, also circa one third of revenue), and a central government grant¹⁸.

In order to ensure that authorities are not unreasonably burdened by having a low local tax base, the central grant (known as Revenue Support Grant for most of the period under study) is calculated by a formula taking account of hundreds of factors including the volume of different services required in each local authority area, local cost structures, and the extra costs associated with sparseness of population, deprivation etc. (see, for example, DCLG, 2012). The original principle behind this system of grant was based upon a Standard Spending Assessment, which is the amount that a given authority needed in its individual circumstances to deliver a standard level of services (Department of the Environment, 1990).

This approach has continued to operate. It is possible, therefore, to assess formula-derived need against the actual local authority finance outturn numbers to generate a measure of efficiency against standard needs. By factoring in a measure of service performance, such as that calculated above as part

18 David Cameron's Conservative government has subsequently started to shift the balance of taxation by reducing central grants and 'repatriating' non-domestic rates, though not by restoring local discretion over tax rates.

of the CPA process, a model of cost-effectiveness can be generated with the following attractive features:

- The output assessment is based upon the expressed service priorities of central government, factored as such into the CPA process;
- The input assessment is equalised for the variations between different types and circumstances of local authorities (Revelli, 2010), again using the government's extensive indicator sets and view of cost drivers; and
- Comprehensive data are available for all local authorities, across multiple years.

The principal measure here is spending in excess of standard. The data used here are those obtained by Revelli from CIPFA data and used to produce his own assessment of local government efficiency (Revelli, 2010). Data are available for 2002–2007, one year less than the CPA data used above. Figure 22 shows the distribution of this measure by year. For comparison, Figure 23 shows the raw level of spending by authorities over the same period – one of generally rising public expenditure in the UK.

6.3.2 *Regression modelling of cost effectiveness and BC results*

A fixed effects panel data model (Model 9, Table 27) was fitted on the same basis as the performance models in Section 6.2.1. Crucially, the service scores from the CPA dataset are included as an independent control variable; as a result, variation in the dependent cost variable reflects cost-effectiveness rather than simply changes in performance. Again, the council's corporate assessment scores and year term were also included as control variables.

As a sanity check on the first model, a companion model (Model 10, Table 27) was again fitted, using spend per capita with no standardisation for local need. This is, in principle, a less effective measure of our variable of interest, because it takes no account of inter-year changes in the situation of each local authority (such as changes in poverty level and population) which are handled by the

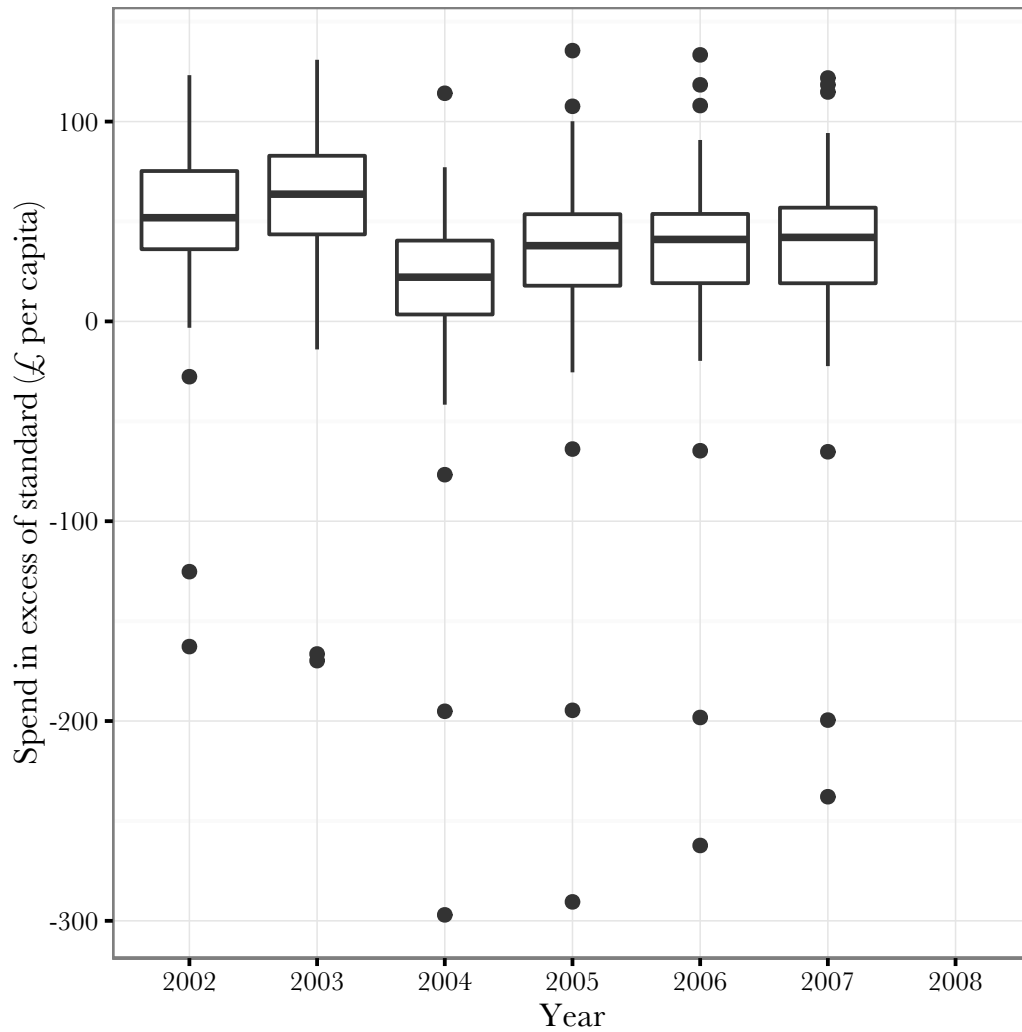


Figure 22: Spend in excess of standard, by year

dependent variable in the main model. Nevertheless, the panel approach does in principle control for variation between authorities which does not change over the period of the study; Model 10 therefore allows us to test a less-sophisticated set of assumptions as a hedge against problems in the methods the government used to calculate the standard spending assessments.

Once again, the results of the analysis provide little support for the thesis that improvements in web performance can transform the cost-effectiveness of delivering public services. Although the model fits are excellent (R^2 is 0.95 in the main model, 0.97 in the companion) the majority of the variation is

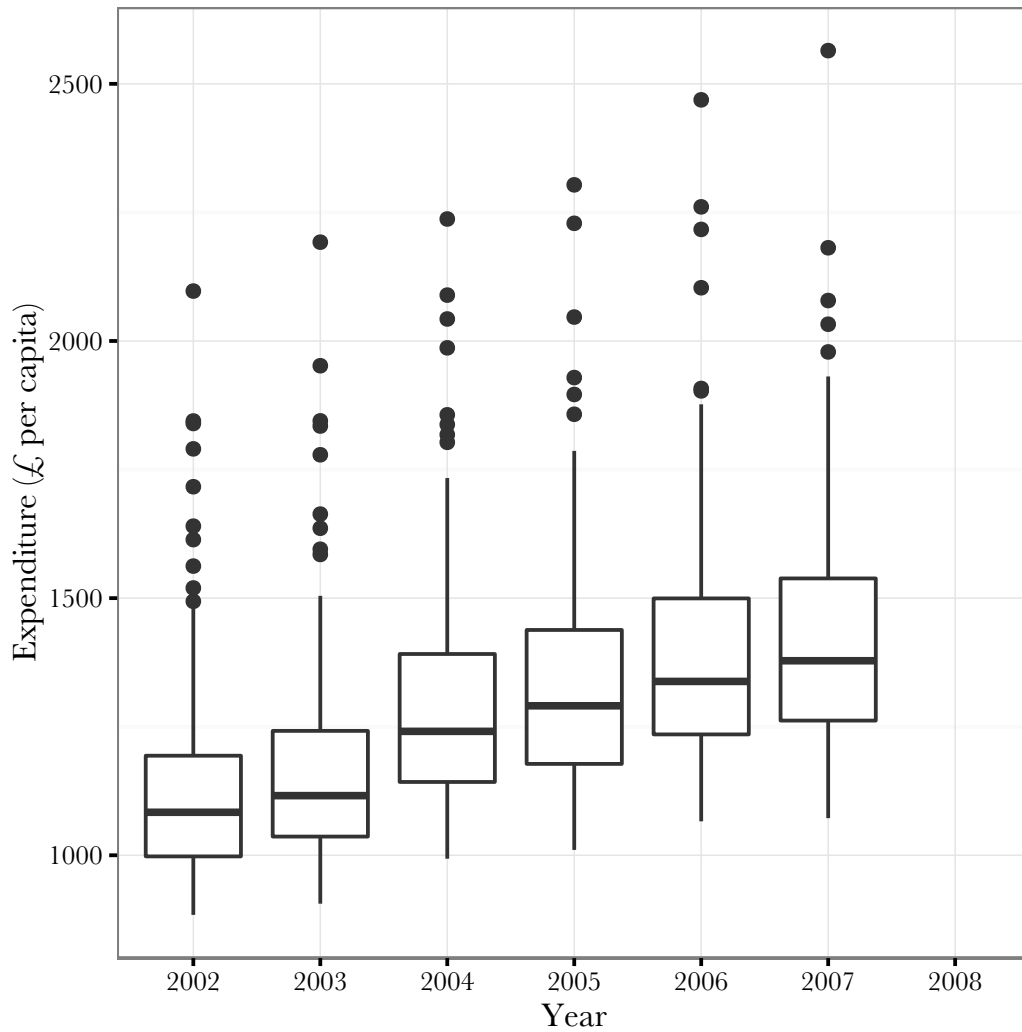


Figure 23: Expenditure, by year

associated with the year control and on the omitted council dummy variables, not the BC variables of interest¹⁹.

There is an effect here, in the direction theorised, but the magnitude of it is tiny: it is £1 per capita per annum, less than 0.1% of council spending. As with the previous section, where the effect was in the opposite direction to that predicted, this is far too small to be confident that the result is a real one,

¹⁹ This very high degree of model fit is not as surprising as it seems. The two major constraints on local government finances are the annual government funding settlement and Council Tax increase cap (which are common across the sector and therefore reflected in the year dummies) and the level of the previous year's revenues (which are reflected in the council dummies).

	Excess spending, £/cap (9)	Spending, £/cap (10)
BC Transactional or better	-1.028 (1.462)	-8.412 (6.314)
2003	6.874*** (1.334)	42.145*** (5.761)
2004	-33.157*** (1.453)	165.096*** (6.276)
2005	-19.219*** (1.358)	214.034*** (5.866)
2006	-16.814*** (1.470)	264.351*** (6.349)
2007	-15.379*** (1.617)	305.282*** (6.985)
CPA corporate assessment	0.363 (1.409)	2.258 (6.086)
CPA services (normalised)	-0.615 (0.749)	2.866 (3.233)
Constant	33.298*** (5.739)	1,425.552*** (24.789)
N	887	887
R ²	0.945	0.971
Adjusted R ²	0.934	0.965
Residual Std. Error	10.954	47.314
F Statistic	81.425***	157.179***

* p < .1; ** p < .05; *** p < .01

Council-specific dummies omitted. Base year is 2002.

Table 27: Regression models: council spending, CPA, and BC, 2002–2007

as it is well within the bounds of what would be caused by measurement error or model specification issues.

Increasing delivery of Internet services comes at an implementation cost. One simple interpretation of these results is that Internet engagement is associated with some cost-savings in service delivery (as the cost of provision doesn't rise, despite the implementation costs) but not sufficiently to result in efficiencies overall. This result would be consistent with earlier work on infor-

matisation reporting that the benefits of such work were often overestimated and the gains difficult to realise (Strassmann, 1990; Willcocks, 1992).

In any case, there is no evidence of substantial cost savings being made through the use of better web provision in English local government in the period modelled. There is insufficient evidence to accept Hypothesis H11 and we must therefore fail to reject the null hypothesis that Internet engagement is not negatively associated with service cost. Although this is formally a null result, the full-population nature of the sample and the highly-sensitive panel design of the model gives us positive evidence that the null hypothesis reflects the underlying truth, above and beyond the simple p-value driven rejection of H11.

6.4 CITIZEN SATISFACTION

In light of the findings of the two previous sections that there is no discernable relationship between Internet engagement and either performance or cost-effectiveness, a question arises as to whether there is another reason that councils have persisted in developing an Internet presence despite this lack of evidence of utility.

As organisations accountable through the ballot box to local citizens, there is an additional goal which local authorities pursue beyond the obvious one of service improvement: the perception of the council in the eyes of citizens. This is seen most clearly by the aftermath of an influential report by Ipsos MORI, arguing that citizens were more satisfied with their local authority if they received an A–Z booklet setting out the range of services provided by the council and giving contact details for each service area, which later became part of a reputation campaign by the Local Government Association (Local Government Association, 2006).

With the Internet, too, it is plausible that the motivation for councils to persist with websites is not to improve their services per se, but rather to

present a public image of a modern and effective authority. If this motivation is well-founded, an effect should be found in a greater favourable rating by citizens of local authorities with a developed web presence.

6.4.1 *Best Value Performance Indicators and citizen satisfaction*

Best Value was a performance improvement method developed by local authorities in the late 1990s and pushed nationally by the Labour government in the early 2000s. It was defined in the 1999 Local Government Act:

“A Best Value authority must make arrangements to secure continuous improvement in the way in which its functions are exercised, having regard to a combination of economy, efficiency and effectiveness.” *Local Government Act 1999*, §3(1)

The underlying concept for driving improvement was that of the “4 Cs”: challenge, compare, compete, and consult. Each aspect of the framework combined to drive self-improvement by local authorities. The approach was firmly grounded in New Public Management (NPM), but this particular expression of it was more holistic than most, looking for collaborative improvements and internal change driven on self-challenge, rather than simply marketisation and contracting out.

As part of the “Compare” part of the framework, the government specified a huge list of Best Value Performance Indicators (BVPIs), a nationally standardised set of performance measures which would allow comparability between different councils, identify areas of good and bad performance, and be usable as a framework with which to drive up service performance (DETR, 1999).

The system was administered by local government’s performance and spending watchdog, the Audit Commission, and indicators were generally to be produced annually. The system persisted from 2000/01 until 2007/08, at which time the set of indicators was replaced by the smaller National Indicator Set. Best Value as a concept had been effectively defunct in local government for several years previously, replaced by a narrower focus on efficiency driven

by the Gershon review (Gershon, 2004). The definition of the indicators was fairly stable over the lifetime of the BVPI set, though new indicators were sometimes added and others removed.

One of the BVPIs was BVPI 3, measuring overall satisfaction with the council. This was measured by survey, commissioned by the local authority concerned, but to national standards²⁰. This provided a headline measure of satisfaction, which could be compared with the large number of other satisfaction measures collected as part of the same surveys. As a good quality measure of council satisfaction, this is ideal for the purposes of this research. Unlike the CPA measures used in the two previous sections, BVPI 3 (like the BC measures) was collected for all principal authorities. As a result, data are available for district councils as well as unitaries, counties and London boroughs.

Unfortunately for our purposes, the cost of commissioning surveys meant that councils were only obligated to collect BVPI 3 and the other satisfaction indicators on a triennial basis, rather than annually with the other measures. As a result, data were collected for 2000/01 (Audit Commission, 2001b,a), 2003/04 (Audit Commission, 2004a) and 2006/07 (Audit Commission, 2007a) before the BVPI series was terminated.

Figure 24 summarises BVPI 3, showing two salient features: firstly that satisfaction markedly fell between 2001 and 2007, and secondly that satisfaction rates for shire districts appear to be notably higher in all three surveys.

20 The surveys were obliged to be based upon a probability sample, using either the Postcode Address File or the electoral register as the sampling frame. There was no central limitation on the survey method used, which could be face-to-face, by telephone or using another method. Responses were on a five-point Likert scale (Very satisfied – Fairly satisfied – Neither satisfied nor dissatisfied – Fairly dissatisfied – Very dissatisfied) and the outturn reported to the Audit Commission was the percentage answering “Fairly satisfied” or better. The operative question for BVPI 3 was “Taking everything into account, how satisfied or dissatisfied are you with the way the authority runs things?”, and followed a series of more detailed questions on individual services. The full survey specification is available (DETR, 1999, ch.14).

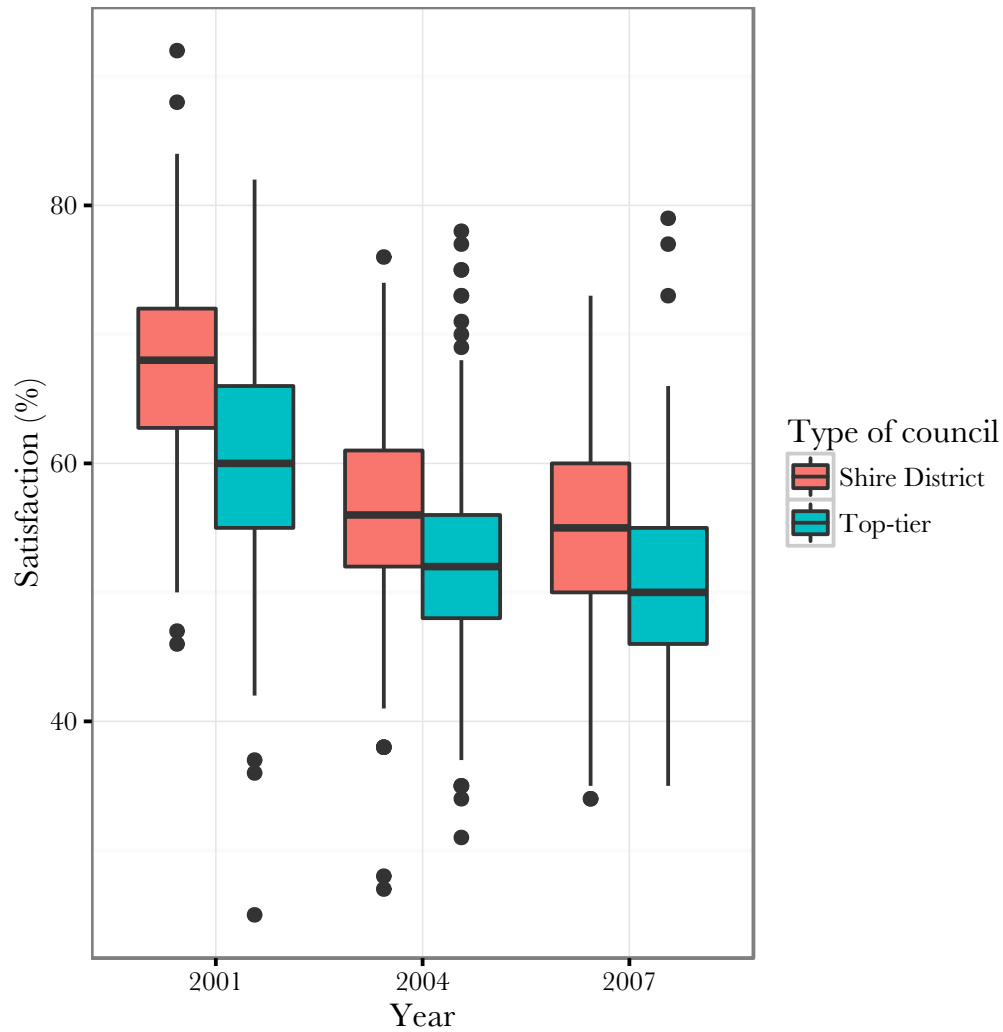


Figure 24: BVPI 3 citizen satisfaction, by year and type of council

6.4.2 Regression modelling of citizen satisfaction and BC results

Hypothesis H12: Local authority engagement with the internet, as measured by BC performance, is positively associated with citizen satisfaction, measured by BVPI 3 scores.

Three regression models have been fitted to assess H12, using the within-authority panel approach as with the previous models in this chapter. The dependent variable in both cases is BVPI 3 and the main independent variable in both is the BC grade. There are two key differences between the models, which reflect limitations of the data available. The first model (Model 11, Table 28)

uses the full population of councils including shire districts. Consequently, it has a much larger number of observations and covers both groups of English local authorities. Unfortunately, it is then impossible to use CPA service quality or corporate assessment ratings as control variables, as this data is not available for shire districts. The second model (Model 12, Table 28) is a replication of the first, but restricted to top-tier authorities only, as a comparator for the third model (Model 13, Table 28) which is also top tier authorities only – but which is more robust as it is possible to control for factors (quality and cost of service) which seem intuitively very likely to have an influence on citizen satisfaction. By fitting all three models, a fuller picture can be obtained.

Unfortunately, the 2001 data is not readily usable in this analysis: as it lies outwith the period of CPA it is not possible to use it alongside CPA-based control variables. Furthermore, there was only one “Transactional”-rated council in the 2001 survey²¹ and there is therefore insufficient data for the key independent variable.

Model 11 in many ways reflects the service quality and cost models. There is essentially no relationship between the judgement of transactional or better for an authority’s website and satisfaction with that authority – the estimate is only three-hundredths of one percentage point in satisfaction. There is, again, statistically-significant variation over time.

Model 12 is equivalent to Model 11, but contains only the top-tier authorities. Results are similar, except that the negligible modelled association of having a transactional website becomes marginally negative when the shire districts are excluded²².

That Model 13, top-tier authorities with controls, is a marginally better fit for the data than Model 12 indicates that the controls add value to the model. The effect of these is to marginally increase the negative association

21 Tameside Metropolitan Borough Council

22 The year term is also no longer statistically significant due to the reduced N , which is of marginal relevance for comparative purposes.

	BVPI 3, All (11)	BVPI 3, Top tier (12)	BVPI 3, Top tier (13)
bc Transactional or better	0.035 (0.793)	-0.696 (1.304)	-0.958 (1.311)
2007	-1.348*** (0.392)	-1.236 (0.770)	-0.472 (1.357)
CPA corporate assessment			2.493* (1.499)
CPA services (normalised)			0.284 (0.745)
Excess spend (£/cap)			-0.070 (0.058)
Constant	68.656*** (3.408)	55.314*** (4.030)	49.953*** (5.429)
N	773	298	296
R ²	0.805	0.752	0.740
Adjusted R ²	0.608	0.500	0.463
Residual Std. Error	4.794	5.471	5.453
F Statistic	4.080***	2.979***	2.674***

* p < .1; ** p < .05; *** p < .01

Council-specific dummies omitted. Base year is 2004.

Table 28: Regression models: citizen satisfaction and bc, 2004/2007

between website quality and satisfaction. As theorised, there is a weak positive relationship between quality of service and satisfaction and between the corporate assessment and satisfaction. Again as expected, there is a weak negative association between spending in excess of standard and satisfaction (every extra one pound per capita is associated with a 0.07 percentage point reduction in satisfaction). None of these results reach the 5% level of statistical significance on a conventional t-test.

Once again, effect sizes here are not large. Coupled with the underlying surveys being subject to sampling error, it implies a need for cautious evaluation. Taking the results at their highest, less than one percentage point of the variation in satisfaction rates is explained by councils' status as having transactional websites, and that in the other direction than that hypothesised.

Given that each individual BVPI 3 data point is itself based on a survey with a $\pm 2\text{--}3\%$ margin of error, this should be treated as noise. In any case, the effect hypothesised is certainly not present in these data. We must fail to reject the null hypothesis and discard H12.

6.5 WHAT EXPLAINS THESE RELATIONSHIPS?

The most obvious explanation for these (lack of) relationships is that improvements in web engagement as measured here are peripheral to the improved delivery of local government services.

With the exception of the online information service itself, there are reasons identified in Section 2.4.7 why local government in particular would be less amenable to service and cost benefits through informatisation than other government services. The obvious interpretation of the relationships (that greater engagement with the Internet is not associated with better services or efficiency) seems fairly robust. There is nothing in the results of the modelling which leads to a different conclusion (the three main models in this chapter are printed in full, including the complete suite of council-specific dummies, in Table 32 in Appendix A).

The results are not the consequence of serial correlation, nor do the results change if we assume the positive effects of web development will be lagged: a set of models were fitted including up to four years of lagged dependent variables²³. No substantive differences emerged (Models 14, 15, and 16 in Table 29).

It is important to be clear about what the results in this chapter show, and what they don't. The data show no associations in the expected direc-

²³ These added lagged versions of the dependent variable as an independent variable (Diebold, 2014). This limits the number of years in the analysis substantially, by requiring prior time periods to be present.

	CPA services (14)	Excess spending (£/cap) (15)	Satisfaction (%) (16)
BC Transactional or better	-0.110 (0.078)	-0.910 (0.749)	-2.140 (1.493)
BC T or better, 1 year lag	-0.059 (0.086)	-0.998 (0.976)	3.529* (1.918)
BC T or better, 2 year lag	-0.103 (0.101)	-0.469 (1.062)	1.928 (2.569)
BC T or better, 3 year lag	-0.100 (0.123)	-0.580 (1.271)	-2.327 (2.783)
BC T or better, 4 year lag	-0.098 (0.145)	2.704 (1.933)	
2006	0.344*** (0.060)	2.410*** (0.493)	
2007	0.598*** (0.066)	3.915*** (0.633)	-0.973 (1.422)
2008	0.961*** (0.081)		
CPA corporate assessment	0.066 (0.081)	0.508 (1.122)	2.295 (1.488)
CPA services (normalised)		-0.565 (0.476)	0.231 (0.737)
Excess spending (£/capita)			-0.062 (0.058)
Constant	-0.756** (0.348)	15.301*** (3.885)	50.195*** (5.594)
N	596	444	296
R ²	0.797	0.995	0.752
Adjusted R ²	0.724	0.992	0.477
Residual Std. Error	0.499	3.753	5.383
F Statistic	10.945***	359.079***	2.734***

* p < .1; ** p < .05; *** p < .01

Council-specific dummies omitted. Base year is 2005.

Table 29: Regression models: BC and CPA, including lagged dependent variables

tions between Internet engagement and either service performance or cost-effectiveness.

This has not been a full-blown causal analysis. In the event that a relationship with a reasonable effect size had been demonstrated, further work would still have been needed to demonstrate that the association was not due to the influence of a common causal agent:

“If we observe change in X to be related to change in $\mathcal{I}$, in most situations we cannot exclude the possibility that the two processes are independent, conditional on an unobserved, time-varying $\mathcal{Z}$. [...] Thus, although panel data represents a considerable improvement on cross-sectional data for understanding dynamic processes, it is important to note its fundamental limitation in unequivocally identifying causal relationships.” Berrington, Smith and Sturgis (2006, p.6)

Although the census approach to data collection means that the results are untainted by sampling error, there are still causes of bias and uncertainty that exist, and which lead to a more cautious conclusion than would otherwise be drawn.

Firstly, the main risk in over-drawing conclusions from the findings above is from specification bias. The main difference between the positive BC and CPA gamma analysis (Figure 21) and the flat final panel regression in Model 6 is the impact of the control variables added, especially the council-specific dummies. Although obvious confounding variables have been controlled for, it's always possible that there are other factors at work which have not been considered, which might have adjusted the results in either direction. Because of the nature of panel modelling, any such confounding variables would have to be time-varying, but that leaves plenty of possibilities.

Secondly, full-population panel data models do not make the dependent variable immune from measurement errors and biases. In particular, this research is based largely upon government judgements of service quality and cash requirements; such judgements are neither politically neutral nor externally quality-controlled, and there are significant incentives for authorities to attempt to gain these measurements in pursuit of additional funding or administrative freedoms. Indeed, the basic interpretation of the significance of the year variables in the regressions is that councils have got better at gaining higher dependent variables as time has gone on. Whether that actually reflects greater performance is, in principle, impossible to prove using this dataset alone.

The key question is whether the lack of link found here can be reasonably attributed to the acknowledged weaknesses of the measures, rather than a lack of underlying relationships of the underlying variables of interest. I conclude that the results support the latter interpretation. Although there is certainly gaming and variability in the measures, any time-invariant biases resulting from this within individual authorities are controlled for by the modelling process. The idea that any time-varying biases would combine to perfectly eliminate an actually existing relationship is less theoretically plausible than the simpler suggestion that no relationship is present.

Overall, the measures are necessarily imperfect, but they are of sufficient quality that they would be expected to detect an effect if one were present. Although further research with alternative measures would be worthwhile, this research gives strong evidence of a lack of association.

6.6 CONCLUSIONS

The panel dataset assembled in this chapter allows for a fairly direct analysis of the relationship between web engagement, service performance, cost, and citizen satisfaction with good-quality proxy measures available for all variables of interest. The full-population approach further allows the analysis to be sensitive, as it avoids sampling error.

Despite this, there is essentially no association between the BC scores and CPA measures of service performance, spending in excess of standard assessment, or BVPI 3 citizen satisfaction. This calls into question utopian models of Internet-driven change and supports instead a more careful and nuanced approach.

Across a period of several years, across the whole population of local authorities, and given a background of central-government directed improvement in online services, no detectable improvement has been identified which is attributable to Internet work. In light of this, it is very difficult to see how the enthusiastic approaches of the early 2000s can be seen as still tenable.

As the BC data are at whole authority level, it has not been possible to break the results down to the levels of individual services. This means that there remains a good deal of uncertainty about the reasons for a lack of overall benefits. Given that the bulk of local services are concrete rather than informational, and rely heavily on face-to-face service delivery, it is completely plausible that the benefits would be limited. For more specific high-volume and transactional areas however, such as council tax administration and planning information, there is clear evidence from other work discussed in Chapter 2 that benefits would be expected.

Consequently, we must take a nuanced view of the scope of the results. They show that, on aggregate, the changes made on the web in the period covered by the data were not associated with improvements. They do not answer the related questions of whether this was because the web is ineffective per se as a means for improving service delivery, whether the problem arises from insufficient realisation of potential benefits (perhaps by not pushing channel shift as hard as it has been subsequent to the end of CPA data), whether the local government combination of expensive local face-to-face services and limited scope for online delivery means that we should simply expect small effect sizes, or whether there are other explanations.

There is an extensive literature on the tradition of over-promising and under-delivery in IT work, explored in Section 2.3.1, so it does seem plausible that the straightforward explanation – no effect was found because no effect is present – is the true one. In any case, the analysis was sufficiently sensitive that any strong effect should have been picked up.

Given the increasing size and scope of local government web provision this finding raises significant questions as to the way forward, and whether local government fits neatly into public sector improvement approaches based on increased informatisation. The implications for the Digital Era Governance (DEG) quasi-paradigm, and broader implications for the local government sector as a whole in light of all the foregoing findings, are discussed in Chapter 7.

DISCUSSION: THE INTERNET AND PUBLIC ADMINISTRATION

To what extent does the experience of
English local government fit the Digital
Era Governance quasi-paradigm?

What opportunities for future political
science research exist in large scale
analysis of web content?

7.1 INTRODUCTION

This chapter assesses the impact of the empirical findings in Chapters 4, 5, and 6 on three increasingly broad areas of academic work.

Section 7.2 analyses the empirical findings as a whole, and makes an argument that parallels the conclusions of *Information Technology in Government* (Margetts, 1999), that a critical approach needs to be taken to digital public administration. The empirical findings of this thesis support the kind of careful, critical-modernist, approach to digitalisation found in that work, despite the significant changes in context and approach that have happened since that work was published, a different focus (local in place of central government), and the centrality of the Internet in place of the more back-office approaches analysed then.

The findings of Margetts (1999), and this thesis, stand in some contrast with the later Digital Era Governance (DEG) thesis, which is more bullish on the benefits of digital technology and less focused on the heterogeneity of circumstances in which technology might be used and the challenges of successful implementation. Section 7.3 looks at DEG itself, as an exemplar of digitally-focused theories of public administration, aiming to replace New Public Management (NPM) as the default toolkit for governments, and assesses its prospects.

Public administration more broadly is the focus of Section 7.4. This section explores explanations within the broad public administration field for the low-key results found for the efficiency benefits of Internet use, and discusses where questions of the Internet and digitalisation may go from here.

Finally, Section 7.5 makes similar supportive but critical arguments in respect of the use of machine learning and other full-population quantitative approaches in political science. The benefits of some of these approaches have been demonstrated in this research, but they come with significant challenges for their effective use within political science. Some of these challenges, which relate to practicalities, ethics, and interpretability, are discussed and solutions proposed.

Finally, it draws lessons for the future use of full-population web crawl data by political science as a whole, and makes recommendations for future research directions.

7.2 EMPIRICAL FINDINGS: AN OVERALL PICTURE

This research has set out to answer an overarching research question, by looking in depth at a number of discrete sub-questions:

Research question:

Sub-question 1: What is the structure, shape, and connectedness of the English local government Internet presence and what

are the relationships between different bodies? What can this tell us about the structure of government in the digital age?

- Sub-question 2: What is the content of the English local government web? What can this tell us about local government and its approach to e-government?
- Sub-question 3: What is the relationship between council administrative performance, cost-effectiveness and use of the Internet? What explains this relationship?
- Sub-question 4: To what extent does the experience of English local government fit the Digital Era Governance quasi-paradigm?
- Sub-question 5: What opportunities for future political science research exist in large scale analysis of web content?

The first three largely empirical sub-questions have been addressed in Chapters 4–6; the latter two more theoretical ones will be addressed by this chapter. This section aims to draw together the empirical conclusions and to assess alternative explanations for them.

The key finding from the analysis of both government web structure in Chapter 4 and content in Chapter 5 is that the local government web is complex and heterogeneous. It is also remarkably lacking in structure, horizontal linkages, and content similarities between local authorities.

The explanation for these findings which arises from the literature review is that local government is, in fact, locally diverse and heterogeneous and that the online variations are reflections of this.

There is some merit in an alternative proposition, though, which is that these online findings don't speak so much to the structure of local government itself so much as they do to a disconnect between local government structure and the structure of the local government web.

A key insight underlying webometrics is that online structures and relationships reflect their offline counterparts, discussed in Section 3.3.3. This insight has proved valuable for studies of university website organisation (such as Hale, Yasseri et al., 2014) as well as looking at online analogues of the journal citation networks which motivate bibliometrics.

The findings from this research are somewhat muddier, though. That offline geography is reflected in online network structures was confirmed by the analyses of linking propensity between neighbouring local authorities in Section 4.3.3, as well as the emergence of weak geographic clustering in the Infomap clustering analysis in Section 4.3.2.

The attempts to extract meaningful extra-geographic clusters from the online link network in Section 4.3.2, however, did not result in a neat hierarchy as might be expected from studying pure web phenomena, but a remarkably flat and uninformative grouping reflecting little information contained within the link content at all. A similar pattern emerges from the grouping of much of the local government web content into only two categories by the Latent Dirichlet Allocation (LDA) process in Section 5.3.4 (though this may reflect difficulties in the modelling process more than the clustering finding does).

Data collection challenges, discussed in Section 7.5.1 below, also make the drawing of direct links between online and offline behaviour more challenging. More difficult to set aside are the empirical findings which focus on the digital directly, rather than as a means of understanding the offline world.

The findings in Sections 5.3.2 and 5.3.3 that much of the local government web is not service focused and the great majority of it is not transactional, for example, are robust. Although there are limitations that arise from the choices made in the web crawling process, discussed in Chapter 3 and returned to below, the effects here are sufficiently stark to be robust to these variations.

Similarly strong are the findings in Chapter 6 that greater engagement with the Internet was not, in English local authorities over the period stud-

ied, associated with greater cost effectiveness, quality of service, or public satisfaction.

The data being analysed are necessarily imperfect instruments which do not perfectly reflect the underlying phenomena of interest, as with all other political science analysis. Nevertheless, the Comprehensive Performance Assessment (CPA) service quality and Better Connected (BC) web service assessments are as robust as we might reasonably hope for. That we have full-population coverage on a panel basis over multiple years also reduces a major source of error, by allowing a fixed-effects approach to control for the time-invariant heterogeneity which is inevitably present when dealing with comparative datasets. This means that the analysis is also robust to misspecification errors.

The alternative explanation in this case is that there *is* some kind of association between web use and service outcomes, but that the uncertainties in the model have conspired to produce a false null result for a full panel series. That any time-varying biases would combine to perfectly eliminate an actually existing relationship is less theoretically plausible than the simpler suggestion that no relationship is present. The more straightforward explanation, that use of the Internet had no measurable effect on service outcomes over the period, is not only supported by a significant body of literature on the under-performance of IT projects as practically implemented, but also by an understanding of the local context of local governments: the most important, and expensive, of the services provided are professional, local, and face-to-face. Consequently, that there are no visible benefits from web technology is a reasonable finding.

The picture drawn by these empirical findings is not the modernist technoutopianism common from the practitioner literature (Section 2.4.3). There is no sign here of technology-driven homogeneity, rather signs of individual organisations pursuing their own goals using technology, and rooted in their own geographical context. As identified by Margetts (1999) in a radically different context, the evidence points towards the value of taking “a more

careful and realistic approach and focus on government, rather than society in general.” (Margetts, 1999, p.169).

Given that these empirical results are somewhat underwhelming adverts for the transformative possibilities of Internet technology, where does that leave digitally-focused public administration theories such as DEG?

7.3 DIGITAL ERA LOCAL GOVERNANCE?

This section begins the process of developing public administration theory to build upon DEG. The particular contribution will be to break down the idea of “digital” as an overarching theoretical construct and look with more scepticism at the contribution that online approaches can make to public administration.

7.3.1 *Digital Era Governance in local practice*

As described in Section 2.2.2, DEG is a mix of the empirical and the normative, described by its authors as an “ideal type” or as a “quasi-paradigm”.

The DEG quasi-paradigm is based on a significant amount of empirical work. Comparative study of several countries’ IT development and contracting practice, together with fuzzy-set Qualitative Comparative Analysis (fsQCA) analysis of various aspects of the procurement situation in those countries underpins the model (an evaluation and critique of this work is in Section 2.2.2). It all builds upon previous theoretical work in the field – most obviously upon the weaknesses of the NPM model it seeks to supplant, but also upon Weberian approaches and upon other theorists.

Nevertheless, there is scope for a substantial broadening and deepening of the empirical and theoretical base for DEG, and this study has attempted to do just that. By looking at a different part of the state (local government) an attempt can be made to validate the general claims of DEG. By offering new

theoretical insights, the limits of what has been claimed by DEG can be further tested.

Implicit in the claim that digitalisation is a suitable normative goal for government, though, is the idea that a digitalised government would be in some way superior to the previous non-digitalised one. This seems intuitively unobjectionable; after all, there are plausible mechanisms for the greater use of technology leading to efficiencies and improvements, and there have been significant changes made in other fields of endeavour as a result of greater use of the Internet.

The empirical findings of this research suggest that that implicit claim may not be as robust as it appears, and therefore undermine the notion that digitalisation is normative. As there seems little justification for this notion beyond the beneficial practical effects that are suggested, this section argues that digitalisation may have been over-sold and should be placed on a more prosaic but defensible footing. In doing so, it contributes to a long line of empirical and theoretical work in informatisation which is sceptical of broad claims of transformation and refocuses attention on the small, the measurable, and the concrete (see Section 2.3.1).

The empirical arguments for DEG as an existing and effective way of improving government are missing two substantial components. Firstly, a detailed picture of how digital services are being provided by governments. Some of this gap is filled by this research, by providing some concrete data to support an understanding of the local portion of the government web.

The second missing component is supporting evidence for service transformation by using digital approaches. This really needs two pieces of evidence: firstly, some showing that digital approaches are actually linked with improved services (Chapter 6 provides some strong counter-evidence); secondly, some evidence that the new digital services provided by governments as part of their improvements are actually used by the public. This will be much harder to obtain, as usage data is closely held by websites and getting a complete picture

may be impossible. Nevertheless, a friendly approach via Socitm and the Government Digital Service (GDS) may provide usage data to cross-reference with the publication data provided in this research¹, or alternatively a broad-brush application of the Freedom of Information Act to server logs may provide sufficient information.

There is little evidence that DEG is a good model to explain the present state of digital public administration in a local context.

The evidence base on which the empirical analyses of DEG rely was gathered from studies of national-level administrations. There are reasons to suspect that the mechanisms identified as part of both the NPM and DEG waves of service reform have played out rather differently in local administrations than they have in central government. In order to develop a fuller understanding of technology-supported developments in public administration, therefore, it's necessary to look at how the differences identified in Chapters 4 and 5, and the findings of Chapter 6, might affect our understanding of a truly local digital era governance framework.

The empirical evidence base on which DEG is built is relatively thin. The first part of Dunleavy et al. (2006a) is an analysis of the properties of IT contracting arrangements, political structures and performance in IT service performance. The most interesting observation from our point of view is that NPM approaches do not appear, in the set of countries studied, to be associated with better government performance.

There is also continued evidence for the perceived importance of some kind of digital component in governments' actual service improvement activities. As discussed in Section 2.2, much work has been ongoing via the GDS in British central government, and Socitm has been supporting and tracking extensive practical e-government work in British local government.

¹ The Socitm Website Performance Service, in particular, could provide a longitudinal panel data source for usage by citizens, as well as user satisfaction data.

The jump from there to the normative part of the DEG ideal type as a model for public administration reform is empirically shakier. In particular, in order to be recommended there should be some evidence that the approach produces positive results in practice. This is not to say that the effectiveness of digitalisation is the only component of DEG, as “[t]he DEG model stresses the importance of digital change, but not in any determinant way, rather assigning equal importance to two other components—organizational reintegration within governments (which fosters disintermediation), and needs-based holism (re-unifying government services around client groups, instead of ‘business processes’)” (Margetts and Dunleavy, 2013, p.2). Nevertheless, without demonstrable benefits from the digital part of the DEG agenda then the value of adopting such a digitally-focused normative model seems questionable.

7.3.2 *Local governance in digital context*

The focus of this section will be to examine the evidence provided for DEG as a model of government reform, in light of the analysis of comparative local (rather than national) government, to see if the policy advice offered seems to have utility beyond the set of seven national governments which inspired it.

It is a trite statement within the public administration field, but local government is different. Local governments in England are not known for billion-pound IT procurements, largely because local budgets and services are not large enough to support them. Although there has been an increasing drive since the late 2000s for back-office services to be “joined up”, including by joint working between differing authorities, there are no single sector-wide IT system procurements as there are in large single-purpose agencies such as Her Majesty’s Revenue and Customs (HMRC) and the Department of Health. There are reasonable grounds, therefore, for believing that English local government IT procurement strategies will be different from those uncovered for central government by Dunleavy et al. (2006a).

Many of the prominent information systems in local government bodies are specialised tools, marketed more or less “off the shelf” to local government customers. In the web crawl analysed in Chapters 4 and 5, for example, the democratic support system ModernGov appeared frequently as the mechanism for local authorities to host their committee agendas and minutes. Similar systems, such as Arena for libraries and Idox for planning applications, exist in other services.

This highlights one of the distinctive features from the IT contracting perspective which differentiates the world of local government from that of central government: a lack of monopsony. When building an IT system to support the Inland Revenue, one of the examples given by Dunleavy et al. (2006a), there is a single customer with a highly bespoke set of requirements. Coupled with the reduction of IT expertise in government caused by that outsourcing, this is a recipe for contractor-led bespoke development.

For local government needs, by contrast, the laying down of national service standards by central government will reduce (if not eliminate) the amount of customisation required. While the procurement process for complex transactional systems will inevitably not reduce to the straightforward purchase of boxed software, it at least means that there is some competitive pressure of the supplier.

7.3.3 *Transformation, digitalisation and public services*

There is a substantial disconnect between the more utopian rhetoric surrounding digital service improvement and the work of actually doing it. As has long been recognised in both academic (Meijer and Zouridis, 2004) and practitioner (Bracken, 2014) writing, significant gains can be made from information technology by the improving of processes but not simply by automation. Indeed, it could not be any other way: to add an additional website dimension to an organisation while allowing the existing information gathering and service

processes to continue unchanged will simply add the cost of a website onto the existing service costs. To automate a broken process simply brings “disaster faster” (Margetts and Willcocks, 1993).

In light of this, it’s necessary to reconsider both the rhetoric surrounding the development of digital channels and some of the promises made by both DEG and enthusiastic digitalists as to the possibilities for Amazon-style transformation of public services through the Internet.

One key claim for DEG in its most recent revision was that trends were leading towards a digitally-focused reintegration of services:

“For the future, the most critical trend for reintegration (particularly for states under austerity pressures) is a move towards an ‘intelligent centre / devolved delivery’ (IC + DD) design in the allocation of functions across tiers of government.” (Margetts and Dunleavy, 2013, p.8)

There is little evidence of that in the UK. There is still no reintegrated local government component to the centralised (and holistic) government service delivery website gov.uk; this is still a floated and contentious ambition rather than something which is actually being implemented. Likewise, there is no visible drive to see local government as a devolved delivery component attached to a (presumably central government-based) informed centre. Rather, central government seems to be continuing its NPM-era drive to reduce the scope of local government services by disaggregation: the latest announcement is that all local-government controlled schools are to be transformed into independent “academies”, with funding and nominal oversight from Whitehall, but significantly increased levels of legal autonomy for the academies’ boards and headteachers (Cook, 2016). The arguments used to justify this are themselves straight out of the “competition” element of NPM: “[ministers] want to break what they see as a monopoly of local authorities, to offer parents a choice between different brands of education from different chains of academy schools.” (Jeffreys, 2016).

There is some budget-driven activity towards the merging of local government back office services, but this is in the tradition of NPM-driven spinning out of tradeable services after the introduction of compulsory competitive tendering. The desire to share back office services and offer those services to other authorities is just as in line with NPM's ambition for competition as it is with any DEG notion of holism. Most of these services, after all, were barely public facing in any case².

The second-wave DEG reintegration of tax and benefits systems using real-time data has, in fact, fallen prey to the traditional curse of badly-specified and procured IT systems identified in the original DEG book: the system supporting Universal Credit is now late, over budget, and too inflexible for the new unified system to be rolled out to most claimants.

Of the other holism elements identified as part of the second wave of DEG, the "big society" changes have been quietly backwatered, the social web and co-production of services are still in their infancy, open-book government is little further forward beyond the release of some non-democratically-critical open datasets (and the much-trailed desire of the government to restrict the Freedom of Information Act does not look promising for the future). The notion of a single citizen account is making some progress, with the GDS trialling a federated authentication system which will allow all of gov.uk to rely on non-government bodies using a common authentication platform, but that is not an impressive return on holism.

7.3.4 *Digitalisation: consequences for DEG and ways forward*

It is in the digitalisation gains where DEG really stands or falls, however, and the evidence from this research is not encouraging. The conclusion of Chapter 6

² The only visible change to most local government citizens of a shared finance department, for example, would be a different address and bank account number for their annual council tax bill. This isn't different in kind from the NPM-driven outsourcing of such services to Capita and others in the last 20 years.

was that there was no link between web development and cost, service quality, or citizen satisfaction in local government over a period of several years. This may or may not apply to central government, for which no comparable data are available, but it at least raises a significant question mark about the practical effectiveness of the digitalisation agenda.

In doing so, it bolsters the claims discussed in Chapter 2 that digitalisation is, in fact, rather marginal to improvements and efficiency changes in public administration:

“It continues to be a useful instrument of administrative and incremental change, but it is no more capable today of bringing about institutional change and administrative reform than it was fifteen years ago.” Kraemer and King (2003, p.5)

These findings link back into the long standing question of the effectiveness of using IT as a means for reducing costs or delivering improvements in services discussed in Chapter 2. In particular, the tension between the Kraemer and King view outlined above and the more positive view by Dunleavy et al. (2006a) seems to resolve for this study in favour of the sceptical but not dismissive attitude of the former. Yes, there are obvious benefits that can arise from the use of digital technology; these have been established since the early days of computing. Equally, there is an extensive tradition of negative or pessimistic findings when these benefits are measured in practice, especially across whole fields (see Section 2.3.1).

There is no evidence, from this study, that digitalisation has led to positive benefits for local government, at least for the period in the 2000s covered by the analysis in Chapter 6. There is also little evidence from Chapters 4 and 5 that the Internet is currently being used by local government in a way which is service-focused, transactional, or tending towards the kind of homogeneous delivery of efficiency gains hoped for by the Blair government or cyber-enthusiastic scholars.

There are a number of possible future scenarios for the development of local government on the web. To take it simply, these can be expressed as “more

digitalisation”, “stagnation in digitalisation”, or “a rollback in digitalisation”. These reflect the digitalisation aspects of the range of future DEG scenarios outlined by Margetts and Dunleavy (2013) and discussed in Section 2.2.2.

I have less confidence than Margetts and Dunleavy that DEG changes can be seen in practice. While DEG2 hypothesises a second and more advanced wave of DEG following the first, it is relatively difficult to see even first wave results in the data analysed here. As a result of the focus on local government, and also of the collection of large-scale quantitative datasets to see what is happening on the government web in aggregate, a different perspective emerges.

In particular, it is difficult to see where the concrete gains have come from in Internet-focused service improvements, particularly in light of the findings in Chapter 6 that no link between Internet use and effectiveness was to be found in the local government dataset.

Undoubtedly there have been successful examples of IT projects which have used the Internet to improve service delivery. The work of the GDS has been built on these, and there are extensive case studies in the e-government literature of success stories (see Section 2.3). Nevertheless, the amount of transactional service delivery in the English local government web is comparatively small (Section 5.3.3).

It is unlikely that, in the heavily cash-constrained world of English local government, that large e-government programs are going to be driven through with anything like the frequency which they were in the 2000s. Not only is management time likely to be heavily focused on managing service reductions, but the benefits of the optimism of the earlier period don't seem to have been delivered. With NPM propositions still being central to much of way local government is being managed (the continuing push for a thin commissioning role being the clearest current trend) this shows few signs of being likely to change³.

³ There is certainly continued reliance on technology as a means of reducing cost by responding to the austerity agenda, but this is likely to continue to manifest as NPM-influenced service

Furthermore, the GDS, the one agency within British central government with the mission (and the political enthusiasm) for digital service transformation, is not in a good place to take forward local government work. The Cabinet Office in which it sits has no direct responsibility for local government (which is managed through the Department for Communities and Local Government), unlike for the Civil Service as a whole.

The heterogeneity in local government which has always been a feature of its governance appears from the evidence of Chapters 4 and 5 to be reflected in its online provision too. This would make a centrally-driven attempt to rationalise local government web provision via the provision of a single local.gov.uk under the GDS even more challenging than its existing issues in rationalising central government (which, given the large number of different services they are attempting to co-ordinate already borders on the heroic).

Finally, we have the specifically local aspect of local government: that so much of its most expensive and valuable services are fundamentally face-to-face and local, and not really amenable to digitalisation (see Section 2.4.7).

I suggest, therefore, that hopes of extensive improvements in digitalisation in local government in the UK seem to reflect the kind of optimism-driven thinking that has defined the more modernist and hyper-modernist propositions of the past.

Given the extensive history of informatisation work failing to deliver substantial gains (as discussed in Section 2.3.1) a more plausible way forward for cash-deprived and risk-averse English local government managers is the very careful introduction of selected elements of technology on a very limited basis.

This looks more like “stagnation in digitalisation” than any other scenario. Given the empirical findings of this thesis, coupled with the centrality of digital approaches to DEG, it also raises doubts that the particular quasi-paradigm it offers is likely to be implemented in practice, at least at a local level. As such,

cost reductions rather than a DEG2-influenced focus on holistic service improvement and digitally-supported reintegration of services.

the future of DEG at local level seems likely to lie somewhere between the first and second of the scenarios outlined by Margetts and Dunleavy:

- “a systematic revival of NPM approaches, and abandonment of or crisis for the DEG quasi-paradigm. So far, this seems to have been limited to short-lived spasms in the UK and elsewhere;
- an austerity squeeze beginning a ‘pause’ of DEG changes that perhaps opens up into an escalating ‘de-modernization’ or permanent lag in the organization of public services. This scenario remains a live one; and
- no more than a short-lived austerity pause in DEG developments, with long-run changes returning to trend, reflecting the need to continuously boost government productivity and curtail relative price increases, and to keep the administration of liberal democratic governments in close touch with digital-related changes in business and civil society.”

(Margetts and Dunleavy, 2013, p.14)

As the authors of DEG have begun to go further than the DEG digitalisation thesis to develop Essentially Digital Governance (EDGE), a new heavily normative model (Dunleavy and Margetts, 2015), it seems that the debate on the effectiveness of IT as a means of delivering service improvements is not yet over. Although DEG is flagged as an “ideal type”, and therefore it isn’t fatal that it can’t be seen in full form in actual practice, the tension between the romantic forward view of DEG and EDGE and the underwhelming empirical results of digitalisation in practice found in this study and others seem difficult to reconcile.

7.4 IMPLICATIONS FOR PUBLIC ADMINISTRATION THEORY

Public administration remains, as outlined in Chapter 2, remarkably unfocused on the digital. The findings of this research show a nuanced digital picture. While the results of digitalisation may not be living up to the optimism of the hyper-modernists, the online dimension of public services is nevertheless huge: more than ten million pages of content in the UK alone, representing a major resource for citizens and a major amount of work by public authorities.

As such, public administration needs to take the Internet seriously. This section will discuss where questions of the Internet and digitalisation may go from here, within the public administration field.

Firstly, it explores explanations within the broad public administration field for the low-key results found for the efficiency benefits of Internet use. Secondly, it looks beyond both DEG and the techno-utopians to suggest ways forward for public administration theory.

7.4.1 *Explanations for the limited impact of the Internet on effectiveness*

It is obvious that some web technology developments offer the opportunity for gains for both customers and service providers. Indeed, the positive effects of greater Internet use in some areas are required for the regression results in Chapter 6 to be flat rather than strictly negative. As money, time, and management efforts have been spent on putting services online, if the activity was inherently unproductive then there would be negative association between the variables, which we have not seen.

Equally, it's perfectly clear that not all interventions will be equally successful. If nothing else, the rates of citizen use of different web applications vary, and the most-used services should provide better value for money as the cost of implementation is amortised over a greater number of interactions. So a more refined question is "which parts of the development of government services online have been productive, and which have not given a net return?"

There is clearly an opportunity for further study of the relationship between e-government activities and service quality, as discussed in Chapter 6. The result of this study are clear, that there is no relationship apparent between website developments and service quality, cost effectiveness, or citizen satisfaction. Nevertheless, this is a counter-intuitive result; the data examined here are only for England and there are a few infelicities in the data available. It would be beneficial, therefore, to repeat the analysis with data from other

jurisdictions, though a different design may be necessary as the UK is unusual in collecting such a detailed uniform set of performance and cost data from its local authorities (Revelli, 2010).

There are several plausible reasons why greater effort spent on Internet activity is not associated, overall, with local government performance. I will offer three potential explanations, which go some way towards explaining the apparent discrepancy and suggest new ways to conceptualise the relationship between the web and performance in a public administration context.

The first explanation is drawn from the existing literature around computerisation discussed in Section 2.3.1, and suggests that the technologies in general are simply prone to generate overestimation of their impacts on business performance. There is extensive evidence going back to the 1970s that computerisation did not, in fact, lead to the greater profitability of private sector enterprises (see the discussion in Section 2.3.1). Reasons demonstrated included a persistent underestimation of the costs of implementing technological change, leading to the approval of poor value-for-money projects and the realisation that in a competitive marketplace systemic reductions in cost can be captured as a consumer surplus rather than resulting in increased profitability. Although the public administration case is somewhat disjoint from the latter, it does seem at least plausible that customer expectations of web provision would rise with the development of the web as a governmental and commercial medium, so at least the lack of increase in satisfaction can be explained as a result of this form of “hedonic treadmill”.

A supporting observation is that the actual gains to public administration from customer-focused IT improvements may be difficult to measure – but that is, in fact, no reason to stop them being made:

“[For valence issues] a rhetorical stance that puts effort into conveying an image of shiny ‘modernity’ and purposive energy might be a logical response. If so, that can also explain why government makeovers in the name of cutting costs and improving quality should put more emphasis on ‘talking the talk’ than ‘walking the

walk’ in the sense of delivering concrete and readily measurable results.” (Hood and Dixon, 2015, pp.191–192)

The second explanation looks in more depth at the rather woolly empirical understandings behind the proposals for greater web use by enthusiastic theorists and government agencies. The quality of critical thought that might have justified such a wide-scale and expensive program of technology development was simply never there. Much of the expressed basis for the program was a belief in the transformative value of the new, straight from the hyper-modernists, coupled with a few positively-written case studies such as those put out by the e-government national projects in the mid-2000s.

The existence of disruptive private sector organisations was also held up as an example for change, while these have been most successful in reality in areas which have been ripe for reinvention through major logistics changes (e.g. Amazon), fundamental changes in information asymmetry (TripAdvisor) or disintermediation (e.g. the insurance comparison sites). For large chunks of local government work, such a proposition was always fanciful; the Internet does not fundamentally change the economics of caring for looked-after children, or of collecting and disposing of household waste, or of delivering school services. These are where the bulk of local government money and effort is spent and it’s therefore reasonable to expect little gain through greater informatisation. Where gains have been had, they haven’t generally been from the use of the public web, but through the use of web technologies to improve the internal information flows and efficiency of scheduling of internal staff.

The third potential explanation is around the relative scale of use of different implementations of government technology. The push to put all services online (see Section 2.4.3) precluded a focus on the most used parts of government, and indeed limited focus on the services most likely to benefit from the introduction of more interactive technology. It has therefore become something of an item of faith in recent years in GDS that much of the previous efforts to deliver (central) government web services were misdirected. The methodology of the BC surveys does correct for this in some measure, especially in later years, by

focusing on areas considered to be particularly important, but this judgement is inherently based upon Socitm's best level of understanding at the time rather than benefiting, as we do, from hindsight.

Likewise, the finding in Chapter 4 that the structure of government online is only loosely linked with the structure of government offline, can be taken as a data point in favour of this explanation. Put shortly, if government had been thoroughly transformed by the influence of informatisation, then the two structures should be much more similar. Likewise, findings from Chapter 5 that local governments' online presence is highly heterogeneous suggest that the influence of technological and centralising pressure towards homogeneity has been limited. If digital reforms had been thoroughgoing, different results would be visible.

I am cautious in advancing the third explanation, because it does have an air of "this time it will be different", which is perhaps the most expensive phrase in business. That the effect from the previous round of e-government work is so limited must generate a degree of cautious scepticism about the possibility that a renewed focus with the benefit of experience will necessarily lead to improvements, at least given the existence of the first two explanations.

A full and robust analysis of the relative strength of these explanations would require further research. It's clear that the official drive for e-government had a distinctively techno-utopian flavour, but a more nuanced picture would probably be available by doing in-depth historical interviews with those involved in e-government projects in local government during the period under study. A text analysis, using similar tools to those discussed in Section 3.5.1 but looking at the copious informal writings of practitioners, would also be informative.

A final observation is that the level of rigour surrounding the evaluation of Internet / e-government / digital approaches in local government has been poor. Like much of the wider public administration research discussed in Chapter 2, the major evaluation reports (such as Local Government Association, 2014)

have relied on an assertion of the positive benefits of technology. For all the reasons discussed above, that is no longer good enough – if it ever was.

7.4.2 *Public administration beyond DEG and hyper-modernist enthusiasm*

The effects of optimism-driven hyper-modernist enthusiasm were explored in Chapter 2 in both public administration (Section 2.3.1) and the policy environment in which English local government sits (Section 2.4.3). Given the lack of support found in this research for either the huge effectiveness gains predicted for Internet-based approaches, or for the inevitability of deeper technology use, some re-evaluation both of public administration approaches and of government policy seems called for.

In many ways, the conclusions of this study are closer to the earlier work of Margetts (1999) than they are to the more enthusiastic later propositions of DEG (Dunleavy et al., 2006a). Like the earlier work, the evidence suggests a careful and critical approach to the use of technology in public administration, and the rejection of both technologically-naïve approaches like NPM and variously utopian and dystopian predictions by technological determinists.

Rather than proposing a digital-centric ideal type such as DEG, therefore, I would focus on more fundamental questions.

From where in a technological change programme are improvements to be found? For some work, such as the desire to replace armies of form-processing tax clerks with digital data inputs and automatic tax calculation, the gains are obvious and measurable. For others, such as the value of local tourism promotion websites, the issue is much more fact-specific. How many people will actually *use* the new service? Not only how much will it cost to set up, but what are the ongoing institutional implications of maintaining this data online?

This question, of the degree of use of online services, does not get enough attention by either practitioners or academics.

On the policy side, I suspect that the reasons are similar to those for the evaluation of other government services: the drive to have got something established, to have made a mark, is much greater than the desire to subsequently put it all on the line by (potentially) proving it was ineffective in practice.

This links in with the notion that modernisation is simply inevitable: that the public expect government to be online simply as part of the modern world. The findings of Section 6.4 suggest that this motivation is not borne out by greater public satisfaction with online local authorities, but that doesn't necessarily mean it's not believed to be important by public officials.

In the local government case, the merits of declaring to elected councillors or the public that a new online service is available are clear; the merits of rigorously evaluating every one of the hundreds of hosted service areas or the tens of thousands of written web pages are less obvious.

For academics, some of the answer must come down to data and its availability. While practitioners and the sites themselves are relatively easy to get hold of and study, the demand and use side is much harder. Either a detailed exploration of council web use would need to be undertaken with members of the public, which is expensive at any degree of scale and very difficult to arrange in such a way that more than a handful of councils' citizens are sampled, or quantitative data on service usage would need to be collected on a very wide scale.

While there are a number of options for this, explored in Section 3.3.4, none of them are particularly easy. Given the relative unpopularity of the Internet with public administration scholars, these resources remain relatively untapped.

It would be wrong to discard the broad-scale picture entirely, and there is a need for a greater number of digitally-aware theories to sit alongside DEG as possible futures for public administration. The work done by GDS, Socitm members, and others remains important.

There is also a certain appeal to the Californian sense of possibility that the hyper-modernists and the utopian theorists of the early Internet bring to the detailed and quotidian world of public administration.

On balance, though, the findings of this thesis suggest that the delivery of online services remains complex, both in principle and in the results as delivered, and we must continue to look at digital reforms with a critical eye.

7.4.3 *Towards a theory of Digital Era Local Governance*

If we reject the wide appeal to digital solutions of DEG on the basis of the discussion above, but accept on the evidence of Chapter 6 and previous e-government work that there is evidence for *some* positive effects of *some* digital service improvements in *some* places, it is nevertheless possible to build a more nuanced theory of Digital Era Local Governance which recognises the strengths of DEG while being clearer about where digital gains are actually expected to lie.

The empirical foundation of such a theory would be the complex and heterogeneous picture of local government's online presence uncovered in Chapters 4 and 5 and discussed above. The focus of the original DEG work on IT contracting as a source of difficulties and potential gains is inapposite in a local government context simply because the kind of large and inflexible contracts noted by Dunleavy et al. (2006a) are not a serious issue in local government, at least in England (see Section 2.4.7).

Taking the three strands of DEG in turn, it is necessary to assess the contribution that needs-based holism, reintegration, and thoroughgoing digitalisation could make to local governments in a very different position than the central governments which inspired the original DEG work.

Needs-based holism still has an obvious attractiveness in a local context. Local governments are large and complex organisations characterised by siloed service delivery arms delivering different services to citizens with limited

horizontal integration. There has also been substantial progress made by local authorities in presenting a single face to the public, not least through the provision of council websites structured around needs rather than replicating internal structures.

Nevertheless, it is unclear whether holism can be empirically expected to increase as an organising principle. Not only are the resource pressures in local government exceptionally tight, but the firm barriers in organisation between local authorities and central government agencies suggest that holism will remain normative and aspirational in at least the near term.

It is not clear that reintegration can be anything like as central to a local theory of public administration, at least in the English context, as it is in DEG. There is no evidence of service reintegration in English local government in actual practice (see Section 7.3.1) and we have no a priori reason to believe that making local governments even larger and more complex will inherently lead to better outcomes. Already councils have powers to work across organisational boundaries for local gains, and this may be a more fruitful route for improvement than further focus on the machinery of local government itself and increased reorganisation in a world where management capacity is already constrained by austerity cuts.

As far as the digitalisation pillar of DEG goes, there are still clear opportunities for service improvement through the better use of technology. As the informatisation literature shows, where there are high-volume routinised transactions being undertaken then better information processing techniques and appropriate technology can bring genuine returns. That there is not a negative association between local government use of the web and performance shown in Chapter 6, despite the obvious costs in money and management time of implementation, shows there must be some gains to be had in a local context. The focus on 'top tasks' by Socitm and others is therefore to be welcomed.

Beyond transactions, though, a distinctive insight from local government identified in this research and potentially of general value in theories of digital

public administration is the value of informational uses of the Internet for citizens. Empirically, as analysed in Sections 5.3.2 and 5.3.3, much of the local government web is not service focused and the great majority of it is not transactional. Normatively, there are reasons explored in Section 2.4 to believe that informational uses are in fact central to the value of local e-government, contrary to the prevailing academic and practitioner focus on transactions.

However, a realistic theory of Digital Era Local Governance cannot take the broad-brush enthusiastic attitude to digitalisation which is characteristic of DEG and (in an even stronger form) of its intended successor EDGE (see Section 2.3.1). Where the costs of implementation are too high, or the expected benefits too modest, we must resist the temptation to place enthusiasm for technology ahead of our better judgement. There should certainly be no expectation that the high costs of delivering large-scale, professionalised, face-to-face services such as education can be substantially reduced.

Given the heterogeneous nature of local government and its web provision, much of the implementation of digital approaches would need to be done council-by-council, which would increase costs. Given the difficulties and significant disadvantages of implementing a single local.gov.uk site identified in Chapters 4 and 5, this is not easily overcome by increased centralisation of local government and its functions.

To know the proper scope of local government digitalisation, public administration researchers would need to deeply understand not only the scope of existing provision, and also to be able to analyse areas where the gains are likely to outweigh the costs, based on an understanding of the user perspective. It would be necessary to undertake a research programme complementing the analyses in Chapters 4 and 5, but focusing directly on the experience of the users of online systems and measuring their volumes in different areas of local government practice. As discussed in Sections 3.3.4 and 7.3.1, there are opportunities in new online data sources and existing datasets for answering this question. An actionable theory of Digital Era Local Governance would

need to be grounded on data; there is no space in a post-austerity world for the kind of techno-optimism which characterised the Blair government's desire to see everything online regardless of evidence of value or citizen use.

7.5 FULL-POPULATION WEB CRAWL DATA IN POLITICAL SCIENCE RESEARCH

There has been a trend in the past few years towards the use of “Big Data” methods to study social phenomena. These have been described as using full-population data in place of traditional sampling methods, using digital trace data in place of that which has been purposively collected, and as a result bringing much larger and more complete datasets to subjects which have previously been data-poor. Big Data approaches have also been criticised as being atheoretical, resulting in over-fitting to the data which happens to be available and identifying of chance correlation rather than causal relationships.

In many ways, this study has taken a Big Data approach to answering some of the questions set, making extensive use of machine learning tools over a large new full-population dataset, and in doing so has highlighted some of the strengths and weaknesses of it within the public administration field.

One of the key issues that has been encountered in this research is the need for more discussion around disciplinary standards for the conduct of web crawling and subsequent data analysis in political science. Although the standards developed in disciplines such as webometrics, information science, and computer science are in large part applicable, there is not as yet a standard road-map for the collection of large web data sets, which could serve both as a guide to researchers getting into this new area and as a way of signposting points of departure for new method development.

7.5.1 *Full-population datasets*

The use of full-population datasets in political science research has been lauded by Dunleavy (2010a), amongst others. This study is based in part upon a crawled collection of web pages in the .gov.uk domain with the aim of being as complete as possible. Nevertheless, as identified in Section 3.4.2, the challenge of reaching a genuine census dataset has not been reached and is not possible in absolute terms.

Firstly, the boundaries of any set of interest are necessarily fuzzily defined. Are there web pages properly relating to local government which are hosted outside their main web domain? If so, they were not included. What about quasi-governmental bodies such as local public-private partnerships, which have their own legal identity (and perhaps website) but are intimately linked with the delivery of local government services? Should such bodies even be included? If so, how are they to be identified? A good example in the present dataset is the handling of FixMyStreet. This is a third-party service, adopted by some councils as a way for citizens to report potholes and other street issues. Oxfordshire hosts a customised instance at fixmystreet.oxfordshire.gov.uk, which is fully crawled. Southampton has a customised instance hosted at the service provider's domain, southampton.fixmystreet.com, which is outside the crawl's scope. Other councils have no commercial relationship with the provider at all, but submissions can still be made for these authorities at the main www.fixmystreet.com portal. Even knowing all this, it is a matter of researcher judgement whether Oxfordshire's site should be included in their figures, or Oxfordshire's and Southampton's, or neither.

Secondly, the boundaries of what constitutes a "page" of content on the web is not well-defined. For example, some councils have links to web database systems which can generate an almost infinite number of ways of presenting

an underlying set of data (perhaps statistical or geographic)⁴. When looking at links, or volume of content, or timescales, should database-generated pages count? If included, they potentially overwhelm the human-created pages in any site; as a result, the usual researcher response is to exclude them (Thelwall, 2003). If excluded, however, the analysis is necessarily incomplete. Existing web crawling software such as Heritrix (Internet Archive, 2013) is generally written for archivists, not for researchers. As a result, it is biased towards collection, not selection. A substantial proportion of the pages collected in a web crawl are junk; this doesn't matter if the resulting archive is to be browsed by URL, as in OpenWayback (International Internet Preservation Consortium, 2014) or passively stored, but is a barrier to systematic reuse for research. Sometimes, a case-by-case manual filter is sufficient to eliminate a significant proportion of the junk⁵. In others, the technical construction of the site makes this extremely difficult⁶.

Thirdly, despite this last point, not all data is available even to a dedicated researcher. Many sites use the Robot Exclusion Standard to prevent web crawlers from accessing some content. In many cases, this is used to prevent indexing of automatically-generated pages, or dynamically-generated parts of sites that would create a trap for crawlers. In others, it is a relatively-ineffective way of hiding particular content from public view.

In either case, a partial view is necessarily taken by researchers relying on full-population collection. Another source of non-collection is pages which are not linked from any other part of the crawl (limited here by using every

4 See <http://lis.darlington.gov.uk>, for example, which offers various sorting options, or the other examples of pathological URLs seen in Section 3.5.2.

5 In this research, the <http://unistats.direct.gov.uk/> crawl scope was carefully set to allow only one possible "route" through the data for the crawler, which should guarantee that each item was only crawled once.

6 For Modern.Gov installations such as <http://democracy.anglesey.gov.uk/>, a varying client-ID in the URL means that each click creates what is effectively in the crawler's eyes a complete new site. It would be possible to deal with this prospectively by creating customised URL canonicalisation rules which handle this, but these rules need to be synchronised across a crawl in order for the Wayback Machine to be effective, so it is not easy to handle this retrospectively. In the present crawl, the offending databases have simply been excluded (with some regret)

registered domain in .gov.uk as a starting seed). Other systematic biases exist with similar research in other fields: even the researchers at Google, with full access to all the traces left by users on their search engine and other services, have no way of collecting those who do *not* use Google as part of their census.

It would be useful to develop a set of metrics to measure the quality of a particular crawl, with reference to the propensity of pages to be found (Hypertext Transfer Protocol (HTTP) status code 200 rather than 404), containing new content using some kind of distance algorithm (many database-driven pages are very similar, with “new” pages generated by slightly moving a map or re-sorting data columns) and completeness against some kind of reference set. Nevertheless, “it is impossible to give an estimate of the fraction of the web that is indexable by robots because dynamically generated pages make the size of the web indeterminate.” (Thelwall, 2003, p.6)

“The key question to be addressed by researchers using commercial search engines or web crawlers to study the web itself is whether the portion covered forms a meaningful entity for study. [...] There are two arguments that the area cover [sic] by crawlers is meaningful and worthy of scholarly attention. Firstly, search engines are important information retrieval tools in themselves. Secondly, correlations established between link count measures and a non-Internet phenomenon (research ratings) for UK universities support the hypothesis that commercial search engines and independent crawlers coverage of this area of [sic] least *is not too arbitrary to be worth investigating*.” Thelwall (2003, p.6-7) (my emphasis)

None of this is a new problem; biased and incomplete datasets have been collected as long as data has been used. Nevertheless, it does illustrate that Big Data methods are not without their methodological challenges. In particular, the theoretical strength of full-population data is weakened if the population is not, in fact, full. It also makes atheoretical claims more difficult to sustain: unless the large but serendipitously-collected data supply is sufficiently robust, it becomes challenging to ascertain if the effect spotted (statistically significant because everything is in a “full-population” study) is a genuine small effect or an artifact of data collection biases.

The usual caveats therefore apply: “fishing” for correlations should be avoided; unexpected results should be triangulated with other sources of knowledge wherever possible.

Some of the issues associated with Big Data do not, on the other hand, apply to web crawl data. Replicability, for example, is a concern when using proprietary data sets associated with trace data (Schroeder, 2014). In this case, not only can the original link dataset be published, but in principle the study could be replicated with a separate dataset collected on the same principle but at a different time period or geographical scope. This would be a worthwhile, if time-consuming, piece of future work.

Likewise, the link analysis carried out in Chapter 4 is based upon the links extracted directly from the pages studied. Although the subsequent network analysis involves a degree of statistical calculation, the method is not dependent on inferring links between otherwise unrelated datasets, as is common in other big data approaches.

Despite these difficulties, the merits of the full-population approach are also highlighted by this research. The analyses in Chapters 4 and 5 would not have been possible in anything like the same form without the collection of a full dataset. The careful use of such methods in political science in the future provides new opportunities and should be encouraged.

7.5.2 *Units of analysis*

It appears from this research that census-based approaches are also not immune from the usual quantitative social science problem of assembling data which is coterminous with the units of analysis of interest. For researchers that deal exclusively in abstractions such as the nation state, sufficient good-quality aggregated data may be available from national statistical bodies and international organisations for this not to be a problem. For the rest of us, research

design is too often a compromise between the data which are desired and those which are available.

This concern applies with particular force to the use of web crawl data. One of the major attractions of full-population web crawls is that the data are available on a very granular level – that of the individual page – and can therefore be aggregated into whatever sized units are theoretically interesting, using the Alternative Document Model (ADM) or otherwise. In this research, for example, third-level domain aggregation was undertaken to reveal web structure at the level of the local authorities.

Unfortunately, the actual unit of analysis here is subtly different; it's not the local authority, but the extent of a local authority's publishing within a particular crawl scope. This is fine if our principal interest is (as with the information scientists who pioneered webometric analysis) the published information itself. In our case, interest is actually in the real-world people, organisations, and policies that lie behind the web. For this, the web is a fascinating new data source with the potential to allow wholly new kinds of analysis, but it is in a very real sense the source of a set of proxy variables, not direct measurements of the analysis units of interest.

Does this draw doubt on the conclusions of Chapters 4 and 5? I would argue not. In this case the finding that there is significant heterogeneity in the presentation of local services online is enough to be confident about the underlying claim (that local government is heterogeneous and this is important for studying electronic public administration).

By contrast, the analyses in Chapter 6 are more satisfying, largely because the website development, service, and cost data have already been prepared and aggregated to local authority level in a defensible (if not perfect) way: simply because local authorities are also the units most important to the practitioners who assemble the data sets.

7.5.3 *The ethics of web data*

Two significant ethical issues have arisen from this research which have implications for future use in political science. These relate back to the discussion in Section 3.4.4.

The first issue relates to the limitations on web crawling that result from “polite” web crawling (that is, following the published robots.txt rules from the sites being crawled). This has inevitably led to incomplete coverage, as was anticipated in advance of the crawl. The wide variation in published rules for similar content was striking: different local authorities ModernGov systems had different access restrictions applied, for example, despite the equivalence of the content.

More concerning was the number of websites with broken robots.txt files which did not actually encode the restrictions that their authors intended them to. As noted in Section 3.4.4, there were numerous examples of non-compliant exclusion files which resulted in both greater restrictions and greater latitude being applied to various sites, depending on which rules were incorrectly coded. Combined with Heritrix’s tendency to incorrectly guess possible outlinks, this led to a significant quantity of unwanted content in the crawl collection (which is a practical problem arising from ethical concern).

It is, perhaps, questionable whether this is a good situation. Future researchers have three choices. One is to place the needs of full-population data ahead of the wishes of the bodies being studied and to crawl “impolitely”. This has the ethical problems discussed earlier, but the significant advantage of providing a uniform dataset. Another is to accept the disconnect between what was presumably intended by the site publisher and the actual machine-coded rule that they promulgated. In this case, more restrictive Heritrix settings would reduce researcher workload. Finally, a deliberately more restrictive approach could be taken. A looser parser could be written for robots.txt files which silently fixes up obvious errors such as erroneous line breaks, and per-

haps also infers “desired” conduct by looking at the restrictions placed on different crawlers⁷.

The second issue is one of copyright, and how it relates to the future use of research data and of replicability. Although the dataset for this thesis was collected after careful ethical consideration, and with approval from an institutional ethics committee, it is far from trivial to make the resulting dataset available for other researchers. The major national repository for political scientists (the UK Data Archive hosted at the University of Essex) has restrictions on the accession of copyright-encumbered materials, and university repositories often have similar restrictions.

The consequence is that the derived datasets for this (and similar) research are fully publishable and the results are therefore replicable by any researcher who cares to try. The network dataset developed in Chapter 4, for example, is now public (Nicholls, 2016). The models from Chapter 5 are also available, but not the raw data that would be required to recreate them or to (for example) run a sensitivity analysis on the modelling assumptions.

This is not sustainable in the longer-term if political science continues along the road of both requiring public data availability for academic integrity and also encouraging the wider use of data science techniques. All textual methods, for example, rely on text which is presumptively copyrighted. The larger the volume of textual input data used in future work, the greater the probability of a copyright holder objecting to the use. Further, the issue of size (the raw crawl data for this thesis alone runs to several terabytes) limits the number of suitable repositories and makes difficult the well-established practice of requesting research data direct from the original researchers: such volumes are too large for an individual or a small research group to commit to keeping secure and available in perpetuity.

⁷ robots.txt files can be written with different rules for different HTTP User-Agents; in cases where the default case is not properly functional, there are often more specific rules for the Googlebot or other web crawlers.

These ethical and practical concerns are not unique to political science. Similar issues have been raised in the context of Internet studies and the use of copyright-encumbered or fully-proprietary datasets. As a disciplinary community, though, these conversations need to be started.

7.5.4 *The practical challenges of large-scale web data analysis*

This section is a discussion of the significant practical difficulties that arise with the use of large-scale data. It exists to signpost issues for the future, as well as to reflect some of the learning from the study.

The first observation is that the kind of quantitative work undertaken in this thesis is not routinely taught to political scientists. As a body, we are frequent users of surveys, elections data, and various other forms of data which are amenable to regression and time-series analyses. The network methods used in Chapter 4 have great potential for analysing other kinds of relational data, but we have no disciplinary models of use or recommended protocols for carrying them out. Hopefully this study has provided some possible answers to the questions of methods which have arisen, but it will need substantial ongoing work for network methods to become well-established in the field.

The second observation is that while handling the volumes of data generated by a substantial webcrawl is not beyond the capacity of a decent desktop computer, the data cleaning and analysis processes are an order-of-magnitude more lengthy than those which arise with more conventional political science data.

Most stages of processing of the data for this thesis took from days to weeks of clock time to complete – without access to grid computing facilities such analysis is inevitably fairly expensive in time. It's also expensive in researcher time: constructing usable data series from the web, in particular, is a technical and meticulous task which (at worst) requires an understanding of the quirks of each of the web servers from which data is being collected, partly because

of the huge variations in content and degrees of standards-adherence that characterise the live web. Research is much easier where standard interfaces to data are available; HTTP servers are not, in practice, as standard as the standards say they ought to be.

Even after the difficulties of data extraction and processing, the volume of data produces limits on the analysis. In Section 5.2.1 the Support Vector Machine (svm) would have been much easier to fit for content classification if it had been possible to extract the substantive content of each page from the navigation frame etc. that surrounds it. For a handful of sites, parsing routines can be written to identify the substantive content and extract it; at full web scale it is practically impossible.

There is, therefore, something of a trade-off of scale for nuance (or, perhaps to phrase it in computer science terms, of completeness for correctness) of a kind which is familiar to political scientists in other areas. The decision to use Heritrix's liberal default link discovery algorithm certainly greatly increased the volume of data cleaning for a relatively minor gain in completeness and was probably not a good trade-off in retrospect. A decision to restrict the scope to a handful of local authorities, or to have restricted the scope of crawling, would have produced a different kind of study – but a much quicker and simpler one.

None of this is to underplay the value of this kind of dataset, or the kind of enthusiasm shown by Dunleavy (2010a) for census-level web data, just that (as with everything else in the world of research) there are costs as well as benefits to its collection.

7.6 CONCLUSIONS

The empirical findings of this thesis provide evidence against the enthusiastically modernist thesis that technology in some ways has, or will, transform the way governments operate. Rather, they suggest that the online world is

not a clear mirror of the offline one (and therefore that efforts to use the web as a proxy for government are problematic).

Chapter 4 produced a new and unique full-population web crawl dataset to analyse the shape and connectivity of online government. It revealed that the structure of online government is not oriented around holistic service delivery. Instead, a heterogeneous picture is built up of different local government agencies developing their website coverage in idiosyncratic ways. Linking patterns being loosely influenced by geography but otherwise remarkably uninformative. This analysis is extended by Chapter 5 uncovering significant heterogeneity in both the volume and topic of content published by councils online.

In neither case do we find much evidence of DEG changes taking place in practice. The DEG notion of digitalisation as a desirable goal was further challenged by the finding that there was no association between Internet development and service performance outlined in Chapter 6.

The DEG thesis is not focused on local government and doesn't seem a particularly good fit for it. The proposed shift towards reintegration seems to be stalled or reversing in English local government, with the continued spinning out of services (including, recently, all council-run schools to private / voluntary academy status). Coupled with the challenges to informatisation in Chapter 6 and the lack of evidence of needs-based holism identified in Chapters 4 and 5, the applicability of DEG to English local government is contested.

More broadly, there is a need to bring local government, as well as the use of technology, into the heart of public administration scholarship. Local government and central government are differently situated, subject to different internal and external pressures, and managed by different people. It is no surprise that they use technology in different (and, in the case of local government, diverse) ways.

Three possible explanations are offered for the empirical findings.

One explanation is that the findings fit well into the general history of computerisation reforms over-costing and under-delivering. This has been repeatedly attested in public administration since the 1970s and represents a compelling possibility here.

Alternatively, the relative lack of impact could be the result of a poor understanding of the underlying mechanisms for change by practitioners during the implementation of digital government; a greater focus on deliverable benefits and a more nuanced scepticism about the applicability of models derived from disparate e-commerce companies could lead to better results in the future.

Finally, a greater focus of effort on the services that would most benefit from informatisation seems called for. The findings do not support the scattershot approach of the Blair government to put all government services online, irrespective of the purposes of each service and the hugely variable value that it can bring to them.

The final argument made in this chapter is for a greater use of full-population datasets and machine learning techniques in public administration and political science more broadly. Nevertheless, the practical, theoretical, and ethical challenges of web scale data are also highlighted. As with the move towards greater online service delivery, enthusiasm for the benefits of new approaches must be married to a sceptical and critical analysis of their strengths and weaknesses as research methods.

CONCLUSIONS

Local governments have been developing online services, aiming to better serve the public and reduce administrative costs. However, the impact of this work, and the links between governments' online and offline activities, remain uncertain.

In attempting direct measurement of the structure and content of local government activities online, this research has generated a substantial new dataset, challenged the assumption that e-government has been associated with improvements to public administration, and developed a number of innovative approaches for handling large-scale full-population data in political science.

The conclusions of this thesis in many ways mirror conclusions from *Information Technology in Government* (Margetts, 1999). Despite a different focus in time period (the 2000s and 2010s instead of the 1970s to the 1980s), in subject (English local government instead of US and UK central government agencies), in object (Internet-based service improvement instead of back-office informatisation), and in method (full-population quantitative analysis instead of smaller-scale mixed-methods work), both studies essentially conclude that information is not a panacea, that there is little evidence of overarching transformation, but that there remain opportunities for the intelligent and sensitive use of technology to improve the delivery of public services.

8.1 THE STRUCTURE OF GOVERNMENT IN THE DIGITAL AGE

A new full-population web crawl of .gov.uk was assembled and a novel webometric analysis conducted to assess the structure, shape, and connectivity of

the English local government online, as measured using councils web presence directly.

English local government online is highly heterogeneous. The structure, shape, and connectivity of different local authorities varies substantially, even when accounting for variations in authority type and population size.

There is remarkably little intrinsic structure to the UK government web. A hierarchical clustering of the whole of .gov.uk using Infomap reveals a geographic association, which is confirmed by the finding that neighbouring local authorities interlink substantially more often than do more distant councils. There is no support, however, for the idea that similar groups of services are strongly connected, or that there is a concerted effort by public authorities to provide holistic and joined-up services between separate service delivery bodies, of the kind advocated by Digital Era Governance (DEG).

There is, overall, little evidence that online structures closely reflect the offline activities of government. An argument is made in Section 4.4 that this is partly due to the extensive number of local government services which are inherently local and offline in scope, and which are not amenable to digitalisation. This raises questions about the blanket application of digital approaches as a means of delivering improvements in services, and about digital-centric theories of public administration such as DEG.

Given that local government offline and online have different content, different strengths, and different service focuses, this chapter concludes that the measurement of government web content only gives us partial information about the overall structure of government in the digital age, and inferences from online to offline activities should be treated with caution.

8.2 LOCAL GOVERNMENT USE OF THE INTERNET: CONTENT AND CONTEXT

The content of the web pages in .gov.uk was directly analysed using novel machine learning and automated classification tools, making use of the unique

opportunity that this dataset provides for direct measurement of the government web at very large scale.

Again, heterogeneity of the volumes and types of content was found. A topic model, trained on the content of the local authority pages, also reveals substantial variation in subject matter between councils.

The most striking finding was the sheer volume of content published by local authorities with no service purpose. Databases of agendas, papers, and minutes of committee meetings make up a very large proportion of all the documents published, as do databases of regulatory documents such as licence grants and planning decisions. The revealed preference of English local authorities is for the importance of public record keeping, not the use of the Internet to support service transformation.

In general, the informational uses of the web (including for the delivery of information about offline services) are more prominent than the transactional service delivery functions which are the focus of most public administration theory. An argument is made that information is more important in local authority e-government work than has previously been identified.

Furthermore, because the relative volume of transactional content is so low while the volume of old democratic minutes is so high, an assessment of the scale of actual use by citizens is an essential component of future research.

8.3 COUNCIL PERFORMANCE AND USE OF THE INTERNET

A unique full-population panel dataset was constructed, comprising official measurements of service quality, cost, and citizen satisfaction, together with an expert measurement of the degree of web development, for each English local authority in the period 2002–2008.

This allowed for a rigorous measurement of the association between e-government activity and service efficiency across a large number of comparable sub-national bodies. The full-population panel approach adopted further

allowed the analysis to be sensitive, by avoiding sampling error and being statistically robust to the heterogeneity that characterises individual local authorities.

There is essentially no association between the Better Connected (BC) measures of web development and either official inspectors' measures of service performance, spending levels, or citizen satisfaction. This challenges utopian models of Internet-driven change and the centrality of digitalisation to public administration reform efforts. An argument is made for a more careful, critical, and nuanced approach to e-government in public administration.

8.4 DIGITAL ERA LOCAL GOVERNMENT

The findings of this research do not resurrect New Public Management (NPM) as a useful model for future public administration reform. Nevertheless, they do raise significant questions about the effectiveness in practice of approaches based around some kind of digital service transformation. The value of broad-brush advocacy of digitalisation as a public service reform strategy is challenged.

The DEG notion of reintegration stands in contradiction to current English local government experience, which is towards a reduced service delivery scope with schools and other council services spun out.

Coupled with the findings above of a lack of evidence of needs-based holism and no finding that web development is associated with higher performance, it is argued that the experience of English local government does not fit well with the DEG quasi-paradigm.

Overall, the importance of digital-centric approaches to local government service improvement is challenged. The path to an empirically-supported and practically useful theory of Digital Era Local Government is outlined, with a case made for a strong focus on needs-based holism, a de-emphasis on the reintegration theme and focus on contracting (which apply more strongly

to central government than English local government on the basis of this research), and a more nuanced approach to digitalisation which discards utopian enthusiasm and focuses much more tightly on the provision of information to citizens alongside high-volume transactional services.

8.5 POLITICAL SCIENCE ANALYSIS OF LARGE-SCALE DATA

Several approaches taken in this research are uncommon in political science, which often relies on small-scale data rather than direct measurement of the subject of interest. The use of direct full-population data has supported the answering of the research question and provided new insight into government structure and behaviour online. Its use is commended in future studies.

Nevertheless, conceptual, ethical, and practical issues are also associated with large-scale web census approaches. These issues, and some solutions for them, are discussed in detail in Chapter 3 and in Section 7.5.

The results of this thesis establish the value of using tools from other disciplines, including machine learning and webometrics, to study phenomena in public administration. Nevertheless, the difficulties experienced in some of the analysis result in the recommendation of a careful approach to future research. It is recommended that full-population quantitative analyses be used more broadly in political science, with awareness of both the opportunities and the challenges associated with them.

8.6 LIMITATIONS, FUTURE RESEARCH AND POLICY IMPLICATIONS

The major limitations of this study relate to the challenges of assembling and analysing a true full-population dataset, as discussed in Section 7.5. The disconnect between online and offline activities identified in Chapters 4 and 5 is an important finding in itself. Nevertheless, it does limit the ability of

researchers to infer offline structures and relationships from online data in a way which is common in webometrics and information science.

The major opportunity for future research identified in the course of this project is in the collection and analysis of user data. There is extensive work in public administration looking at the inner workings of public authorities and the motivations of public servants, including a good amount within the e-government field. This thesis and a limited amount of other work has set out to provide an evidence base for analysing the objective results of e-government activity. There remains far too little work done on the use of e-government services by citizens and the effect that has on efficiency and service quality. Recommendations are made in Section 3.3.4 for ways to take this work forward.

Practically, it is not clear that the proposal to develop a central local.gov.uk domain to host shared local government online services is well-founded. The variations in what is currently being hosted by local authorities are very wide and there is no evidence of online convergence of a kind which would support the delivery of a single platform. Moreover, it is not clear in light of the analysis in this research that such a move would provide a return in service quality or cost to justify its expense and disruption.

The opportunities for service improvements through informatisation undoubtedly exist; we have extensive evidence from individual project successes that technology, applied in the right way to the right problems, can support the development public administration. New insights to understand the government web and support this work can be found by using the tools highlighted in this project. Full-population web crawling is novel in political science and there remains a huge amount that could be learned from this new dataset alone.

What we may not do is to continue asserting that technology can improve public administration – at least without careful attention being paid to the means by which gains might be had and the specific opportunities that might

deliver them. With the benefits of the last wave of e-government uncertain at least in Britain, it is incumbent upon us all to improve the evidence base for e-government work before advocating further broad-brush digitalisation that is likely to be neither efficient nor effective.



APPENDIX: DATA TABLES

A.1 ISOLATED .GOV.UK DOMAINS

domain	size	type
decymru-tan.gov.uk	2,080	Fire authorities
bedsfire.gov.uk	165	Fire authorities
yorkshirelca.gov.uk	114	Joint local government bodies
surreyfirst.gov.uk	75	Joint local government bodies
kingtontowncouncil.gov.uk	631	Parish, Town and Community councils
northinksey-pc.gov.uk	590	Parish, Town and Community councils
bridport-tc.gov.uk	474	Parish, Town and Community councils
hawardencommunitycouncil.gov.uk	431	Parish, Town and Community councils
yatton-pc.gov.uk	424	Parish, Town and Community councils
longashtonparishcouncil.gov.uk	366	Parish, Town and Community councils
billericaytowncouncil.gov.uk	364	Parish, Town and Community councils
seaham.gov.uk	354	Parish, Town and Community councils
wilsdenparishcouncil.gov.uk	320	Parish, Town and Community councils
leominstertowncouncil.gov.uk	313	Parish, Town and Community councils
wiltontowncouncil.gov.uk	288	Parish, Town and Community councils
salcombetowncouncil.gov.uk	280	Parish, Town and Community councils
brockworth-pc.gov.uk	279	Parish, Town and Community councils
silverdale-pc-lancs.gov.uk	278	Parish, Town and Community councils
broughtonandbrettoncommunitycouncil.gov.uk	272	Parish, Town and Community councils
horsforthtowncouncil.gov.uk	264	Parish, Town and Community councils
lytchetminsterandupton-tc.gov.uk	262	Parish, Town and Community councils
bearstedparishcouncil.gov.uk	248	Parish, Town and Community councils
sitlingtonparishcouncil.gov.uk	244	Parish, Town and Community councils
stanley-tc.gov.uk	232	Parish, Town and Community councils
westmoors-pc.gov.uk	222	Parish, Town and Community councils
biltoninainstywithbickerton-pc.gov.uk	220	Parish, Town and Community councils
heybridgeparishcouncil.gov.uk	210	Parish, Town and Community councils
knaresboroughtowncouncil.gov.uk	209	Parish, Town and Community councils
woughtoncommunitycouncil.gov.uk	199	Parish, Town and Community councils
rossingtonparishcouncil.gov.uk	196	Parish, Town and Community councils
waltoncommunitycouncil.gov.uk	182	Parish, Town and Community councils

domain	size	type
kimboltonandstonely-pc.gov.uk	178	Parish, Town and Community councils
nelson-mid-glam.gov.uk	170	Parish, Town and Community councils
newlandparishcouncil.gov.uk	163	Parish, Town and Community councils
swillingtonparishcouncil.gov.uk	163	Parish, Town and Community councils
cyngortrefcaerfyrddin.gov.uk	148	Parish, Town and Community councils
brymptonparishcouncil.gov.uk	146	Parish, Town and Community councils
saltfordparishcouncil.gov.uk	145	Parish, Town and Community councils
yeovilwithoutparishcouncil.gov.uk	143	Parish, Town and Community councils
kippaxparishcouncil.gov.uk	129	Parish, Town and Community councils
sturminsternewton-tc.gov.uk	129	Parish, Town and Community councils
greatbursteadsouthgreenparishcouncil.gov.uk	121	Parish, Town and Community councils
odcombeparishcouncil.gov.uk	120	Parish, Town and Community councils
saltneytowncouncil.gov.uk	112	Parish, Town and Community councils
chivelstoneparishcouncil.gov.uk	104	Parish, Town and Community councils
langport-tc.gov.uk	104	Parish, Town and Community councils
christowparishcouncil.gov.uk	96	Parish, Town and Community councils
berinsfield-pc.gov.uk	95	Parish, Town and Community councils
newcastleemlyntowncouncil.gov.uk	94	Parish, Town and Community councils
poolparishcouncil.gov.uk	90	Parish, Town and Community councils
southkirkbyandmoorthorpetowncouncil.gov.uk	90	Parish, Town and Community councils
wivenhoe.gov.uk	89	Parish, Town and Community councils
abergavennytowncouncil.gov.uk	84	Parish, Town and Community councils
corfemullen-pc.gov.uk	84	Parish, Town and Community councils
mudfordparishcouncil.gov.uk	82	Parish, Town and Community councils
congresbury-pc.gov.uk	79	Parish, Town and Community councils
ilchesterparishcouncil.gov.uk	69	Parish, Town and Community councils
bradwell-pc.gov.uk	62	Parish, Town and Community councils
aberporthcommunitycouncil.gov.uk	61	Parish, Town and Community councils
poultonparishcouncil-gloucs.gov.uk	61	Parish, Town and Community councils
parbold-pc.gov.uk	60	Parish, Town and Community councils
rawdoparishcouncil.gov.uk	56	Parish, Town and Community councils
comeytrowe-pc.gov.uk	53	Parish, Town and Community councils
stainforthtowncouncil.gov.uk	42	Parish, Town and Community councils
halesandheckingham-pc.gov.uk	41	Parish, Town and Community councils
buckinghampark-pc.gov.uk	37	Parish, Town and Community councils
welton-by-lincoln-pc.gov.uk	33	Parish, Town and Community councils
glynneathtowncouncil.gov.uk	14	Parish, Town and Community councils
scottishflooddefences.gov.uk	11	Scottish government

Table 30: .gov.uk: isolates

A.2 DISTRIBUTION OF WEB PAGE TOPICS BY COUNCIL

Council	Topics											TOTAL
	0	1	2	3	4	5	6	7	8	9	No topic	
Allerdale	1	30	175	9,193	0	7	0	2,102	0	656	1,618	13,782
Amber Valley	2	0	1	521	0	0	0	636	1	3	141	1,305
Arun	0	0	0	91	0	0	0	17	0	1	47	156
Ashfield	0	5	0	1,328	2	0	0	2,389	0	0	68	3,792
Ashford	17	3	6	747	2	11	0	1,521	0	6	442	2,755
Aylesbury Vale	2	3	9	3,958	0	49	0	2,020	2,127	72	409	8,649
Babergh	53	3	18	2,274	0	1,906	0	541	0	112	493	5,400
Barking and Dagenham	7	474	3,501	30,162	3	226	6	3,666	3	1,331	7,257	46,636
Barnet	6	8	794	4,928	0	5	2	6,274	2	471	358	12,848
Barnsley	5	2,523	535	10,248	0	118	1	2,669	8	189	6,177	22,473
Barrow-in-Furness	0	21	25	2,954	0	1	0	387	0	79	287	3,754
Basildon	1	1	131	989	0	8	1	692	6	30	1,306	3,165
Basingstoke and Deane	0	25	29	3,427	0	65	1	484	0	338	351	4,720
Bassetlaw	0	20	4	2,829	0	61	0	846	0	56	909	4,725
Bath and North East Somerset	1	159	63	14,490	0	1,574	0	538	4,270	1,113	1,351	23,559
Bedford	0	4	11	1,443	11	2	2	3,595	1	16	1,719	6,804

Council	Topics											TOTAL
	0	1	2	3	4	5	6	7	8	9	No topic	
Bexley	2	175	5	6,631	0	5	0	2,324	1	211	645	9,999
Birmingham	0	5	10	6,449	2	5,326	1	4,432	27	249	440	16,941
Blaby	2	23	2	4,208	0	4	0	689	1	201	882	6,012
Blackburn with Darwen	2	0	11	1,186	0	18	0	516	0	29	159	1,921
Blackpool	1	45	80	14,762	146	21	0	979	1	1,176	775	17,986
Bolsover	2	2	0	910	0	193	0	4,623	0	56	171	5,957
Bolton	0	0	9	1,181	9	11	0	1,950	0	35	273	3,468
Boston	0	840	86	5,486	0	0	0	1,379	1	345	418	8,555
Bournemouth	11	5	275	4,398	0	35	0	2,859	0	326	424	8,333
Bracknell Forest	10	141	33	12,894	0	113	0	2,230	0	543	1,407	17,371
Bradford	3	104	30	8,813	1	259	0	3,057	3	233	567	13,070
Braintree	4	45	37	3,695	0	186	1	8,895	4	159	309	13,335
Breckland	1	24	142	8,183	0	226	0	678	0	464	652	10,370
Brent	1	266	138	11,747	3	37	955	4,117	4	886	1,003	19,157
Brentwood	18	4	1,198	3,787	0	28	0	1,104	0	277	753	7,169
Brighton and Hove	53	69	2,604	26,074	10	337	1	2,046	1	1,953	2,919	36,067
Bristol, City of	30	3	3	1,848	1	10	1	697	277	84	409	3,363
Broadland	1	30	94	9,262	0	57	0	12,648	1	331	1,022	23,446
Bromley	27	8	86	12,664	7,694	48	0	8,268	26	712	3,161	32,694
Bromsgrove	1	21	4	7,236	117	5	1	631	0	351	1,301	9,668
Broxbourne	8	680	22	3,084	0	7	119	3,949	1	80	130	8,080

Council	Topics											TOTAL
	0	1	2	3	4	5	6	7	8	9	No topic	
Broxtowe	0	23	2	975	0	0	0	2,334	0	4	179	3,517
Buckinghamshire	7,100	4	160	4,300	1	39	0	2,885	0	108	3,042	17,639
Burnley	1	24	42	4,547	1,401	77	0	923	1	759	573	8,348
Bury	1	11	385	6,246	7	33	0	2,337	1	331	1,319	10,671
Calderdale	0	7	27	2,697	5	34	0	6,806	2	162	458	10,198
Cambridge	1	0	107	3,980	0	92	0	1,693	0	479	370	6,722
Cambridgeshire	0	5	24	2,404	0	280	0	4,281	1	71	683	7,749
Camden	0	0	111	1,322	0	0	0	13,619	0	22	358	15,432
Cannock Chase	0	42	8	6,062	0	6	1	8,309	10	194	598	15,230
Canterbury	0	1	79	957	0	10	0	934	0	142	152	2,275
Carlisle	0	1,783	22	2,120	0	110	0	4,027	0	140	217	8,419
Castle Point	0	1	2	833	1	2	0	266	0	62	117	1,284
Central Bedfordshire	2	10	55	7,296	0	230	2	1,357	0	968	1,136	11,056
Charnwood	1	79	96	8,932	1	355	107	1,576	23	375	961	12,506
Chelmsford	2	78	24	4,095	19	55	2	4,964	2	82	350	9,673
Cheltenham	2	14	145	7,345	1	100	0	7,874	0	315	604	16,400
Cherwell	2	96	17	6,585	0	312	11	1,602	0	263	1,622	10,510
Cheshire East	2	383	100	20,168	6	127	0	1,574	3	562	1,760	24,685
Cheshire West and Chester	0	5	3	4,942	0	1	0	1,599	0	15	85	6,650
Chesterfield	0	0	2	171	0	0	0	1,950	0	0	28	2,151
Chichester	0	234	7	680	0	0	0	1,421	0	2	27	2,371

Council	Topics										No topic	TOTAL
	0	1	2	3	4	5	6	7	8	9		
Chiltern	4	121	12	3,973	0	4	0	922	3	93	5,095	10,227
Chorley	24	65	332	6,197	4	720	0	301	1	448	1,900	9,992
City of London	1,759	2	123	12,547	0	66	2	15,935	0	761	1,716	32,911
Colchester	0	60	2	968	0	0	0	1,939	0	0	31	3,000
Copeland	3	14	643	4,955	1	12	0	550	17,314	355	993	24,840
Corby	2	1,575	40	4,941	0	133	0	2,569	0	349	174	9,783
Cornwall	10	15	85	13,054	1	188	0	3,691	1	331	1,945	19,321
Cotswold	2	6,194	11	651	0	183	0	1,468	0	29	222	8,760
County Durham	6	4,412	303	28,011	0	119	0	1,521	5	765	2,015	37,157
Coventry	3	41	48	21,334	12	65	2	26,952	0	301	2,445	51,203
Craven	0	152	4	1,021	0	0	0	3,660	0	0	23	4,860
Crawley	19	33	97	8,563	2	35	0	3,244	0	195	2,296	14,484
Croydon	2	421	833	14,123	0	102	1	3,389	0	411	2,166	21,448
Cumbria	10	18	231	15,472	4	259	0	16,811	2	975	2,645	36,427
Dacorum	0	41	66	993	0	11	0	1,421	0	58	180	2,770
Darlington	7,617	1,119	97	19,906	0	286	0	3,584	0	1,252	1,958	35,819
Dartford	2	32	110	10,185	1	240	0	1,029	12	690	976	13,277
Daventry	1	17	18	511	0	0	0	312	0	4	10	873
Derby	0	4,114	19	2,512	9	76	0	2,253	0	734	981	10,698
Derbyshire	16	3	10	3,706	0	68	0	4,844	0	1,534	1,223	11,404
Derbyshire Dales	0	0	1	307	1,329	0	0	356	2	0	2	1,997

Council	Topics											TOTAL
	0	1	2	3	4	5	6	7	8	9	No topic	
Devon	4	13	81	12,992	3	78	0	17,613	4	512	2,216	33,516
Doncaster	0	20	16	10,118	1	26	0	1,615	0	259	1,001	13,056
Dorset	0	0	0	0	1	0	0	2	0	0	17	20
Dover	2	15	132	4,567	0	29	0	1,161	0	110	644	6,660
Dudley	1	4,738	27	2,884	2	301	0	2,766	1	86	714	11,520
Ealing	5	13	530	7,871	0	95	2	10,446	1	275	2,616	21,854
East Cambridgeshire	0	41	662	5,249	0	281	0	1,838	6,179	181	621	15,052
East Devon	2	41	4	3,355	0	33	0	1,747	2	175	955	6,314
East Dorset	1	723	149	8,353	0	2	0	5	0	52	340	9,625
East Hampshire	2	7	4,698	5,007	0	0	0	129	0	249	1,249	11,341
East Hertfordshire	0	63	131	8,331	0	570	0	127	2	548	1,078	10,850
East Lindsey	2	246	12	8,234	1	199	0	909	5	490	1,498	11,596
East Northamptonshire	0	39	18	4,663	0	256	2	1,687	21	204	748	7,638
East Riding of Yorkshire	12	9	49	3,009	0	141	107	1,006	2	308	6,948	11,591
East Staffordshire	0	241	58	1,491	0	46	0	966	0	30	301	3,133
East Sussex	1	37	51	17,498	0	147	0	3,808	0	577	756	22,875
Eastbourne	5	12	160	4,598	0	50	0	1,167	0	279	191	6,462
Eastleigh	0	2	20	957	0	2	0	986	0	147	165	2,279
Eden	0	1,301	1	1,965	0	0	0	1,085	0	16	8	4,376
Elmbridge	0	22	37	8,518	0	133	0	1,365	0	132	403	10,610
Enfield	3	27	230	14,493	0	299	12	15,219	16	413	2,787	33,499

Council	Topics											TOTAL
	0	1	2	3	4	5	6	7	8	9	No topic	
Epping Forest	1	2,016	223	14,686	8	81	0	1,848	1	1,016	1,242	21,122
Epsom and Ewell	0	1	900	3,908	0	52	0	1,493	2	187	1,562	8,105
Erewash	1,764	3	83	4,053	0	18	2	571	8	236	848	7,586
Essex	0	2,387	17,947	9,999	4	60	1	1,692	1	349	2,233	34,673
Exeter	0	0	0	473	0	0	0	3,085	0	0	15	3,573
Fareham	0	83	28	6,780	5	33	0	6,959	0	55	138	14,081
Fenland	3	61	4	4,651	0	14	0	811	13	48	159	5,764
Forest Heath	1	13	29	1,400	0	39	0	8,979	0	103	305	10,869
Forest of Dean	4	2	6	1,226	0	151	1	239	0	83	351	2,063
Fylde	0	4	56	2,588	0	1	0	1,427	47	67	1,265	5,455
Gateshead	7	2,909	242	30,135	6,811	49	0	1,673	2	1,754	5,306	48,888
Gedling	0	26	5	5,797	0	22	2	1,334	2	230	244	7,662
Gloucester	2	1,112	238	10,226	2	196	0	4,712	12	290	1,939	18,729
Gloucestershire	1	15	33	8,784	2	343	1	8,587	0	281	847	18,894
Gosport	0	359	15	1,360	1	0	0	821	0	5	21	2,582
Gravesham	4	35	82	7,051	0	328	0	517	1	245	2,421	10,684
Great Yarmouth	0	13	11	1,095	0	7	0	236	0	74	154	1,590
Greenwich	7	1	9	1,494	0	8	3	8,071	1	131	245	9,970
Guildford	1	194	92	6,412	0	94	0	1,120	5	267	1,241	9,426
Hackney	2	1	493	5,436	0	66	0	2,740	0	119	600	9,457
Halton	14	73	173	12,012	11	96	0	1,893	1	356	1,599	16,228

Council	Topics											TOTAL
	0	1	2	3	4	5	6	7	8	9	No topic	
Hambleton	3	26	0	2,708	1	107	0	1,081	0	55	333	4,314
Hammersmith and Fulham	7	7	154	13,720	0	207	0	4,927	6	515	1,434	20,977
Hampshire	278	419	1,628	33,367	2,736	126	1	745	0	13,997	1,646	54,943
Harborough	4	3,168	182	6,035	1	37	0	6,363	5	390	845	17,030
Haringey	13	23	395	21,885	1	190	2	7,148	3	795	5,114	35,569
Harlow	0	66	15	3,034	0	22	0	3,344	0	184	613	7,278
Harrogate	0	716	64	32,928	0	45	0	2,174	2	5,165	9,461	50,555
Harrow	63	49	340	14,343	0	95	0	3,329	2	400	3,122	21,743
Hart	4	3	7	717	0	14	1	714	1	28	161	1,650
Hartlepool	7	40	29	5,879	0	27	3	13,743	0	182	329	20,239
Hastings	0	4	43	3,940	0	42	0	583	0	156	435	5,203
Havant	5	331	55	2,583	0	58	0	838	0	91	619	4,580
Havering	1	2	59	5,081	0	15	2	2,680	0	139	1,158	9,137
Herefordshire, County of	6	9	177	8,020	0	23	1	8,024	0	198	1,722	18,180
Hertsmere	1	51	272	7,430	0	10	0	807	2	550	992	10,115
High Peak	0	2	10	1,771	0	41	0	6,848	3	121	612	9,408
Hillingdon	3	8	5	2,365	0	2	0	8,381	3	79	222	11,068
Hinckley and Bosworth	57	1	130	3,753	0	88	1	8,204	0	178	751	13,163
Horsham	2	103	3	3,645	2	127	0	3,648	0	81	648	8,259
Hounslow	0	71	345	23,949	0	1,084	0	3,755	20	893	4,171	34,288
Huntingdonshire	29	48	22	21,360	4	290	0	2,512	1	952	1,306	26,524

Council	Topics										No topic	TOTAL
	0	1	2	3	4	5	6	7	8	9		
Hyndburn	1	27	11	5,111	0	187	2	4,279	0	111	643	10,372
Ipswich	0	18	84	5,331	0	37	0	1,845	3	294	686	8,298
Isle of Wight	0	0	0	292	0	1	0	0	0	0	39	332
Islington	323	127	57	8,017	1	105	415	2,791	11	127	572	12,546
Kensington and Chelsea	1	5	82	4,977	0	52	0	3,294	3	102	720	9,236
Kent	5	13	218	12,978	12	94	0	2,226	6	1,048	1,606	18,206
Kettering	12	107	45	6,707	0	65	4	16,980	1	369	1,166	25,456
King's Lynn and West Norfolk	2	97	66	6,597	0	483	0	953	3	184	1,188	9,573
Kingston upon Hull, City of	0	4,017	3	220	0	2	0	55	0	25	33	4,355
Kingston upon Thames	0	0	10	1,021	0	20	0	5,953	4	19	132	7,159
Kirklees	8	2	3	3,442	2	30	0	2,300	1	379	959	7,126
Knowsley	1	1	4	1,111	0	5	0	624	0	57	229	2,032
Lambeth	5	6	42	12,497	2	110	0	5,839	8	590	1,444	20,543
Lancashire	16	47	263	15,890	6	514	0	3,044	1	530	3,380	23,691
Lancaster	11	59	132	13,588	0	103	0	1,622	45	1,159	1,886	18,605
Leeds	20	69	42	31,206	21	347	0	2,084	14	1,314	2,961	38,078
Leicester	1	12	2,933	19,161	22	28	1	15,128	0	326	11,166	48,778
Leicestershire	34	103	1,175	35,711	5	265	0	16,036	4	2,271	6,149	61,753
Lewes	0	1,973	23	4,120	0	76	0	526	1	235	446	7,400
Lewisham	1	188	45	18,301	1	270	0	2,458	2	448	2,882	24,596
Lichfield	59,913	89	75	7,908	14	82	0	6,501	1	1,804	969	77,356

Council	Topics											TOTAL
	0	1	2	3	4	5	6	7	8	9	No topic	
Lincoln	0	47	34	16,117	0	93	0	2,910	2	632	893	20,728
Lincolnshire	26	567	1,168	22,606	1	95	0	19,308	6	455	2,378	46,610
Liverpool	0	2	24	3,190	2	12	0	2,051	0	399	175	5,855
Luton	1	7,932	13	4,337	0	28	0	3,087	0	330	2,135	17,863
Maidstone	0	1	1	141	0	329	0	387	0	6	15	880
Maldon	0	49	48	6,390	0	10	0	6,810	4	260	3,876	17,447
Malvern Hills	4	0	13	1,188	112	18	0	831	1	83	427	2,677
Manchester	5	67	53	14,762	0	81	0	15,926	32	189	923	32,038
Mansfield	1	3	19	1,009	0	0	0	4,362	0	30	257	5,681
Medway	0	68	137	11,714	1	876	0	8,468	0	667	1,006	22,937
Melton	0	109	86	5,100	0	7	0	989	3	357	683	7,334
Mendip	0	9	12	536	0	3	0	683	0	4	1,890	3,137
Merton	20	47	571	16,996	3	684	2,043	9,694	0	884	3,474	34,416
Mid Devon	0	73	2	2,122	0	0	0	1,063	0	1	27	3,288
Mid Suffolk	0	0	8	1,277	0	26	0	713	0	69	142	2,235
Mid Sussex	3	7	22	2,481	2	281	0	828	5	54	211	3,894
Middlesbrough	1	30	125	13,794	0	38	0	2,160	2	380	1,864	18,394
Milton Keynes	5	4,297	297	25,717	0	681	0	1,627	2	2,621	2,827	38,074
Mole Valley	3	134	146	3,541	0	13	0	5,477	0	59	772	10,145
New Forest	0	107	157	31,998	0	228	0	2,926	0	680	3,330	39,426
Newark and Sherwood	2	0	12	1,367	0	116	1	1,049	0	93	391	3,031

Council	Topics											TOTAL
	0	1	2	3	4	5	6	7	8	9	No topic	
Newcastle upon Tyne	4,268	18	128	13,823	12,156	42	1	8,802	3	602	4,244	44,087
Newcastle-under-Lyme	2	49	46	5,012	0	347	0	267	2	192	415	6,332
Newham	0	0	5	1,033	1	29	0	1,747	0	50	171	3,036
Norfolk	0	15	7	4,874	1	13	0	1,024	9	59	163	6,165
North Devon	3	11	3	1,725	1	39	0	771	1	99	337	2,990
North Dorset	0	0	1	4	0	10	0	1	0	0	2	18
North East Derbyshire	0	49	3	215	0	0	0	1,085	0	0	2	1,354
North East Lincolnshire	5	2,380	89	4,589	0	56	0	5,401	81	94	3,126	15,821
North Hertfordshire	23	97	307	8,414	1	313	0	1,343	4	751	986	12,239
North Kesteven	0	0	2	247	0	52	0	2,066	0	46	98	2,511
North Lincolnshire	16,470	2,929	153	7,602	6	97	1	9,247	5	582	171	37,263
North Norfolk	0	46	189	3,208	0	67	0	227	0	241	216	4,194
North Somerset	3	147	49	5,675	1	102	0	1,139	2	405	24,718	32,241
North Tyneside	0	6	31	305	0	0	0	18	0	0	35	395
North Warwickshire	4	13	7	2,106	1	1	3	11,962	3	77	247	14,424
North West Leicestershire	2	19	24	6,507	2	47	75	2,595	25	380	632	10,308
North Yorkshire	4	2,979	34	25,032	4	129	0	3,178	1	1,406	8,561	41,328
Northampton	0	1	8	1,243	7	9	0	5,773	0	118	442	7,601
Northamptonshire	2	1,222	73	6,645	2	92	0	1,448	2	342	2,008	11,836
Northumberland	18	32	382	16,418	0	294	0	922	15	879	4,442	23,402
Norwich	1	10	191	11,344	1	103	0	1,610	3	318	559	14,140

Council	Topics											TOTAL
	0	1	2	3	4	5	6	7	8	9	No topic	
Nottingham	4	16	21	12,702	1	254	0	7,434	4	181	879	21,496
Nottinghamshire	3	13	33	9,870	64	71	1	14,484	1	212	2,082	26,834
Nuneaton and Bedworth	0	17	5	1,954	0	7	0	4,903	0	48	265	7,199
Oadby and Wigston	4	10	16	4,459	0	27	35	1,769	2	356	904	7,582
Oldham	2	58	14	10,678	2	78	4	6,332	0	381	3,077	20,626
Oxford	10	18	115	12,050	0	121	3	3,224	2	656	3,238	19,437
Oxfordshire	108	85	66	15,026	8,097	182	17,478	10,956	4	774	1,634	54,410
Pendle	10	58	3,902	5,145	1,275	95	2	18,731	0	208	7,770	37,196
Peterborough	29	2	796	3,358	1	24	0	3,453	0	291	1,105	9,059
Plymouth	10	37	4,082	18,742	4	281	21	8,181	4	423	5,438	37,223
Poole	0	0	0	13	0	0	0	89	0	0	0	102
Portsmouth	7	180	115	23,053	2	173	0	5,700	2	1,109	3,756	34,097
Preston	0	1	38	1,427	0	17	0	1,403	0	46	108	3,040
Reading	100	2	92	5,590	1	247	0	7,293	11	512	1,758	15,606
Redbridge	84	5	9	2,292	13,231	566	0	4,498	0	72	28	20,785
Redcar and Cleveland	0	561	77	7,929	0	73	0	1,506	0	451	737	11,334
Redditch	0	40	16	5,134	32	298	0	32	0	477	360	6,389
Reigate and Banstead	191	37	52	6,725	0	117	0	1,136	14	220	1,438	9,930
Ribble Valley	0	13	111	4,567	0	40	0	5,977	0	243	969	11,920
Richmond upon Thames	1	28	103	6,282	0	76	0	4,617	0	208	1,158	12,473
Richmondshire	0	118	16	746	0	5	0	3,383	1	13	75	4,357

Council	Topics											TOTAL
	0	1	2	3	4	5	6	7	8	9	No topic	
Rochdale	0	38	16	8,090	0	77	0	1,775	0	478	713	11,187
Rochford	14	2,131	77	1,741	0	14	0	831	7	35	165	5,015
Rossendale	7,727	43	321	7,363	0	23	0	2,556	1	238	741	19,013
Rother	0	65	1	1,063	0	1	0	1,759	0	8	43	2,940
Rotherham	4	119	72	27,222	34	84	0	6,880	11	642	3,376	38,444
Rugby	6	24	24	1,663	0	1	2	6,409	2	43	422	8,596
Runnymede	1	253	207	5,448	5	54	0	1,265	0	229	2,466	9,928
Rushcliffe	0	171	40	2,899	0	1,500	0	1,388	1	81	711	6,791
Rushmoor	0	0	18	681	0	9	0	2,054	0	58	14,630	17,450
Rutland	5	94	141	5,413	0	257	0	1,444	0	449	3,232	11,035
Ryedale	7	20,312	141	6,398	0	682	0	1,159	6	325	4,284	33,314
Salford	19	51	468	29,107	4,898	74	1	4,525	7	1,467	3,153	43,770
Sandwell	0	9,641	19	1,636	0	8	0	6,593	0	100	222	18,219
Scarborough	4	95	43	13,312	0	328	1	1,620	1	373	3,350	19,127
Sedgemoor	0	1	0	442	0	0	0	461	0	0	7	911
Sefton	0	8	75	11,416	0	18	0	1,311	7	233	865	13,933
Selby	15	2	221	3,258	0	234	0	84	4	147	774	4,739
Sevenoaks	80	86	146	9,989	0	170	0	1,035	2	667	1,318	13,493
Sheffield	15	14	39	7,248	3	119	0	4,851	0	321	1,249	13,859
Shepway	9	17	23	5,849	0	115	0	1,705	11	554	1,196	9,479
Shropshire	0	1,494	244	11,804	0	72	0	4,330	0	462	966	19,372

Council	Topics											TOTAL
	0	1	2	3	4	5	6	7	8	9	No topic	
Slough	1	5,435	28	11,167	0	77	0	2,619	1	403	1,492	21,223
Solihull	1	163	746	23,161	0	308	0	5,042	2	1,597	13,751	44,771
Somerset	16	875	106	19,214	568	33	0	3,750	3	707	8,462	33,734
South Bucks	0	2	0	185	0	0	0	600	0	1	0	788
South Cambridgeshire	17	0	205	4,608	0	633	0	8,019	5	182	6,647	20,316
South Derbyshire	21	3	23	2,483	0	42	2	973	0	77	184	3,808
South Gloucestershire	4	20	1,443	12,145	2	106	0	5,570	1	440	4,111	23,842
South Hams	0	50	93	11,359	0	28	0	1,935	0	1,019	371	14,855
South Holland	16	529	87	10,321	0	13	0	1,505	0	422	2,863	15,756
South Kesteven	4	13	15	1,213	0	2	0	1,973	0	16	781	4,017
South Lakeland	3	31	118	11,913	0	354	0	720	0	520	4,833	18,492
South Norfolk	2	21	300	8,130	0	112	0	985	0	460	445	10,455
South Northamptonshire	1	151	131	7,792	0	209	1	1,236	0	526	962	11,009
South Oxfordshire	4	28	12	5,143	0	92	14	975	3	395	633	7,299
South Ribble	1	9	170	3,084	0	7	0	1,560	0	185	395	5,411
South Somerset	29	31	41	14,788	0	94	0	830	3	816	1,972	18,604
South Staffordshire	4	519	304	10,119	1	107	0	1,211	0	102	1,429	13,796
Southampton	48	58	131	15,514	0	249	0	2,713	4	1,112	1,563	21,392
Southend-on-Sea	62	49	167	12,104	1	72	0	5,717	5	853	1,736	20,766
Southwark	0	31	258	17,562	0	256	0	15,707	12	604	4,587	39,017
Spelthorne	0	104	0	777	0	0	0	2,010	0	2	97	2,990

Council	Topics											TOTAL
	0	1	2	3	4	5	6	7	8	9	No topic	
St Albans	0	9	44	888	0	12	0	1,055	0	17	2,610	4,635
St Edmundsbury	1	9	12	7,820	0	42	0	1,183	0	115	787	9,969
St. Helens	4	93	138	6,631	0	17	0	3,729	10	331	1,046	11,999
Stafford	0	323	83	6,323	0	24	0	4,283	4	133	624	11,797
Staffordshire	8	6	20	11,430	0	210	0	4,383	2	439	708	17,206
Staffordshire Moorlands	1	9	0	1,445	0	41	0	1,782	3	103	737	4,121
Stevenage	0	8	48	4,155	3,786	13	0	1,155	0	348	479	9,992
Stockport	0	191	67	10,391	0	274	0	3,692	2	342	1,277	16,236
Stockton-on-Tees	8	77	74	14,675	0	49	0	352	0	316	2,806	18,357
Stoke-on-Trent	0	0	4	1,238	1	7	0	3,029	0	25	199	4,503
Stratford-on-Avon	2	19	22	8,557	0	753	0	1,700	7	363	1,547	12,970
Stroud	3	78	20	10,626	0	32	0	619	2	342	1,422	13,144
Suffolk	1	104	97	6,768	0	8	0	26,228	1	93	913	34,213
Suffolk Coastal	3	42	354	3,538	1	105	0	357	4	126	979	5,509
Sunderland	0	4	2	1,092	0	0	0	1,979	0	3	62	3,142
Surrey	0	13	4,112	34,819	8	641	0	8,902	5	1,994	13,133	63,627
Surrey Heath	3	36	78	6,546	0	9	0	1,032	3	223	952	8,882
Sutton	0	0	2	1,662	0	0	0	2,824	0	1	56	4,545
Swale	22	106	596	9,092	1	100	0	1,505	1	321	1,177	12,921
Swindon	6	428	155	32,626	0	445	0	1,899	1	1,441	4,625	41,626
Tameside	0	31	89	7,332	1	2,523	0	3,214	13	51	1,801	15,055

Council	Topics											TOTAL
	0	1	2	3	4	5	6	7	8	9	No topic	
Tamworth	1	6	15	2,070	0	9	0	767	0	125	404	3,397
Tandridge	1	1,516	4	1,068	0	1	0	827	4	137	129	3,687
Taunton Deane	8	3	780	1,462	0	42	0	420	0	45	362	3,122
Teignbridge	0	13	0	427	3	0	0	4,074	0	3	9	4,529
Telford and Wrekin	11	12	23	3,529	0	197	0	16,955	22	339	2,467	23,555
Tendring	10	18	84	3,059	1	8	0	5,938	0	175	2,944	12,237
Test Valley	3	0	776	3,120	0	56	1	1,874	0	163	2,042	8,035
Tewkesbury	8	32	47	6,895	0	55	0	686	0	687	2,861	11,271
Thanet	24	9	39	6,460	0	25	0	1,922	0	214	986	9,679
Three Rivers	0	14	76	1,165	0	2	0	1,320	0	20	1,906	4,503
Thurrock	0	5,791	47	2,028	0	39	0	4,026	0	55	36	12,022
Tonbridge and Malling	17	134	1,589	26,537	0	55	1	1,253	23	3,031	4,625	37,265
Torbay	10	17	33	11,638	0	125	0	14,055	0	451	1,539	27,868
Torridge	0	33	0	486	0	0	0	3,018	3	3	59	3,602
Tower Hamlets	30	75	29	11,688	2	116	1	2,749	2	719	1,564	16,975
Trafford	0	11	137	9,437	21,590	32	2	2,595	1	111	503	34,419
Tunbridge Wells	1	70	99	8,846	0	42	0	61	1	583	697	10,400
Uttlesford	0	894	1	675	2	0	0	808	0	0	39	2,419
Vale of White Horse	14	18	51	6,200	0	130	7	984	8	521	1,973	9,906
Wakefield	51	0	4	2,373	1	31	0	1,266	0	88	890	4,704
Walsall	24	4,383	17	11,307	0	13	0	3,338	0	259	3,610	22,951

Council	Topics											TOTAL
	0	1	2	3	4	5	6	7	8	9	No topic	
Waltham Forest	6	27	103	13,727	1	42	0	6,416	7	927	1,318	22,574
Wandsworth	2	66	294	21,668	27	257	5	16,017	0	986	1,346	40,668
Warrington	0	38	93	4,322	1	70	1	12,083	0	102	854	17,564
Warwick	695	308	12	1,314	0	54	3	3,006	4	102	322	5,820
Warwickshire	9	1,520	177	34,861	2	73	152	5,967	29,542	1,493	5,710	79,506
Watford	1	15	2	572	0	0	0	3,301	0	10	8	3,909
Waveney	28	14	199	6,643	0	63	0	5,731	6	406	1,805	14,895
Waverley	4	34	167	7,421	0	13	0	7,246	0	471	3,269	18,625
Wealden	0	4	0	1,016	0	0	0	3,722	0	5	51	4,798
Wellingborough	4,818	35	18	4,197	0	27	1	1,837	0	130	387	11,450
Welwyn Hatfield	0	21	0	571	0	0	0	773	0	0	18	1,383
West Berkshire	0	143	113	8,862	1	286	0	7,311	4	365	1,066	18,151
West Devon	0	3	0	113	0	0	0	873	0	0	2	991
West Dorset	0	0	0	10	0	0	0	6	0	0	16	32
West Lancashire	3	403	14	3,029	1	13	0	6,380	0	27	318	10,188
West Lindsey	1	547	10	9,213	0	54	0	1,623	32	481	5,328	17,289
West Oxfordshire	7	676	96	2,450	0	96	20	1,302	0	29	893	5,569
West Somerset	1	300	18	6,782	7	37	0	1,458	0	285	560	9,448
West Sussex	49	84	41	9,820	0	257	0	10,044	1	225	611	21,132
Westminster	6	28	254	63,356	559	475	1	2,389	1	2,136	16,834	86,039
Weymouth and Portland	0	0	3	174	0	0	0	17	1	1	33	229

Council	Topics											TOTAL
	0	1	2	3	4	5	6	7	8	9	No topic	
Wigan	2	3	45	2,995	2	58	0	1,524	1	95	296	5,021
Wiltshire	31,781	38	282	29,060	0	296	1	2,364	1	968	3,145	67,936
Winchester	2	1	787	6,107	0	124	1	1,948	2	185	1,162	10,319
Windsor and Maidenhead	1	34	225	9,137	0	1,239	0	5,572	0	411	1,584	18,203
Wirral	6	1	97	5,007	3	24	0	1,790	3	203	1,223	8,357
Woking	1	4	25	1,928	0	3	0	1,511	0	97	194	3,763
Wokingham	0	48	9	1,466	0	0	0	1,973	0	0	84	3,580
Wolverhampton	0	0	0	107	0	0	0	2,449	0	3	12	2,571
Worcester	2	30	8	7,636	0	83	0	427	1	322	542	9,051
Worcestershire	12	8	14	8,728	110	131	0	4,911	4	656	2,041	16,615
Wychavon	2	0	4	1,358	0	92	0	840	0	47	963	3,306
Wycombe	2	63	4	3,600	0	4	0	1,021	1	209	671	5,575
Wyre	14	2	130	3,882	0	56	0	10,479	0	146	2,118	16,827
Wyre Forest	13	57	220	3,621	27	64	0	857	0	134	544	5,537
York	19	100	228	23,672	1	2,492	2	18,607	6	1,484	5,556	52,167
TOTAL	146,895	142,985	86,290	2,672,604	87,155	52,874	21,674	1,258,745	60,730	142,116	553,045	5,225,113

Table 31: Distribution of web page topics by council

A.3 PERFORMANCE REGRESSION MODELS INCLUDING COUNCIL DUMMY VARIABLES

	CPA services (normalised)	Excess spending (£/cap)	Satisfaction (%)
	(6)	(9)	(13)
BC Transactional or better	-0.001 (0.064)	-1.028 (1.462)	-0.958 (1.311)
CPA corporate assessment	0.390*** (0.055)	0.363 (1.409)	2.493* (1.499)
CPA services (normalised)		-0.615 (0.749)	0.284 (0.745)
Excess spending (£/capita)			-0.070 (0.058)
2003	0.513*** (0.066)	6.874*** (1.334)	
2004	0.864*** (0.067)	-33.157*** (1.453)	
2005	0.295*** (0.068)	-19.219*** (1.358)	
2006	0.581*** (0.070)	-16.814*** (1.470)	
2007	0.783*** (0.074)	-15.379*** (1.617)	-0.472 (1.357)
2008	1.145*** (0.077)		
Barnet	0.380 (0.303)	32.143*** (6.380)	6.580 (5.921)
Barnsley	0.708** (0.307)	2.522 (6.490)	-9.515* (5.699)
Bath and North East Somerset	0.120 (0.307)	20.663*** (6.468)	-8.300 (5.701)
Bedfordshire	-0.722** (0.306)	62.051*** (6.433)	-11.401* (6.594)
Bexley	1.279*** (0.315)	27.470*** (6.864)	2.480 (6.402)
Birmingham	0.095 (0.303)	7.434 (6.347)	1.904 (5.516)
Blackburn with Darwen	0.926*** (0.317)	31.901*** (6.840)	-13.134** (6.349)
Blackpool	1.012*** (0.305)	12.552* (6.478)	-5.864 (5.810)
Bolton	0.922*** (0.315)	22.982*** (6.760)	-7.484 (6.277)
Bournemouth	0.748**	23.069***	-0.484

	CPA services (normalised)	Excess spending (£/cap)	Satisfaction (%)
	(6)	(9)	(13)
	(0.306)	(6.450)	(5.819)
Bracknell Forest	-0.023	-18.828***	-1.748
	(0.308)	(6.582)	(6.049)
Bradford	0.096	-9.771	-10.500*
	(0.305)	(6.468)	(5.768)
Brent	0.107	2.773	-4.826
	(0.303)	(6.390)	(5.555)
Brighton and Hove	0.068	13.840**	-6.703
	(0.303)	(6.390)	(5.618)
Bristol, City of	-0.626**	41.878***	-0.791
	(0.304)	(6.396)	(5.989)
Bromley	0.332	-15.821**	0.389
	(0.304)	(6.412)	(5.735)
Buckinghamshire	0.511	27.218***	-3.757
	(0.310)	(6.564)	(6.042)
Bury	0.881***	10.195	-3.735
	(0.303)	(6.424)	(5.571)
Calderdale	0.796***	27.616***	-7.677
	(0.304)	(6.450)	(5.934)
Cambridgeshire	0.263	8.612	-2.620
	(0.310)	(6.658)	(6.089)
Camden	1.224***	47.756***	-0.477
	(0.314)	(6.798)	(6.570)
Cheshire	1.150***	39.570***	7.684
	(0.305)	(6.498)	(5.948)
City of London	2.042***		
	(0.307)		
Cornwall	0.209	8.922	-4.905
	(0.307)	(6.488)	(5.678)
Coventry	-0.242	36.936***	-6.508
	(0.303)	(6.371)	(5.726)
Croydon	0.217	12.028*	-12.451**
	(0.310)	(6.601)	(5.971)
Cumbria	-0.403	45.540***	-1.870
	(0.306)	(6.417)	(6.191)
Darlington	0.693**	-4.500	-1.956
	(0.310)	(6.617)	(5.882)
Derby	0.907***	-6.833	-2.554
	(0.304)	(6.455)	(5.662)
Derbyshire	1.226***	44.196***	-3.577
	(0.312)	(6.815)	(6.556)
Devon	0.372	34.347***	-1.616
	(0.304)	(6.403)	(5.988)
Doncaster	-0.618**	-10.251	-10.572*
	(0.306)	(6.479)	(5.744)

	CPA services (normalised)	Excess spending (£/cap)	Satisfaction (%)
	(6)	(9)	(13)
Dorset	1.025*** (0.306)	59.727*** (6.535)	0.747 (6.867)
Dudley	1.090*** (0.302)	-12.600** (6.391)	-5.587 (5.752)
Durham	0.907*** (0.303)	51.402*** (6.438)	-2.137 (6.382)
Ealing	-0.320 (0.306)	18.982*** (6.441)	-9.570 (5.910)
East Riding of Yorkshire	1.300*** (0.306)	36.625*** (6.520)	0.685 (5.862)
East Sussex	0.477 (0.303)	62.363*** (6.390)	12.338* (6.909)
Enfield	0.277 (0.306)	27.723*** (6.421)	-4.654 (5.884)
Essex	0.649** (0.305)	39.021*** (6.462)	-5.167 (6.049)
Gateshead	0.592* (0.313)	56.184*** (6.778)	11.298 (6.871)
Gloucestershire	0.515* (0.304)	33.738*** (6.428)	-6.576 (5.928)
Greenwich	-0.092 (0.306)	12.430* (6.466)	0.302 (5.682)
Hackney	-0.561* (0.308)	23.199*** (6.478)	-12.265** (5.811)
Halton	0.610* (0.312)	-14.219** (6.615)	-2.206 (5.913)
Hammersmith and Fulham	0.189 (0.313)	20.329*** (6.766)	-6.037 (5.992)
Hampshire	0.942*** (0.315)	15.220** (6.802)	2.308 (6.079)
Haringey	-0.208 (0.304)	48.412*** (6.403)	-6.776 (6.361)
Harrow	-0.276 (0.306)	57.023*** (6.416)	-5.552 (6.707)
Hartlepool	1.081*** (0.314)	58.599*** (6.854)	-2.776 (6.926)
Havering	-0.496 (0.302)	61.688*** (6.348)	-8.313 (6.640)
Herefordshire, County of	-0.353 (0.306)	19.224*** (6.467)	-9.266 (5.752)
Hertfordshire	0.713** (0.312)	25.988*** (6.770)	3.445 (6.183)
Hillingdon	0.227 (0.303)	40.450*** (6.346)	0.355 (6.060)
Hounslow	0.503* (0.303)	55.386*** (6.346)	-6.440 (6.060)

	CPA services (normalised)	Excess spending (£/cap)	Satisfaction (%)
	(6)	(9)	(13)
	(0.305)	(6.446)	(6.537)
Isle of Wight	0.276	30.197***	-2.134
	(0.303)	(6.358)	(5.824)
Islington	-0.376	2.489	-8.298
	(0.303)	(6.390)	(5.546)
Kensington and Chelsea	1.713***	-72.121***	8.800
	(0.314)	(6.854)	(7.595)
Kent	0.744**	22.612***	2.867
	(0.317)	(6.825)	(6.239)
Kingston upon Hull, City of	-0.014	-3.606	8.629
	(0.309)	(6.542)	(5.815)
Kingston upon Thames	0.913***	77.260***	6.398
	(0.305)	(6.484)	(7.436)
Kirklees	1.156***	16.219**	-7.268
	(0.307)	(6.561)	(5.831)
Knowsley	0.928***	21.652***	5.267
	(0.312)	(6.687)	(6.131)
Lambeth	-0.742**	-26.238***	-11.101*
	(0.304)	(6.427)	(5.746)
Lancashire	0.837***	47.690***	-3.922
	(0.307)	(6.508)	(6.310)
Leeds	0.801***	-16.126**	6.790
	(0.305)	(6.481)	(5.771)
Leicester	0.420	2.800	-4.731
	(0.304)	(6.445)	(5.710)
Leicestershire	1.414***	25.188***	-5.893
	(0.303)	(6.483)	(5.817)
Lewisham	0.178	13.767**	-4.274
	(0.307)	(6.496)	(5.769)
Lincolnshire	-0.165	6.608	-6.549
	(0.303)	(6.359)	(5.547)
Liverpool	-0.406	46.746***	2.248
	(0.303)	(6.415)	(5.961)
Luton	-0.039	-12.392*	-11.400**
	(0.307)	(6.466)	(5.752)
Manchester	-0.356	12.155*	1.553
	(0.306)	(6.467)	(5.664)
Medway	0.054	-32.382***	-9.705
	(0.304)	(6.396)	(5.930)
Merton	0.105	39.824***	-2.490
	(0.303)	(6.367)	(5.879)
Middlesbrough	0.987***	12.152*	-2.506
	(0.312)	(6.639)	(6.007)
Milton Keynes	-0.208	-0.619	-2.184
	(0.302)	(6.348)	(5.506)

	CPA services (normalised)	Excess spending (£/cap)	Satisfaction (%)
	(6)	(9)	(13)
Newcastle upon Tyne	0.262 (0.304)	49.165*** (6.438)	6.252 (6.130)
Newham	-0.039 (0.306)	-14.558** (6.465)	-2.353 (5.747)
Norfolk	0.187 (0.306)	27.704*** (6.468)	-3.237 (5.923)
Northamptonshire	-0.184 (0.305)	9.096 (6.416)	-13.365*** (5.653)
North East Lincolnshire	-0.957*** (0.306)	41.050*** (6.458)	-3.332 (6.028)
North Lincolnshire	1.042*** (0.303)	49.204*** (6.464)	-3.491 (6.105)
North Somerset	0.642** (0.306)	13.621** (6.441)	-4.203 (5.771)
North Tyneside	0.489 (0.308)	40.904*** (6.481)	1.095 (6.042)
Northumberland	0.517* (0.307)	52.787*** (6.477)	-4.928 (6.299)
North Yorkshire	1.308*** (0.305)	31.871*** (6.486)	0.150 (6.053)
Nottingham	0.371 (0.305)	32.529*** (6.433)	-1.121 (5.915)
Nottinghamshire	0.946*** (0.306)	68.089*** (6.487)	-5.326 (6.894)
Oldham	-0.092 (0.308)	50.338*** (6.463)	-14.577*** (6.291)
Oxfordshire	0.551* (0.305)	32.232*** (6.422)	-3.278 (5.955)
Peterborough	0.241 (0.305)	3.151 (6.412)	-7.182 (5.631)
Plymouth	-0.641** (0.306)	-4.095 (6.476)	-5.292 (5.684)
Poole	0.735** (0.303)	12.992** (6.422)	3.220 (5.623)
Portsmouth	-0.191 (0.307)	-19.907*** (6.465)	-7.256 (5.737)
Reading	-0.461 (0.305)	39.867*** (6.435)	-6.020 (5.974)
Redbridge	1.004*** (0.304)	14.380** (6.452)	-4.675 (5.734)
Redcar and Cleveland	0.882*** (0.303)	42.545*** (6.428)	-8.460 (5.875)
Richmond upon Thames	0.735** (0.305)	103.482*** (6.454)	9.727 (8.308)
Rochdale	0.239	18.908***	-11.128*

	CPA services (normalised)	Excess spending (£/cap)	Satisfaction (%)
	(6)	(9)	(13)
	(0.305)	(6.419)	(5.685)
Rotherham	0.855***	14.492**	-4.083
	(0.305)	(6.448)	(5.729)
Rutland	0.272	90.630***	-1.210
	(0.306)	(6.425)	(7.515)
Salford	0.208	53.011***	-1.668
	(0.302)	(6.337)	(6.158)
Sandwell	-0.342	11.859*	1.188
	(0.304)	(6.387)	(5.542)
Sefton	0.230	35.399***	0.199
	(0.305)	(6.404)	(5.832)
Sheffield	-0.125	32.264***	-7.275
	(0.317)	(6.797)	(6.301)
Shropshire	1.312***	37.633***	-2.148
	(0.311)	(6.690)	(6.706)
Slough	-0.138	-20.726***	-14.041**
	(0.306)	(6.466)	(5.780)
Solihull	0.756**	-11.109*	6.415
	(0.303)	(6.418)	(5.652)
Somerset	1.235***	25.469***	-3.184
	(0.305)	(6.490)	(5.921)
Southampton	0.338	7.650	-1.177
	(0.305)	(6.444)	(5.634)
Southend-on-Sea	-0.221	-31.822***	-11.032*
	(0.305)	(6.471)	(6.055)
South Gloucestershire	0.372	28.156***	-2.879
	(0.309)	(6.508)	(5.999)
South Tyneside	0.837***	26.571***	3.876
	(0.303)	(6.388)	(5.780)
Southwark	0.443	-3.343	-6.476
	(0.303)	(6.376)	(5.630)
Staffordshire	0.682**	15.755**	-2.229
	(0.306)	(6.443)	(5.742)
St. Helens	0.945***	32.236***	-1.429
	(0.312)	(6.697)	(6.245)
Stockport	0.510*	42.960***	0.535
	(0.307)	(6.480)	(6.043)
Stockton on Tees	1.120***	20.292***	-0.321
	(0.314)	(6.780)	(6.220)
Stoke-on-Trent	-0.895***	-6.628	-16.495***
	(0.306)	(6.511)	(5.694)
Suffolk	0.678**	31.095***	-6.476
	(0.312)	(6.686)	(6.490)
Sunderland	1.292***	2.412	-7.463
	(0.314)	(6.834)	(6.179)

	CPA services (normalised)	Excess spending (£/cap)	Satisfaction (%)
	(6)	(9)	(13)
Surrey	0.356 (0.303)	34.572*** (6.416)	-0.146 (6.018)
Sutton	0.813** (0.317)	28.105*** (6.820)	2.157 (6.322)
Swindon	-0.330 (0.305)	0.418 (6.433)	-6.373 (5.541)
Tameside	0.795** (0.314)	15.497** (6.754)	-4.855 (5.991)
Telford and Wrekin	0.721** (0.317)	10.841 (6.825)	-6.315 (6.114)
Thurrock	-0.011 (0.303)	-19.353*** (6.350)	-8.013 (5.652)
Torbay	0.033 (0.308)	15.201** (6.464)	-8.801 (5.691)
Tower Hamlets	1.020*** (0.306)	-32.130*** (6.481)	-17.543*** (6.146)
Trafford	0.312 (0.303)	-36.363*** (6.367)	-2.755 (6.083)
Wakefield	0.751** (0.305)	-15.863** (6.468)	-7.945 (5.741)
Walsall	-0.283 (0.315)	33.494*** (6.676)	-12.320** (5.881)
Waltham Forest	-0.191 (0.306)	38.654*** (6.435)	-8.244 (6.133)
Wandsworth	1.495*** (0.314)	-199.959*** (6.816)	-2.538 (13.387)
Warrington	0.966*** (0.307)	-13.653** (6.518)	-4.374 (5.854)
Warwickshire	0.609** (0.304)	37.148*** (6.432)	1.187 (5.976)
West Berkshire	0.349 (0.305)	46.649*** (6.407)	7.204 (6.083)
Westminster	1.647*** (0.314)	-256.482*** (6.857)	-11.572 (17.123)
West Sussex	1.073*** (0.304)	27.560*** (6.479)	-2.404 (5.983)
Wigan	0.945*** (0.317)	15.865** (6.846)	-9.000 (6.150)
Wiltshire	0.235 (0.303)	30.837*** (6.399)	1.181 (5.772)
Windsor and Maidenhead	-0.095 (0.305)	-9.998 (6.465)	-1.844 (5.851)
Wirral	0.741** (0.306)	36.659*** (6.453)	1.994 (5.832)
Wokingham	0.103	40.890***	2.797

	CPA services (normalised)	Excess spending (£/cap)	Satisfaction (%)
	(6)	(9)	(13)
	(0.306)	(6.468)	(5.954)
Wolverhampton	0.596*	30.375***	12.148**
	(0.306)	(6.443)	(5.785)
Worcestershire	1.471***	23.066***	-1.971
	(0.305)	(6.553)	(5.915)
York	0.499	-12.912**	-6.995
	(0.305)	(6.478)	(5.765)
Constant	-2.131***	33.298***	49.953***
	(0.251)	(5.739)	(5.429)
N	1,042	887	296
R ²	0.725	0.945	0.740
Adjusted R ²	0.676	0.934	0.463
Residual Std. Error	0.565	10.954	5.453
F Statistic	14.932***	81.425***	2.674***

*p < .1; **p < .05; ***p < .01

Base year is 2002. Base council is Barking and Dagenham.

Table 32: Performance regression models including council dummy variables

BIBLIOGRAPHY

-
- Aldhous, Peter and Jeremy Howard (2012). 'Specialist knowledge is useless and unhelpful'. In: *New Scientist* 2893. URL: <http://www.newscientist.com/article/mg21628930.400-specialist-knowledge-is-useless-and-unhelpful.html?full=true> (visited on 30/11/2014).
- Alfano, Giuseppe (2011). 'Adapting bureaucracy to the Internet. The case of Venice local government'. In: *Information Polity* 16.1, pp. 5–22. URL: <http://doi.org/10.3233/IP-2011-0226>.
- Andersen, Kim N., Helle Z. Henriksen, Rony Medaglia, James N. Danziger, Møyfrid K. Sannarnes and Mette Enemærke (2010). 'Fads and facts of e-government: a review of impacts of e-government (2003–2009)'. In: *International Journal of Public Administration* 33.11, pp. 564–579. URL: <http://doi.org/10.1080/01900692.2010.517724>.
- Andrews, Rhys (2004). 'Analysing deprivation and local authority performance: the implications for CPA'. In: *Public Money & Management* 24.1, pp. 19–26. URL: <http://doi.org/10.1111/j.1467-9302.2004.00389.x>.
- Andrews, Rhys, George A. Boyne, Jennifer Law and Richard M. Walker (2005). 'External constraints on local service standards: the case of Comprehensive Performance Assessment in English local government'. In: *Public Administration* 83.3, pp. 639–656. URL: <http://doi.org/10.1111/j.0033-3298.2005.00466.x>.
- Andrews, Rhys, George A. Boyne and Richard M. Walker (2006). 'Strategy content and organizational performance: an empirical analysis'. In: *Public Administration Review* 66.1, pp. 52–63. URL: <http://doi.org/10.1111/j.1540-6210.2006.00555.x>.

- Andrews, Rhys and Steven Van de Walle (2013). 'New Public Management and citizens' perceptions of local service efficiency, responsiveness, equity and effectiveness'. In: *Public Management Review* 15.3, pp. 762–783. URL: <http://doi.org/10.1080/14719037.2012.725757>.
- Apache Software Foundation (2013). *Apache Tika version 3.1.1*. URL: <http://tika.apache.org/> (visited on 11/07/2014).
- Audit Commission (2001a). *2000/01 district data*. URL: <http://archive.audit-commission.gov.uk/auditcommission/SiteCollectionDocuments/Downloads/Ladata00-01DISTRICTS.xls> (visited on 30/06/2014).
- (2001b). *2000/01 single tier and county council data*. URL: <http://archive.audit-commission.gov.uk/auditcommission/SiteCollectionDocuments/Downloads/Ladata00-01exclDistrictCouncils.xls> (visited on 30/06/2014).
- (2002a). *Comprehensive Performance Assessment: Scores and analysis of performance for single tier and county councils in England, December 2002*. London: Audit Commission, p. 20. URL: <http://archive.audit-commission.gov.uk/auditcommission/SiteCollectionDocuments/InspectionOutput/CPAScores2002.pdf> (visited on 04/06/2014).
- (2002b). *The final CPA assessment framework for single tier and county councils, 2002*. London: Audit Commission, p. 19. URL: <http://archive.audit-commission.gov.uk/auditcommission/SiteCollectionDocuments/Consultation/cpaAssFramework2002.pdf> (visited on 17/06/2014).
- (2003). *Comprehensive Performance Assessment: Scores and analysis of performance for single tier and county councils in England, 2003*. London: Audit Commission, p. 21. URL: <http://archive.audit-commission.gov.uk/auditcommission/SiteCollectionDocuments/InspectionOutput/CPAScores2003.pdf> (visited on 04/06/2014).
- (2004a). *2003/04 Best Value user satisfaction data*. URL: <http://archive.audit-commission.gov.uk/auditcommission/SiteCollectionDocuments/Downloads/UserSatisfactionDataAndQuartiles.xls> (visited on 30/06/2014).

- (2004b). *Comprehensive Performance Assessment: Scores and analysis of performance for single tier and county councils in England, 2004*. London: Audit Commission, p. 22. URL: <http://archive.audit-commission.gov.uk/auditcommission/SiteCollectionDocuments/InspectionOutput/CPAScores2004.pdf> (visited on 04/06/2014).
- (2005a). *Comprehensive Performance Assessment: Scores and analysis of performance for district councils in England, 2003/04*. London: Audit Commission, p. 44. URL: <http://archive.audit-commission.gov.uk/auditcommission/SiteCollectionDocuments/InspectionOutput/CPA06Sep2005REP.pdf> (visited on 31/05/2014).
- (2005b). *CPA – The Harder Test: Scores and analysis of performance in single tier and county councils 2005*. London: Audit Commission, p. 40. URL: <http://archive.audit-commission.gov.uk/auditcommission/SiteCollectionDocuments/InspectionOutput/cpascores2005.pdf> (visited on 04/06/2014).
- (2006). *CPA – The Harder Test: Scores and analysis of performance in single tier and county councils 2006*. London: Audit Commission, p. 60. URL: <http://archive.audit-commission.gov.uk/auditcommission/SiteCollectionDocuments/InspectionOutput/CPATheHarderTest2006.pdf> (visited on 04/06/2014).
- (2007a). *2006/07 user survey Best Value data*. URL: http://archive.audit-commission.gov.uk/auditcommission/SiteCollectionDocuments/Downloads/2006_07UserSurveyBVPIswithDistricts.xls (visited on 30/06/2014).
- (2007b). *CPA – The Harder Test: Scores and analysis of performance in single tier and county councils 2007*. London: Audit Commission, p. 56. URL: <http://archive.audit-commission.gov.uk/auditcommission/SiteCollectionDocuments/InspectionOutput/CPATheHarderTest2007.pdf> (visited on 04/06/2014).
- (2008). *CPA – The Harder Test: Scores and analysis of performance in single-tier and county councils 2008*. London: Audit Commission, p. 68. URL: <http://archive.audit-commission.gov.uk/auditcommission/SiteCollectionDocuments/AnnualReports/2009/05032009CPATheHarderTestREP.pdf> (visited on 08/05/2014).

- Baldersheim, Harald and Morten Øgård (2008). 'Innovation in e-government: analysis of municipal web pages in the Nordic countries'. In: *Information Polity* 13.3, pp. 125–137. URL: <http://content.iospress.com/articles/information-polity/ip000133> (visited on 01/04/2016).
- Barlow, John Perry (1996). *A declaration of the independence of cyberspace*. Electronic Frontier Foundation. URL: <https://www EFF.org/cyberspace-independence> (visited on 04/04/2016).
- Bekkers, Victor (2007). 'Modernization, public innovation and information and communication technologies: the emperor's new clothes?' In: *Information Polity* 12.3, pp. 103–107. URL: <http://content.iospress.com/articles/information-polity/ip000122> (visited on 01/04/2016).
- Bélanger, France and Lemuria Carter (2008). 'Trust and risk in e-government adoption'. In: *The Journal of Strategic Information Systems* 17.2, pp. 165–176. URL: <http://doi.org/10.1016/j.jsis.2007.12.002>.
- Benbasat, Izak, Albert S. Dexter, Donald H. Drury and Robert C. Goldstein (1984). 'A critique of the stage hypothesis: theory and empirical evidence'. In: *Communications of the ACM* 27.5, pp. 476–485. URL: <http://doi.org/10.1145/358189.358076>.
- Benoit, Kenneth and Michael Laver (2007). 'Benchmarks for text analysis: A response to Budge and Pennings'. In: *Electoral Studies* 26.1, pp. 130–135. URL: <http://doi.org/10.1016/j.electstud.2006.04.001>.
- Berrington, Ann, W.F. Smith and Patrick Sturgis (2006). *An overview of methods for the analysis of panel data*. NCRM/007. Southampton, UK: ESRC National Centre for Research Methods, p. 57. URL: <http://eprints.ncrm.ac.uk/415/1/MethodsReviewPaperNCRM-007.pdf> (visited on 11/06/2014).
- Beynon-Davies, Paul and Steve Martin (2004). 'Electronic local government and the modernisation agenda: progress and prospects for public service improvement'. In: *Local Government Studies* 30.2, pp. 214–229. URL: <http://doi.org/10.1080/0300303042000267245>.

- Bird, Steven, Ewan Klein and Edward Loper (2009). *Natural language processing with Python*. Sebastopol, CA: O'Reilly. ISBN: 978-0-596-51649-9. URL: <http://www.nltk.org/book>.
- Bonsón, Enrique, Lourdes Torres, Sonia Royo and Francisco Flores (2012). 'Local e-government 2.0: Social media and corporate transparency in municipalities'. In: *Government Information Quarterly* 29.2, pp. 123–132. URL: <http://doi.org/10.1016/j.giq.2011.10.001>.
- Boyne, George A., Oliver James, Peter John and Nicolai Petrovsky (2010). 'Does public service performance affect top management turnover?' In: *Journal of Public Administration Research and Theory* 20 (suppl 2), pp. i.261–i.279. URL: <http://doi.org/10.1093/jopart/muq024>.
- Bracken, Mike (2014). *Revealing the hidden side of transformation*. Government Digital Service. URL: <https://gds.blog.gov.uk/2014/12/17/revealing-the-hidden-side-of-transformation/> (visited on 17/12/2014).
- Brandsen, Taco and Victor Pestoff (2006). 'Co-production, the third sector and the delivery of public services'. In: *Public Management Review* 8.4, pp. 493–501. URL: <http://doi.org/10.1080/14719030601022874>.
- Braumoeller, Bear F. (2015). 'Guarding against false positives in Qualitative Comparative Analysis'. In: *Political Analysis* 23.4, pp. 471–487. URL: <http://doi.org/10.1093/pan/mpv017>.
- Bright, Jonathan and Tom Nicholls (2014). 'The life and death of political news: measuring the impact of the audience agenda using online data'. In: *Social Science Computer Review* 32.2, pp. 170–181. URL: <http://ora.ox.ac.uk/objects/uuid:5cb38355-7c82-4541-8504-4fdd07d21f2b> (visited on 25/11/2013).
- Brown, M. Maureen (2007). 'Understanding e-government benefits: an examination of leading-edge local governments'. In: *The American Review of Public Administration* 37.2, pp. 178–197. URL: <http://doi.org/10.1177/0275074006291635>.
- Brüderl, Josef (2005). 'Panel data analysis'. URL: <http://www.sowi.uni-mannheim.de/lessm/veranst/Panelanalyse.pdf>.

- Brügger, Niels (2013). 'Historical network analysis of the Web'. In: *Social Science Computer Review* 31.3, pp. 306–321. URL: <http://doi.org/10.1177/0894439312454267>.
- Brynjolfsson, Erik and Lorin M. Hitt (2003). 'Computing productivity: firm-level evidence'. In: *The Review of Economics and Statistics* 85.4, pp. 793–808. URL: <http://www.jstor.org/stable/3211806> (visited on 09/04/2015).
- Cabinet Office (1999a). *Central local information age government concordat*. URL: https://web.archive.org/web/20030811015635/http://www.e-envoy.gov.uk/Resources/OtherPublicationsArticle/fs/en?CONTENT_ID=4000244&chk=SiEqeR (visited on 04/04/2016).
- (1999b). *Modernising government*. Cm. 4310. London: TSO. 66 pp.
- (2005). *Transformational government, enabled by technology: annual report 2006*. Cm. 6970. London: TSO. 52 pp. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/228602/6970.pdf.
- (2013). *List of .gov.uk domain names as at 1 October 2013*. URL: <https://www.gov.uk/government/publications/list-of-gov-uk-domain-names> (visited on 27/11/2013).
- Carter, Lemuria and France Bélanger (2005). 'The utilization of e-government services: citizen trust, innovation and acceptance factors'. In: *Information Systems Journal* 15.1, pp. 5–25. URL: <http://doi.org/10.1111/j.1365-2575.2005.00183.x>.
- Chadwick, Andrew and Christopher May (2003). 'Interaction between states and citizens in the age of the Internet: "e-Government" in the United States, Britain, and the European Union'. In: *Governance* 16.2, pp. 271–300. URL: <http://doi.org/10.1111/1468-0491.00216>.
- Chapman, Richard A. (1999). 'The importance of Modernising Government'. In: *Teaching Public Administration* 19.1, pp. 1–18. URL: <http://tpa.sagepub.com/content/19/1/1> (visited on 06/10/2013).

- Clark, Tom S. and Drew A. Linzer (2015). 'Should i use fixed or random effects?' In: *Political Science Research and Methods* 3.2, pp. 399–408. URL: http://journals.cambridge.org/article_S2049847014000326 (visited on 21/07/2015).
- Clarke, Michael and John Stewart (1994). 'The local authority and the new community governance'. In: *Local Government Studies* 20.2, pp. 163–176. URL: <http://doi.org/10.1080/03003939408433719>.
- Conte, R., N. Gilbert, G. Bonelli, C. Cioffi-Revilla, G. Deffuant, J. Kertesz, V. Loreto, S. Moat, J.-P. Nadal, A. Sanchez, A. Nowak, A. Flache, M. San Miguel and D. Helbing (2012). 'Manifesto of computational social science'. In: *The European Physical Journal Special Topics* 214.1, pp. 325–346. URL: <http://doi.org/10.1140/epjst/e2012-01697-8>.
- Cook, Chris (2016). *Every school to become an academy, ministers to announce*. BBC News. URL: <http://www.bbc.co.uk/news/education-35814215> (visited on 28/03/2016).
- Copeland, Eddie, Sarah Fink and Cameron Scott (2014). *Technology manifesto*. London: Policy Exchange, p. 32. URL: <http://www.policyexchange.org.uk/images/publications/technology%20manifesto.pdf> (visited on 01/04/2016).
- Cordella, Antonio and Federico Iannacci (2010). 'Information systems in the public sector: the e-Government enactment framework'. In: *The Journal of Strategic Information Systems* 19.1, pp. 52–66. URL: <http://doi.org/10.1016/j.jsis.2010.01.001>.
- Cullen, Rowena and Caroline Houghton (2000). 'Democracy online: an assessment of New Zealand government web sites'. In: *Government Information Quarterly* 17.3, pp. 243–267. URL: [http://doi.org/10.1016/S0740-624X\(00\)00033-2](http://doi.org/10.1016/S0740-624X(00)00033-2).
- Czerniawski, Stefan (2014). *Locally centralising the centrally local*. Public Strategist. URL: <http://publicstrategist.com/2014/07/locally-centralising-the-centrally-local/> (visited on 27/07/2014).
- Daigle, Leslie (2004). *RFC 3912: WHOIS protocol specification*. URL: <https://tools.ietf.org/html/rfc3912> (visited on 06/02/2016).

- Dan, Sorin and Christopher Pollitt (2015). 'NPM can work: An optimistic review of the impact of New Public Management reforms in central and eastern Europe'. In: *Public Management Review* 17.9, pp. 1305–1332. URL: <http://doi.org/10.1080/14719037.2014.908662>.
- Danziger, James N., William H. Dutton, R. Kling and K.L. Kraemer (1982). *Computers and politics: high technology in American local governments*. New York: Columbia University Press. 280 pp. ISBN: 0-231-04888-2.
- DCLG (2007). *Indices of Deprivation 2007*. URL: <http://webarchive.nationalarchives.gov.uk/20100410180038/http://communities.gov.uk/communities/neighbourhoodrenewal/deprivation/deprivation07/> (visited on 23/05/2014).
- (2010a). *English indices of deprivation 2010*. URL: <https://www.gov.uk/government/publications/english-indices-of-deprivation-2010> (visited on 23/05/2014).
- (2010b). *Local authority revenue expenditure and financing England: 2007 to 2008 individual local authority data*. URL: <https://www.gov.uk/government/statistics/local-authority-revenue-expenditure-and-financing-england-2007-to-2008-individual-local-authority-data> (visited on 15/03/2016).
- (2012). *Calculating the 2013-14 Formula Funding*. London: Department for Communities and Local Government, p. 157. URL: <http://www.localcommunities.gov.uk/finance/1314/1314LGFRFormFunding.pdf> (visited on 18/06/2014).
- (2013). *Single Data List 2014/15*. London: Department for Communities and Local Government, p. 5. URL: <https://www.gov.uk/government/policies/making-local-councils-more-transparent-and-accountable-to-local-people/supporting-pages/single-data-list> (visited on 08/05/2014).
- (2016). *Local authority revenue expenditure and financing in England: 2014 to 2015 final outturn - Publications*. URL: <https://www.gov.uk/government/statistics/local-authority-revenue-expenditure-and-financing-in-england-2014-to-2015-final-outturn> (visited on 15/03/2016).

- DeLone, William H. and Ephraim R. McLean (2003). 'The DeLone and McLean model of information systems success: a ten-year update'. In: *Journal of Management Information Systems* 19.4, pp. 9–30. URL: <http://doi.org/10.1080/07421222.2003.11045748>.
- Department of the Environment (1990). *The Revenue Support Grant Distribution Report (England)*. HC 049. London: HMSO. 44 pp. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/235799/0049.pdf (visited on 18/06/2014).
- Detlor, Brian, Maureen E. Hupfer, Umar Ruhi and Li Zhao (2013). 'Information quality and community municipal portal use'. In: *Government Information Quarterly* 30.1, pp. 23–32. URL: <http://doi.org/10.1016/j.giq.2012.08.004>.
- DETR (1999). *Best Value and Audit Commission Performance Indicators for 2000/2001 Volume One: The Performance Indicators including The Publication of Information Direction 1999 (England)*. London: Department of the Environment, Transport and the Regions. 130 pp. URL: <http://archive.audit-commission.gov.uk/auditcommission/subwebs/publications/studies/studyPDF/NEW1040.pdf> (visited on 30/06/2014).
- Diebold, Francis X. (2014). 'Econometrics: streamlined, applied and e-aware'. URL: <http://www.ssc.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>.
- D'Orazio, Vito, Steven T. Landis, Glenn Palmer and Philip Schrodtt (2014). 'Separating the wheat from the chaff: applications of automated document classification using Support Vector Machines'. In: *Political Analysis* 22.2, pp. 224–242. URL: <http://pan.oxfordjournals.org/content/22/2/224> (visited on 10/03/2015).
- Drechsler, Wolfgang and Tiina Randma-Liiv (2015). 'In some Central and Eastern European countries, some NPM tools may sometimes work: a reply to Dan and Pollitt's "NPM can work"'. In: *Public Management Review*, pp. 1–7. URL: <http://doi.org/10.1080/14719037.2015.1114137>.
- DTLR (2000). *Indices of deprivation 2000*. URL: <http://webarchive.nationalarchives.gov.uk/20120919132719/http://www.communities.gov.uk/archived/general->

- content/communities/indicesofdeprivation/indicesofdeprivation/ (visited on 28/06/2014).
- DTLR (2001). *Strong local leadership, quality public services*. Cm. 5327. London: TSO. 139 pp. ISBN: 978-0-10-153272-3. URL: <http://web.archive.org/web/20030211042039/http://www.local-regions.odpm.gov.uk/sll/index.htm> (visited on 28/06/2014).
- Duivenboden, Hein van and Marcel Thaens (2008). 'ICT-driven innovation and the culture of public administration: a contradiction in terms?' In: *Information Polity* 13.3, pp. 213–232. URL: <http://content.iospress.com/articles/information-polity/ip000157> (visited on 05/02/2016).
- Dunleavy, Patrick (2010a). 'New Worlds in Political Science'. In: *Political Studies* 58.2, pp. 239–265. URL: <http://doi.org/10.1111/j.1467-9248.2009.00834.x>.
- (2010b). *What is the Cameron-Clegg governance strategy? Zombie 'new public management' cannot work in the face of massive public expenditure cutbacks*. British Politics and Policy at LSE. URL: <http://blogs.lse.ac.uk/politicsandpolicy/archives/4313> (visited on 02/12/2013).
- (2016). "'Big data" and policy learning'. In: *Methods that matter: social science and evidence-based policymaking*. Ed. by Gerry Stoker and Mark Evans. Forthcoming. Bristol: The Policy Press. URL: https://www.academia.edu/23754121/Big_data_and_policy_learning (visited on 31/03/2016).
- Dunleavy, Patrick and Leandro Carrera (2013). *Growing the productivity of government services*. Cheltenham: Edward Elgar. 365 pp. ISBN: 978-0-85793-498-7.
- Dunleavy, Patrick and Helen Margetts (2015). 'Design principles for Essentially Digital Governance'. In: 111th Annual Meeting of the American Political Science Association. San Francisco, CA. URL: https://www.researchgate.net/publication/282867465_Design_Principles_for_Essentially_Digital_Governance (visited on 19/04/2016).
- Dunleavy, Patrick, Helen Margetts, Simon Bastow and Jane Tinkler (2006a). *Digital Era Governance: IT corporations, the state, and e-government*. Oxford: Oxford University Press. 289 pp.

- (2006b). ‘New Public Management is dead – long live Digital-Era Governance’. In: *Journal of Public Administration Research and Theory* 16.3, pp. 467–494. URL: <http://doi.org/10.1093/jopart/mui057>.
- Dutton, William H. and Grant Blank (2011). *Next Generation Users: The Internet in Britain*. OXIS 2011. Oxford: Oxford Internet Institute, p. 64. URL: http://oxis.oii.ox.ac.uk/sites/oxis.oii.ox.ac.uk/files/content/files/publications/oxis2011_report.pdf (visited on 03/07/2014).
- Dutton, William H., Grant Blank and Darja Groselj (2013). *Cultures of the Internet: The Internet in Britain*. OXIS 2013. Oxford: Oxford Internet Institute, p. 64. URL: http://oxis.oii.ox.ac.uk/sites/oxis.oii.ox.ac.uk/files/content/files/publications/OxIS_2013.pdf (visited on 03/07/2014).
- Dutton, William H., Corinna di Gennaro and Andrea Millwood Hargrave (2005). *The Internet in Britain*. OXIS 2005. Oxford: Oxford Internet Institute, p. 60. URL: http://oxis.oii.ox.ac.uk/sites/oxis.oii.ox.ac.uk/files/content/files/publications/oxis2005_report.pdf (visited on 03/07/2014).
- Dutton, William H. and Ellen J. Helsper (2007). *The Internet in Britain: 2007*. OXIS 2007. Oxford: Oxford Internet Institute, p. 83. URL: http://oxis.oii.ox.ac.uk/sites/oxis.oii.ox.ac.uk/files/content/files/publications/oxis2007_report.pdf (visited on 03/07/2014).
- Dutton, William H., Ellen J. Helsper and Monica M. Gerber (2009). *The Internet in Britain: 2009*. OXIS 2009. Oxford: Oxford Internet Institute, p. 75. URL: http://oxis.oii.ox.ac.uk/sites/oxis.oii.ox.ac.uk/files/content/files/publications/oxis2009_report.pdf (visited on 03/07/2014).
- Ellison, Nick and Michael Hardey (2014). ‘Social media and local government: citizenship, consumption and democracy’. In: *Local Government Studies* 40.1, pp. 21–40. URL: <http://doi.org/10.1080/03003930.2013.799066>.
- Escher, Tobias, Helen Margetts, Ingemar Cox and Vaclav Petricek (2006). ‘Governing from the centre: comparing the nodality of digital governments’. In: Annual Meeting of the American Political Science Association (APSA). Philadelphia, Pennsylvania. URL: <http://doi.org/10.2139/ssrn.1755762>.

- Fishenden, Jerry and Mark Thompson (2013). 'Digital government, open architecture, and innovation: why public sector IT will never be the same again'. In: *Journal of Public Administration Research and Theory* 23.4, pp. 977–1004. URL: <http://doi.org/10.1093/jopart/mus022>.
- Foreshew-Cain, Stephen (2015). *How digital and technology transformation saved £1.7bn last year*. URL: <https://gds.blog.gov.uk/2015/10/23/how-digital-and-technology-transformation-saved-1-7bn-last-year/> (visited on 23/10/2015).
- Fountain, Jane E. (2001). *Building the virtual state : information technology and institutional change*. Washington, DC: Brookings Institution Press. 251 pp. ISBN: 0-8157-0078-4.
- Game, Chris (2006). 'Comprehensive performance assessment in English local government'. In: *International Journal of Productivity and Performance Management* 55.6, pp. 466–479. URL: <http://doi.org/10.1108/17410400610682497>.
- Ganapati, Sukumar and Christopher G. Reddick (2012). 'Open e-government in US state governments: survey evidence from Chief Information Officers'. In: *Government Information Quarterly* 29.2, pp. 115–122. URL: <http://doi.org/10.1016/j.giq.2011.09.006>.
- General Services Administration (2014). *DAP: Digital Analytics Program*. URL: <http://www.digitalgov.gov/services/dap/> (visited on 10/07/2014).
- Georgiou, Ion (2014). 'Seeing the forest for the trees: an atlas of the politics–administration dichotomy'. In: *Public Administration Review* 74.2, pp. 156–175. URL: <http://doi.org/10.1111/puar.12163>.
- Gershon, Peter (2004). *Releasing resources for the front line: independent review of public sector efficiency*. London: HM Treasury. 66 pp. ISBN: 1-84532-032-8. URL: http://webarchive.nationalarchives.gov.uk/20100407181317/http://www.hm-treasury.gov.uk/d/efficiency_review120704.pdf (visited on 26/06/2014).
- Government Digital Service (2014). *Digital Transformation*. URL: <https://www.gov.uk/transformation> (visited on 02/08/2014).

- gov.uk (2016). *List of councils in England*. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/491463/List_of_councils_in_England.pdf (visited on 03/11/2016).
- GRASS Development Team (2015). *Geographic Resources Analysis Support System (GRASS GIS) software, Version 7.0.2*. USA: Open Source Geospatial Foundation. URL: <http://grass.osgeo.org>.
- Grimmer, Justin (2011). ‘An introduction to Bayesian inference via variational approximations’. In: *Political Analysis* 19.1, pp. 32–47. URL: <http://doi.org/10.1093/pan/mpq027>.
- Grimmer, Justin and Brandon M. Stewart (2013). ‘Text as data: the promise and pitfalls of automatic content analysis methods for political texts’. In: *Political Analysis* 21.3, pp. 267–297. URL: <http://doi.org/10.1093/pan/mps028>.
- Gulati, Girish J. “Jeff”, Christine B. Williams and David J. Yates (2014). ‘Predictors of on-line services and e-participation: a cross-national comparison’. In: *Government Information Quarterly* 31.4, pp. 526–533. URL: <http://doi.org/10.1016/j.giq.2014.07.005>.
- Hale, Scott A., Helen Margetts and Taha Yasseri (2013). ‘Petition growth and success rates on the UK No. 10 Downing Street website’. In: *Proceedings of the 5th Annual ACM Web Science Conference*. WebSci ’13. New York, NY, USA: ACM, pp. 132–138. ISBN: 978-1-4503-1889-1. URL: <http://doi.org/10.1145/2464464.2464518>.
- Hale, Scott A., Taha Yasseri, Josh Cowls, Eric T. Meyer, Ralph Schroeder and Helen Margetts (2014). ‘Mapping the UK webspace: fifteen years of British universities on the web’. In: *Proceedings of the 2014 ACM Conference on Web Science*. WebSci ’14. New York, NY, USA: ACM, pp. 62–70. URL: <http://doi.org/10.1145/2615569.2615691>.
- Hanneman, Robert and Mark Riddle (2005). ‘Introduction to social network methods’. Online textbook. University of California, Riverside. URL: http://faculty.ucr.edu/~hanneman/nettext/Introduction_to_Social_Network_Methods.pdf (visited on 20/03/2015).

- Heeks, Richard and Savita Bailur (2007). 'Analyzing e-government research: perspectives, philosophies, theories, methods, and practice'. In: *Government Information Quarterly* 24.2, pp. 243–265. URL: <http://doi.org/10.1016/j.giq.2006.06.005>.
- Henman, Paul, Rob Ackland and Tim Graham (2014a). 'Community structure in e-Government hyperlink networks'. In: *Proceedings of the European Conference on e-Government*. 14th European Conference on e-Government. Braov, Romania, pp. 135–143. ISBN: 978-1-909507-36-4. URL: <http://espace.library.uq.edu.au/view/UQ:331454> (visited on 05/08/2014).
- (2014b). 'Social policy on the Web: the online institutional structure of social policy domains in the UK'. In: *Proceedings of the European Conference on e-Government*. 14th European Conference on e-Government. Braov, Romania, pp. 126–134. ISBN: 978-1-909507-36-4. URL: <http://espace.library.uq.edu.au/view/UQ:331474> (visited on 05/08/2014).
- Hoffman, Matthew, Francis R. Bach and David M. Blei (2010). 'Online learning for latent dirichlet allocation'. In: *Advances in neural information processing systems*, pp. 856–864. URL: <http://papers.nips.cc/paper/3902-online-learning-for-latentdirichlet-allocation!> (visited on 24/03/2016).
- Holmberg, Kim (2009). *Webometric network analysis: mapping cooperation and geopolitical connections between local government administration on the web*. Åbo: Åbo Akademis Förlag. 291 pp. ISBN: 978-951-765-510-1.
- Hood, Christopher and Ruth Dixon (2015). *A government that worked better and cost less?: evaluating three decades of reform and change in UK central government*. Oxford: Oxford University Press. ISBN: 978-0-19-968702-2.
- Hood, Christopher and Helen Margetts (2007). *The tools of government in the digital age*. 2nd edition. Basingstoke: Palgrave Macmillan. 232 pp. ISBN: 978-0-230-00144-2.
- International Internet Preservation Consortium (2014). *OpenWayback*. Version 2.0.0-beta. URL: <http://www.netpreserve.org/netpreserve.org/tools/openwayback>.

- Internet Archive (2011). *Basic Crawl Job Settings – Heritrix*. URL: <https://web.archive.jira.com/wiki/display/Heritrix/Basic+Crawl+Job+Settings> (visited on 10/06/2014).
- (2013). *Heritrix 3.1.1*. URL: <http://builds.archive.org:8080/maven2/org/archive/heritrix/heritrix/3.1.1/>.
- (2014). *Release Notes – Heritrix 3.2.0*. URL: <https://web.archive.jira.com/wiki/display/Heritrix/Release+Notes+-+Heritrix+3.2.0> (visited on 06/07/2014).
- Irani, Zahir, Peter E.D. Love, Tony Elliman, Steve Jones and Marinos Themistocleous (2005). ‘Evaluating e-government: learning from the experiences of two UK local authorities’. In: *Information Systems Journal* 15.1, pp. 61–82. URL: <http://doi.org/10.1111/j.1365-2575.2005.00186.x>.
- ISO (2008). ‘ISO 28500: Information and documentation – the WARC file format’. Working draft. Geneva. URL: http://bibnum.bnf.fr/warc/WARC_ISO_28500_version1_latestdraft.pdf (visited on 24/06/2014).
- James, Oliver and Peter John (2007). ‘Public management at the ballot box: performance information and electoral support for incumbent English local governments’. In: *Journal of Public Administration Research and Theory* 17.4, pp. 567–580. URL: <http://doi.org/10.1093/jopart/mul020>.
- Jeffreys, Branwyn (2016). *Academy plans create unexpected allies*. BBC News. URL: <http://www.bbc.co.uk/news/education-35899483> (visited on 28/03/2016).
- Joseph, Rhoda C. (2013). ‘A structured analysis of e-government studies: trends and opportunities’. In: *Government Information Quarterly* 30.4, pp. 435–440. URL: <http://doi.org/10.1016/j.giq.2013.05.006>.
- Kapucu, Naim, Qian Hu and Sana Khosa (2014). ‘The state of network research in public administration’. In: *Administration & Society* early access. URL: <http://doi.org/10.1177/0095399714555752>.
- Krackardt, David (1987). ‘QAP partialling as a test of spuriousness’. In: *Social Networks* 9.2, pp. 171–186. URL: [http://doi.org/10.1016/0378-8733\(87\)90012-8](http://doi.org/10.1016/0378-8733(87)90012-8).

- Kraemer, Kenneth L., William H. Dutton and Alana Northrop (1981). *The management of information systems*. New York: Columbia University Press. 416 pp. ISBN: 0-231-04886-6.
- Kraemer, Kenneth L. and John Leslie King (2003). *Information technology and administrative reform: will the time after e-government be different?* Irvine, CA: Center for Research on Information Technology and Organizations, UC Irvine, p. 19. URL: <http://repositories.cdlib.org/crito/government/337>.
- Krippendorff, Klaus (2004). 'Reliability in content analysis'. In: *Human Communication Research* 30.3, pp. 411–433. URL: <http://doi.org/10.1111/j.1468-2958.2004.tb00738.x>.
- Krogslund, Chris, Donghyun Danny Choi and Mathias Poertner (2014). 'Fuzzy sets on shaky ground: parameter sensitivity and confirmation bias in fsQCA'. In: *Political Analysis*, mpu016. URL: <http://pan.oxfordjournals.org/content/early/2014/11/04/pan.mpu016> (visited on 07/11/2014).
- Laffin, Martin and Christianne Ormston (2013). 'Disconnected communities? ICT, policy learning and the lessons for central–local relations'. In: *Public Money & Management* 33.3, pp. 185–191. URL: <http://doi.org/10.1080/09540962.2013.785703>.
- Lambert, Frank (2013). 'Seeking electronic information from government resources: a comparative analysis of two communities' web searching of municipal government websites'. In: *Government Information Quarterly* 30.1, pp. 99–109. URL: <http://doi.org/10.1016/j.giq.2012.07.007>.
- Lancichinetti, Andrea and Santo Fortunato (2009). 'Community detection algorithms: a comparative analysis'. In: *Physical Review E* 80.56117, pp. 1–11. URL: <http://doi.org/10.1103/PhysRevE.80.056117>.
- Larsson, Hannu and Åke Grönlund (2014). 'Future-oriented eGovernance: the sustainability concept in eGov research, and ways forward'. In: *Government Information Quarterly* 31.1, pp. 137–149. URL: <http://doi.org/10.1016/j.giq.2013.07.004>.

- Laver, Michael, Kenneth Benoit and John Garry (2003). 'Extracting policy positions from political texts using words as data'. In: *American Political Science Review* 97.2, pp. 311–331. URL: <http://doi.org/10.1017/S0003055403000698>.
- Layne, Karen and Jungwoo Lee (2001). 'Developing fully functional e-government: A four stage model'. In: *Government Information Quarterly* 18.2, pp. 122–136. URL: [http://doi.org/10.1016/S0740-624X\(01\)00066-1](http://doi.org/10.1016/S0740-624X(01)00066-1).
- Lecy, Jesse D., Ines A. Mergel and Hans Peter Schmitz (2014). 'Networks in public administration: current scholarship in review'. In: *Public Management Review* 16.5, pp. 643–665. URL: <http://doi.org/10.1080/14719037.2012.743577>.
- Lehmberg, Oliver, Robert Meusel and Christian Bizer (2014). 'Graph structure in the Web: aggregated by pay-level domain'. In: *Proceedings of the 2014 ACM Conference on Web Science*. WebSci '14. New York, NY, USA: ACM, pp. 119–128. ISBN: 978-1-4503-2622-3. URL: <http://doi.org/10.1145/2615569.2615674>.
- Lenk, Klaus (2007). 'Reconstructing public administration theory from below'. In: *Information Polity* 12.4, pp. 207–212. URL: <http://content.iospress.com/articles/information-polity/ip000126> (visited on 01/04/2016).
- LGIU (2014). *Technology and transformation in town halls*. London: Local Government Information Unit, p. 28. URL: <http://www.lgiu.org.uk/wp-content/uploads/2014/07/LGIU-Tech-Report-A5-v3b2.pdf> (visited on 15/07/2014).
- Linders, Dennis (2012). 'From e-government to we-government: defining a typology for citizen coproduction in the age of social media'. In: *Government Information Quarterly* 29.4, pp. 446–454. URL: <http://doi.org/10.1016/j.giq.2012.06.003>.
- Lipton, Jacqueline (2003). 'Mixed metaphors in cyberspace: property in information and information systems'. In: *Loyola University Chicago Law Journal* 35, p. 235. URL: <http://ssrn.com/abstract=493642> (visited on 11/07/2014).
- Local Data Panel (2010). *Publishing itemised local authority expenditure - advice for comment*. URL: <https://data.gov.uk/blog/publishing-itemised-local-authority-expenditure-advice-comment> (visited on 02/02/2016).

- Local Democracy, Economic Development and Construction Act 2009*. URL: <http://www.legislation.gov.uk/ukpga/2009/20/> (visited on 10/01/2016).
- Local Government Act 1999*. URL: <http://www.legislation.gov.uk/ukpga/1999/27/> (visited on 26/06/2014).
- Local Government Association (2006). *Local government reputation campaign: delivering for people and places*. London: Local Government Association. ISBN: 1-84049-522-7. URL: <http://www.reputation.lga.gov.uk/lga/aio/251829> (visited on 30/06/2014).
- (2014). *Transforming public services using technology and digital approaches*. L14-303. London: Local Government Association, p. 44. URL: www.local.gov.uk/documents/10180/11553/Transforming+public+services+using+technology+and+digital+approaches/ab9af2bd-9b68-4473-ac17-bbddf2adec05 (visited on 08/12/2015).
- Lockwood, Ben and Francesco Porcelli (2013). 'Incentive schemes for local government: theory and evidence from Comprehensive Performance Assessment in England'. In: *American Economic Journal: Economic Policy* 5.3, pp. 254–286. URL: <http://doi.org/10.1257/pol.5.3.254>.
- Lodge, Martin and Derek Gill (2011). 'Toward a new era of administrative reform? The myth of post-NPM in New Zealand'. In: *Governance* 24.1, pp. 141–166. URL: <http://doi.org/10.1111/j.1468-0491.2010.01508.x>.
- Lucas, Christopher, Richard A. Nielsen, Margaret E. Roberts, Brandon M. Stewart, Alex Storer and Dustin Tingley (2015). 'Computer-assisted text analysis for comparative politics'. In: *Political Analysis* 23.2, pp. 254–277. URL: <http://doi.org/10.1093/pan/mpu019>.
- Lui, Marco and Timothy Baldwin (2012). 'langid.py: An off-the-shelf language identification tool'. In: *Proceedings of the ACL 2012 system demonstrations*. Stroudsburg, PA, USA: Association for Computational Linguistics, pp. 25–30. URL: <http://dl.acm.org/citation.cfm?id=2390470.2390475> (visited on 05/02/2016).

- Manning, Christopher D., Prabhakar Raghavan and Hinrich Schütze (2008). *Introduction to information retrieval*. Cambridge: Cambridge University Press. 482 pp. ISBN: 978-0-521-86571-5.
- Manoharan, Aroon, Marc Fudge and Yueping Zheng (2014). 'Global clusters in municipal e-governance: a longitudinal study'. In: *International Journal of Public Administration*, pp. 1–16. URL: <http://doi.org/10.1080/01900692.2014.931422>.
- Margetts, Helen (1999). *Information Technology in government: Britain and America*. London: Routledge. 226 pp. ISBN: 978-0-203-02094-4.
- Margetts, Helen and Patrick Dunleavy (2013). 'The second wave of digital-era governance: a quasi-paradigm for government on the Web'. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371.1987, p. 20120382. URL: <http://doi.org/10.1098/rsta.2012.0382>.
- Margetts, Helen and Gerry Stoker (2010). 'Experiments and quasi-experiments'. In: *Theories and methods in political science*. Ed. by David Marsh and Gerry Stoker. 3rd. London: Palgrave Macmillan. ISBN: 978-0-230-57627-8.
- Margetts, Helen and Leslie Willcocks (1993). 'Information technology in public services: disaster faster?' In: *Public Money & Management* 13.2, pp. 49–56. URL: <http://doi.org/10.1080/09540969309387763>.
- Martin, Steve (2002). 'The modernization of UK local government: markets, managers, monitors and mixed fortunes'. In: *Public Management Review* 4.3, pp. 291–307. URL: <http://doi.org/10.1080/14616670210151595>.
- Meijer, Albert (2007). 'Why don't they listen to us? Reasserting the role of ICT in public administration'. In: *Information Polity* 12.4, pp. 233–242. URL: <http://content.iospress.com/articles/information-polity/ip000127> (visited on 01/04/2016).
- Meijer, Albert and Marcel Thaens (2009). 'Public information strategies: making government information available to citizens'. In: *Information Polity* 14.1, pp. 31–45. URL: <http://doi.org/10.3233/IP-2009-0167>.

- Meijer, Albert and Stavros Zouridis (2004). 'E-government as institutional transformation'. In: IRMA Annual Conference. New Orleans, LA, pp. 1–9. URL: https://www.researchgate.net/profile/Albert_Meijer/publication/27702319_E-government_as_Institutional_Transformation/links/53f210960cf272810e4c8e3e.pdf (visited on 29/03/2016).
- Monroe, Burt L. (2013). 'The five Vs of big data political science: introduction to the virtual issue on big data in political science'. In: *Political Analysis* Virtual Issue on Big Data in Political Science, pp. 1–9. URL: <http://www.oxfordjournals.org/page/5658/6> (visited on 30/11/2014).
- Moon, M. Jae and Donald F. Norris (2005). 'Does managerial orientation matter? the adoption of reinventing government and e-government at the municipal level'. In: *Information Systems Journal* 15.1, pp. 43–60. URL: <http://doi.org/10.1111/j.1365-2575.2005.00185.x>.
- Moore, Mark (1997). *Creating public value: strategic management in government*. Harvard University Press. 412 pp. ISBN: 978-0-674-17558-7.
- Morgeson, Forrest V. and Sunil Mithas (2009). 'Does e-government measure up to e-business? Comparing end user perceptions of US federal government and e-business web sites'. In: *Public Administration Review* 69.4, pp. 740–752. URL: <http://doi.org/10.1111/j.1540-6210.2009.02021.x>.
- Nasre, Meghana, Matteo Pontecorvi and Vijaya Ramachandran (2013). 'Betweenness centrality – incremental and faster'. In: *arXiv:1311.2147 [cs]*. URL: <http://arxiv.org/abs/1311.2147> (visited on 08/02/2016).
- National Audit Office (1999). *Government on the Web*. HC 87. London: TSO. 117 pp. URL: <http://www.nao.org.uk/wp-content/uploads/1999/12/990087.pdf> (visited on 10/07/2014).
- (2002). *Government on the Web II*. HC 764. London: TSO. URL: <http://www.nao.org.uk/wp-content/uploads/2002/04/0102764.pdf> (visited on 10/07/2014).
- (2007). *Government on the Internet: progress in delivering information and services online – research report*. London: National Audit Office, p. 89. URL: <http://www>.

- nao.org.uk/wp-content/uploads/2007/07/0607529_Research.pdf (visited on 11/07/2014).
- Nicholls, Tom (2013). 'An empirical analysis of Internet top-level domain policy'. In: *Journal of Information Policy* 3, pp. 464–484. URL: <http://doi.org/10.5325/jinfopoli.3.2013.0464>.
- (2014). *Warctika*. Version 1.0. URL: <https://github.com/pmyteh/warctika>.
- (2016). *Full-population web crawl of .gov.uk: URLs, links, and third-level domain graph*. Data Collection; forthcoming.
- Nicholls, Tom and Jonathan Bright (2016). *handclassifier*. Version 0.5.0. URL: <https://github.com/oxfordinternetinstitute/handclassifier>.
- Nograšek, Janja and Mirko Vintar (2014). 'E-government and organisational transformation of government: black box revisited?' In: *Government Information Quarterly* 31.1, pp. 108–118. URL: <http://doi.org/10.1016/j.giq.2013.07.006>.
- Nolan, Richard L. (1973). 'Managing the computer resource: a stage hypothesis'. In: *Communications of the ACM* 16.7, pp. 399–405. URL: <http://doi.org/10.1145/362280.362284>.
- Norris, Donald F. and Christopher G. Reddick (2013). 'Local e-government in the United States: transformation or incremental change?' In: *Public Administration Review* 73.1, pp. 165–175. URL: <http://doi.org/10.1111/j.1540-6210.2012.02647.x>.
- O'Donnell, Gus (2013). 'Better government'. In: *The Political Quarterly* 84.3, pp. 380–387. URL: <http://doi.org/10.1111/j.1467-923X.2013.12037.x>.
- ODPM (2003). *Local e-government: process evaluation of the implementation of electronic local government in England - research summary*. London: Office of the Deputy Prime Minister. 5 pp.
- (2004). *Indices of Deprivation 2004*. URL: <http://webarchive.nationalarchives.gov.uk/20100410180038/http://www.communities.gov.uk/archived/general-content/communities/indicesofdeprivation/216309/> (visited on 23/05/2014).

- Oliveira, Gustavo Henrique Maultasch and Eric W. Welch (2013). 'Social media use in local government: linkage of technology, task, and organizational context'. In: *Government Information Quarterly* 30.4, pp. 397–405. URL: <http://doi.org/10.1016/j.giq.2013.05.019>.
- O'Neill, Rose (2009). 'The transformative impact of e-government on public governance in New Zealand'. In: *Public Management Review* 11.6, pp. 751–770. URL: <http://doi.org/10.1080/14719030903318939>.
- Ordnance Survey (2014). *Boundary-Line*. URL: http://parlvid.mysociety.org/os/bdline_gb-2014-05.zip (visited on 22/01/2016).
- Ordnance Survey Northern Ireland (2016). *Open Data Largescale Boundaries – Local Government Districts 1993*. URL: http://osni.spatial-ni.opendata.arcgis.com/datasets/6cfafa7abbbe4d798cdb74f41bc61921_7 (visited on 13/04/2016).
- Page, Lawrence, Sergey Brin, Rajeev Motwani and Terry Winograd (1999). *The PageRank citation ranking: bringing order to the web*. Technical Report 1999-66. Stanford InfoLab. URL: <http://ilpubs.stanford.edu:8090/422/>.
- Park, Han Woo and Mike Thelwall (2003). 'Hyperlink analyses of the world wide web: a review'. In: *Journal of Computer-Mediated Communication* 8.4. URL: <http://doi.org/10.1111/j.1083-6101.2003.tb00223.x>.
- Parks, Roger B., Paula C. Baker, Larry Kiser, Ronald Oakerson, Elinor Ostrom, Vincent Ostrom, Stephen L. Percy, Martha B. Vandivort, Gordon P. Whitaker and Rick Wilson (1981). 'Consumers as coproducers of public services: some economic and institutional considerations'. In: *Policy Studies Journal* 9.7, pp. 1001–1011. URL: <http://doi.org/10.1111/j.1541-0072.1981.tb01208.x>.
- Parsons, Wayne (2002). 'From muddling through to muddling up - evidence based policy making and the modernisation of British government'. In: *Public Policy and Administration* 17.3, pp. 43–60. URL: <http://doi.org/10.1177/095207670201700304>.

- Parvez, Zahid (2006). 'Informatization of local democracy: a structuration perspective'. In: *Information Polity* 11.1, pp. 67–83. URL: <http://content.iospress.com/articles/information-polity/ip000089> (visited on 01/04/2016).
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay (2011). 'Scikit-learn: machine learning in Python'. In: *Journal of Machine Learning Research* 12, pp. 2825–2830.
- Pestoff, Victor, Stephen P. Osborne and Taco Brandsen (2006). 'Patterns of co-production in public services'. In: *Public Management Review* 8.4, pp. 591–595. URL: <http://doi.org/10.1080/14719030601022999>.
- Petricek, Vaclav, Tobias Escher, Ingemar J. Cox and Helen Margetts (2006). 'The web structure of e-government: developing a methodology for quantitative evaluation'. In: *Proceedings of the 15th International Conference on World Wide Web*. www '06. New York, NY, USA: ACM, pp. 669–678. ISBN: 978-1-59593-323-2. URL: <http://doi.org/10.1145/1135777.1135875>.
- Quigley, Kevin (2004). 'The emperor's new computers: y2k (re)visited'. In: *Public Administration* 82.4, pp. 801–829. URL: <http://onlinelibrary.wiley.com/doi/10.1111/j.0033-3298.2004.00420.x/abstract> (visited on 24/10/2011).
- Quinn, Kevin M, Burt L Monroe, Michael Colaresi, Michael H Crespín and Dragomir R Radev (2010). 'How to analyze political attention with minimal assumptions and costs'. In: *American Journal of Political Science* 54.1, pp. 209–228. ISSN: 1540-5907.
- Ragin, Charles C. (2000). *Fuzzy-set social science*. Chicago: University of Chicago Press. ISBN: 978-0-226-70277-3.
- Rashman, Lyndsay and Zoe Radnor (2005). 'Learning to improve: approaches to improving local government services'. In: *Public Money & Management* 25.1, pp. 19–26. URL: <http://doi.org/10.1111/j.1467-9302.2005.00446.x>.
- Reinwald, Anja and Pernille Kraemmergaard (2012). 'Managing stakeholders in transformational government – a case study in a Danish local gov-

- ernment'. In: *Government Information Quarterly* 29.2, pp. 133–141. URL: <http://www.sciencedirect.com/science/article/pii/S0740624X1200007X> (visited on 11/10/2013).
- Revelli, Federico (2010). 'Spend more, get more? An inquiry into English local government performance'. In: *Oxford Economic Papers* 62.1, pp. 185–207. URL: <http://oep.oxfordjournals.org/content/62/1/185> (visited on 17/05/2014).
- Richardson, Leonard (2015). *Beautiful Soup*. Version 4.4.1. URL: <https://pypi.python.org/pypi/beautifulsoup4> (visited on 22/03/2016).
- robotstxt.org (2007). *About / robots.txt*. URL: <http://www.robotstxt.org/robotstxt.html>.
- Rosvall, M., D. Axelsson and C. T. Bergstrom (2009). 'The map equation'. In: *arXiv:0906.1405 [physics]*. arXiv: 0906.1405. URL: <http://arxiv.org/abs/0906.1405> (visited on 07/12/2014).
- Rufin, R., C. Medina and J.C. Sánchez Figueroa (2012). 'Moderating factors in adopting local e-government in Spain'. In: *Local Government Studies* 38.3, pp. 367–385.
- Schroeder, Ralph (2014). 'Big Data and the brave new world of social media research'. In: *Big Data & Society* 1.2, p. 2053951714563194. URL: <http://bds.sagepub.com/content/1/2/2053951714563194> (visited on 20/04/2016).
- Senyucel, Zorlu (2007). 'Assessing the impact of e-government on providers and users of the is function: a structuration perspective'. In: *Transforming Government: People, Process and Policy* 1.2, pp. 131–144. ISSN: 1750-6166. DOI: 10.1108/17506160710751968. (Visited on 31/05/2014).
- Shackleton, Peter, Julie Fisher and Linda Dawson (2004). 'Internal and external factors impacting on e-government maturity: a local government case study'. In: *Journal of Information Technology Cases and Applications* 6.4, pp. 36–50.
- Siau, Keng and Yuan Long (2005). 'Synthesizing e-government stage models: a meta-synthesis based on meta-ethnography approach'. In: *Industrial Management + Data Systems* 105.3, pp. 443–458. URL: <http://doi.org/10.1108/02635570510592352>.

- Socitm (1999). *Well Connected? A snapshot of local authority websites*. Better Connected 1999. Northampton, UK: Socitm. 85 pp. ISBN: 1-899330-41-0.
- (2000). *Better Connected? A year 2000 snapshot of local authority websites*. Better Connected 2000. Northampton, UK: Socitm. 92 pp. ISBN: 1-899330-60-7.
- (2001). *Better Connected 2001? A snapshot of local authority websites*. Better Connected 2001. Northampton, UK: Socitm. 104 pp. ISBN: 1-899330-72-0.
- (2002). *Better Connected 2002? A snapshot of all local authority websites*. Better Connected 2002. Northampton, UK: Socitm. 136 pp. ISBN: 1-899330-84-4.
- (2003). *Better Connected 2003: A snapshot of all local authority websites*. Better Connected 2003. Northampton, UK: Socitm. 168 pp. ISBN: 1-904414-02-8.
- (2004). *Better Connected 2004: A snapshot of all local authority websites*. Better Connected 2004. Northampton, UK: Socitm. 212 pp. ISBN: 1-904414-17-6.
- (2005). *Better Connected 2005: A snapshot of all local authority websites*. Better Connected 2005. Northampton, UK: Socitm. 232 pp. ISBN: 1-904414-33-8.
- (2006). *Better Connected 2006: A snapshot of all local authority websites*. Better Connected 2006. Northampton, UK: Socitm. 226 pp. ISBN: 1-904414-48-6.
- (2007). *Better Connected 2007: A snapshot of all local authority websites*. Better Connected 2007. Northampton, UK: Socitm. 244 pp. ISBN: 1-904414-63-x.
- (2008). *Better Connected 2008: A snapshot of all local authority websites*. Better Connected 2008. Northampton, UK: Socitm. 244 pp. ISBN: 1-904414-78-8.
- (2009). *Better Connected 2009: A snapshot of all local authority websites*. Better Connected 2009. Northampton, UK: Socitm. 244 pp. ISBN: 1-904414-91-5.
- (2010). *Better Connected 2010: A snapshot of all local authority websites*. Better Connected 2010. Northampton, UK: Socitm. 212 pp. ISBN: 978-1-907608-00-1.
- (2011). *Better Connected 2011: A snapshot of all local authority websites*. Better Connected 2011. Northampton, UK: Socitm. 297 pp. ISBN: 978-1-907608-12-4.
- (2012). *Better Connected 2012: A snapshot of all local authority websites*. Better Connected 2012. Northampton, UK: Socitm. 265 pp. ISBN: 978-1-907608-19-3.

- Socitm (2013a). *Better Connected 2013: A snapshot of all local authority websites*. Better Connected 2013. Northampton, UK: Socitm. 248 pp. ISBN: 978-1-907608-28-5.
- (2013b). *Website Performance Survey 2013: summary data*. Unpublished dataset.
- (2014). *No to single local government website, yes to common platform for local government tools and applications, says Socitm*. URL: <http://www.socitm.net/press/no-single-local-government-website-yes-common-platform-local-government-tools-and-applications> (visited on 19/06/2014).
- Socitm and Improvement and Development Agency (2002). *Local e-government now 2002: there for the taking*. Northampton: Socitm. 192 pp. ISBN: 1-899330-89-5. URL: <https://www.rbkc.gov.uk/COMMITTEES/Document.ashx?czJKcaeA i5tUFL1DTL2UE4zNRBcoShgo=9wso55SP2cKun0mgYOPmBzfzRSM 9y1GLy3HcmiLUvrSvR7ycku3fg%3D%3D> (visited on 27/06/2014).
- Socitm, Improvement and Development Agency and Local Government Association (2003). *Casting the net wider: local e-democracy 2003*. Northampton: Socitm. 20 pp. ISBN: 1-84049-348-8.
- Sørum, Hanne, Rony Medaglia and Kim N. Andersen (2009). 'Assessment of website quality: Scandinavian web awards right on track?' In: *Electronic government*. Ed. by Maria A. Wimmer, Hans J. Scholl, Marijn Janssen and Roland Traunmüller. Lecture Notes in Computer Science 5693. Springer Berlin Heidelberg, pp. 198–209. ISBN: 978-3-642-03515-9. URL: http://link.springer.com/chapter/10.1007/978-3-642-03516-6_17 (visited on 14/05/2015).
- Stewart, John (2003). *Modernising British local government: an assessment of Labour's reform programme*. Basingstoke, Hampshire: Palgrave Macmillan. 288 pp. ISBN: 978-0-333-96465-1.
- Strassmann, Paul A. (1990). *The business value of computers: an executive's guide*. New Canaan, CT: Information Economics Press. 530 pp. ISBN: 0-9620413-2-7.

- (1997). *The squandered computer: evaluating the business alignment of information technologies*. New Canaan, CT: Information Economics Press. 426 pp. ISBN: 0-9620413-1-9.
- Tat-Kei Ho, Alfred (2002). 'Reinventing local governments and the e-government initiative'. In: *Public Administration Review* 62.4, pp. 434–444. URL: <http://doi.org/10.1111/0033-3352.00197>.
- The Legal Deposit Libraries (Non-Print Works) Regulations 2013*. URL: <http://www.legislation.gov.uk/ukdsi/2013/9780111533703> (visited on 06/07/2014).
- Thelwall, Mike (2003). 'A free database of university web links: data collection issues'. In: *Cybermetrics* 6/7.2, p. 11. URL: <http://www.cybermetrics.info/articles/v6i1p2.pdf> (visited on 11/07/2014).
- (2009). *Introduction to webometrics: quantitative web research for the social sciences*. Synthesis Lectures on Information Concepts, Retrieval, and Services. San Rafael, CA: Morgan and Claypool Publishers. 126 pp. ISBN: 978-1-59829-993-9.
- (2013). 'Webometrics and social web research methods: an updated and extended version of 'Introduction to webometrics', including chapters to appear in a forthcoming social web methods book'. URL: <http://www.scit.wlv.ac.uk/~cm1993/papers/IntroductionToWebometricsAndSocialWebAnalysis.pdf> (visited on 10/05/2014).
- Thelwall, Mike and David Stuart (2006). 'Web crawling ethics revisited: cost, privacy, and denial of service'. In: *Journal of the American Society for Information Science and Technology* 57.13, pp. 1771–1779. URL: <http://doi.org/10.1002/asi.20388>.
- Trotta, Marta, Danila Scaroza, Alessandro Hinna and Luca Gnan (2011). 'Can information systems facilitate the integration of New Public Management and Public Governance? Evidence from an Italian public organization'. In: *Information Polity* 16.1, pp. 23–34. URL: <http://doi.org/10.3233/IP-2011-0223>.

- Van Ryzin, Gregg G. (2013). 'Service quality, administrative process, and citizens' evaluation of local government in the US'. In: *Public Management Review* 17.3, pp. 425–442. URL: <http://doi.org/10.1080/14719037.2013.841456>.
- Walker, Richard M. and George A. Boyne (2006). 'Public management reform and organizational performance: an empirical assessment of the UK Labour government's public service improvement strategy'. In: *Journal of Policy Analysis and Management* 25.2, pp. 371–393. URL: <http://doi.org/10.1002/pam.20177>.
- Weinberger, Kilian, Anirban Dasgupta, John Langford, Alex Smola and Josh Attenberg (2009). 'Feature hashing for large scale multitask learning'. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ICML '09. New York, NY, USA: ACM, pp. 1113–1120. ISBN: 978-1-60558-516-1. URL: <http://doi.org/10.1145/1553374.1553516>.
- West, Darrell M. (2004). 'E-government and the transformation of service delivery and citizen attitudes'. In: *Public Administration Review* 64.1, pp. 15–27. URL: <http://doi.org/10.1111/j.1540-6210.2004.00343.x>.
- Whalen, Ryan (2011). 'The structure of federal eGovernment: using hyperlinks to analyze the .gov domain'. In: Political Networks Conference. Ann Arbor, MI, p. 22. URL: http://opensiuc.lib.siu.edu/pnconfs_2011/8/ (visited on 05/08/2014).
- Willcocks, Leslie (1992). 'Evaluating information technology investments: research findings and reappraisal'. In: *Information Systems Journal* 2.4, pp. 243–268. URL: <http://doi.org/10.1111/j.1365-2575.1992.tb00081.x>.
- (1994). 'Managing information systems in UK public administration: issues and prospects'. In: *Public Administration* 72, pp. 13–32.
- Wilson, David and Chris Game (2011). *Local government in the United Kingdom*. 5th. Basingstoke: Palgrave Macmillan. 463 pp. ISBN: 978-0-230-24639-3.
- Wooldridge, Jeffrey M. (2010). *Econometric analysis of cross section and panel data*. 2nd ed. Cambridge, Mass: MIT Press. 1064 pp. ISBN: 978-0-262-23258-6.

- Wynen, Jan and Koen Verhoest (2013). 'Do NPM-type reforms lead to a cultural revolution within public sector organizations?' In: *Public Management Review* 17.3, pp. 356–379. URL: <http://doi.org/10.1080/14719037.2013.841459>.
- Yannas, P. and G. Lappas (2007). 'Evaluating local e-government: an analysis of Greek prefecture websites'. In: *2nd International Conference on Digital Information Management, ICDIM*. Vol. 1. ISBN: 978-1-4244-1475-8. URL: <http://doi.org/10.1109/ICDIM.2007.4444232>.
- Yıldız, Mete and Ayegül Saylam (2013). 'E-government discourses: an inductive analysis'. In: *Government Information Quarterly* 30.2, pp. 141–153. URL: <http://doi.org/10.1016/j.giq.2012.10.007>.