Acknowledgements

I am grateful to the Project Director, Professor Elizabeth Shepherd, for her oversight and guidance, to Helen Sumping for her assistance with the annotated bibliography, and Megan Venter for her assistance with research into civic technologies.
Table of Contents

Acknowledgements 2
1. Introduction 4
2. Literature Search 7
3. Annotated Bibliography 30
4. Literature Review 72
References 79
1. Introduction

The Open Government Data Literature Review project produced an analysis of key literature about open government data and the recordkeeper’s role in order to establish a framework for further research.

The area of open government data is fairly new for recordkeepers and it is important that we understand the issues better than is the case at present. Although related to freedom of information/access to information and to public information and records management, open data has some distinct characteristics which merit investigation. For example, the relationships between open government data, public records, freedom of information, linked data and the role of recordkeepers and of data scientists is not fully understood; issues of provenance and co-creation as government functions are increasingly delivered by commercial and third sector bodies in partnership with the public sector; and the consequences for the recordkeepers need to be examined. A first step towards such new research is a systematic literature review.

The literature review produced by the Open Government Data Literature Review project consists of three separate elements:

Literature search

The literature search comprises a list of the published works related to the scope of this project. The search strategy included searches of the following search engines and databases:

- Google and Google Scholar
- University of London Library catalogue
- Web of Science
- SCOPUS

The following search terms were used in each:

- ‘open government data’
- ‘open data’
• open government data + records management
• open data + records management
• open government data + archives
• open data + archives
• open government data + information management
• open data + information management

A specific review of the Journal of Community Informatics was also conducted, given its profile for publishing significant works on open government data.

Additionally, the websites of a number of organisations were reviewed for relevant publications. These were:

• The Open Government Partnership
• The Open Data Institute
• The Open Knowledge Foundation
• The Sunlight Foundation

Finally, a Google search for civic technologies revealed a number of relevant sites, of which the following are described in the Literature Review as examples:

• GotToVote: http://gottovote.cc/
• Fix my Street: https://www.fixmystreet.com
• They Work For You: http://www.theyworkforyou.com
• Transparent Chennai: http://www.transparentchennai.com/
• NHS Citizen: https://www.nhscitizen.org.uk
• Quién te financia: http://quientefinancia.cl/

Annotated bibliography

The annotated bibliography presents significant contributions to the formally published scholarly literature, together with their published abstracts.
Literature review proper

The literature review proper is an analysis of the key works from the literature search. This review covers books, journals, official reports, guides and civic technology sites, but cannot track all relevant media, such as social media.

Given that open government data is a rapidly expanding area of practice and scholarship, it is recommended that the literature search and bibliography should be used as the basis for a wiki style resource that can continue to develop over time.
2. Literature Search


InterPARES Trust  EU02: Open Government Data Literature Review


InterPARES Trust  EU02: Open Government Data Literature Review


InterPARES Trust EU02: Open Government Data Literature Review


InterPARES Trust EU02: Open Government Data Literature Review


https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0CC0QFjABahUKEwifvOWt7fv1AhWCvxoKHfgpCLw&url=https%3A%2F%2Fcirandas.net%2Falandtygel%2Fgaleria%2Fbudgetdata.pdf&usg=AFQjCNJGljceYiNMrXmDCeSQx-uhceuq5g


3. Annotated Bibliography

This annotated bibliography includes the items in the literature search results, with their published abstracts.


Open Government Data (OGD) is a technology platform adopted by governments to meet new society demands. This theme is based on the assumptions of transparency, collaboration and participation. It is believed that from these initiatives, it is possible to generate activities of economic and social value through the use of public data by society. However, some challenges emerge for the effective achievement of these objectives. They include a series of factors, among which the article highlights: technical problems related to the handling of information within government structures, political principles of each government, priorities of government actions, legal issues and aspects related mainly to the interest and use of this information by society. Thus, the main purpose of this study is to identify and describe the activities developed along the value chain of open government data.


The use of Open Government Data (OGD) involves multiple activities developed by networks of users with different capacities and goals, along a value chain. These users, and also government agents supplying OGD recognize benefits, motivations, barriers,
facilitating and inhibiting factors in the process. This paper surveyed the perceptions of Brazilian OGD users and government agents, in order to provide information for the improvement of OGD supply and use.


The trend of opening government data, in order to be used for scientific, commercial and political purposes, has resulted in the development of numerous e-infrastructures providing public sector information (PSI). The big investments that have been made in this direction necessitate a deeper understanding and assessment of the value they produce. This paper presents a methodology for evaluating PSI e-infrastructures, which is based on the estimation of multiple value models corresponding to their main stakeholder groups: data users and data providers. Each value model consists of several value dimensions and their corresponding value measures, organized in three levels (associated with efficiency, effectiveness and future behaviour), and also the relations among them. The proposed methodology allows a comprehensive assessment of the various types of value generated by a PSI e-infrastructure for each stakeholder group, and also the interconnections among them. This enables a better understanding of the whole value generation mechanism and a rational definition of improvements priorities.


Two important trends in government that are emerging in the recent years have been on one hand the exploitation of the Web 2.0 social media, supporting a more extensive interaction and collaboration with citizens, and on the other hand the opening of government data to the citizens through the Internet, in order to be used for scientific, commercial and political purposes. However, there has been limited attempt of
integrating them. Using a design science approach a second generation of open government data (OGD) platforms has been developed, which offer to the users both the 'classical' first generation functionalities, and also a comprehensive set of additional novel Web 2.0 features. The latter aim to provide support to the users in order to generate value from ODG. They enable users to become 'prosumers', both producing and consuming data. These novel capabilities for performing various types of processing, information and knowledge exchange, and collaboration were found to be useful and valuable by users in a first evaluation.


In the past couple of years we have witnessed an increased interest among governments around the world to release raw data to their citizens. This interest was augmented by converting the raw data into semantic representation to insure that the data can be shared and reused across applications. In this paper the authors present the attempt to convert Saudi Open Government Data (OGD) into Resource Description Framework (RDF) format by reusing existing ontologies. The advantage of performing such a process is twofold: (1) contribute to the vision of the semantic web and linked open data initiative and (2) enable data exchange and linking with other semantic resources.


The emergence of a global cyberinfrastructure is rapidly increasing the ability of scientists to produce, manage, and use data, leading to new understanding and modes of scientific inquiry that depend on broader data access. As research becomes increasingly global, data intensive, and multifaceted, it is imperative to address national and international data access and sharing issues systematically in a policy arena that transcends national jurisdictions. The authors of this Policy Forum summarize key findings of an international group that studied these issues on behalf of the OECD, and
argue that an international framework of principles and guidelines for data access is needed to better realize this potential. They provide a framework for locating and analysing where improvements can be made in data access regimes, and highlight several topics that require further examination to better inform future policies.


The article argues, drawing on interviews, event observations, academic and policy literature, that the UK’s Open Government Data (OGD) initiative should be understood as part of a deep seated social and political struggle with significant processes of co-optation and contestation over outcomes underway. The OGD initiative’s intersections with both the PSI re-use industry and the UK government’s programme of forced ‘austerity’ and marketisation of public services are problematised. Civil society advocates’ vulnerabilities within this context are discussed and a number of recommendations are offered for the progressive shaping of OGD based on egalitarian principles.


The article adopts a neo-Gramscian analytical framework developed in the field of International Political Economy in order to analyze the relationship between an online collective of civil society actors and U.K. government policymakers in the case of the UK’s Open Government Data (OGD) initiative. The aim of the article is to consider the neo-Gramscian notion of trasformismo as a useful conceptual tool for exploring the relations between the OGD advocates and policymakers within the U.K. state. Empirical evidence is presented which suggests that the notion of trasformismo is able to illuminate some of the political processes of absorption, adaptation, and distortion which have emerged during the development of the UK’s OGD initiative, and which
have functioned to restrict the counterhegemonic potential of OGD in order to shape the initiative toward a distinctly neoliberal framework for action.


The article analyses the role of Open Government Data policy as part of the broader public policy agenda of the UK government. A thematic analysis of interview, observational and policy documentation is presented which suggests that since 2010 the Open Government Data agenda has been used strategically by the UK's centre-right coalition government to progress a range of controversial policies, which are aimed at the continuation of the neoliberal form of state through its current crisis. Specifically, the relationship between Open Government Data policy and the neoliberal objectives of the marketisation of public services and privatisation of public assets, the leveraging of financial markets and the pharmaceutical industry, and the embedding of OGD into a broader agenda aimed at rebuilding trust in political elites are analysed. These findings are examined in relation to Braman's (2006, 2011) arguments regarding the strategic implementation of information policy by Governments in the exercising of state power, and the development of the 'informational state'.


The literature provides a wide range of techniques to assess and improve the quality of data. Due to the diversity and complexity of these techniques, research has recently focused on defining methodologies that help the selection, customization, and application of data quality assessment and improvement techniques. The goal of this article is to provide a systematic and comparative description of such methodologies. Methodologies are compared along several dimensions, including the methodological phases and steps, the strategies and techniques, the data quality dimensions, the types of
data, and, finally, the types of information systems addressed by each methodology.
The article concludes with a summary description of each methodology.


Providing a broad investigation into the significance of the "information age" to contemporary government, this study examines available perspectives on the relationship between information and communication technology and social change, applying this to government practice in the UK.


Given that the Obama Administration still relies on many strategies we would think of as on the side of secrecy, it seems that the only lasting transparency legacy of the Obama Administration will be data-driven or e-transparency as exemplified by the web interface ‘data.gov’. As the data-driven transparency model is exported and assumes an ascendant position around the globe, it is imperative that we ask what kind of publics, subjects, and indeed, politics it produces. Open government data is not just a matter concerning accountability but seen as a necessary component of the new ‘data economy’. To participate and benefit from this info-capitalist-democracy, the data subject is called upon to be both auditor and entrepreneur. This article explores the implications of responsibilisation, outsourcing, and commodification on the contract of representational democracy and asks if there are other forms of transparency that might better resist neoliberal formations and re-politicise the public sphere.

Many government organizations publish a variety of data on the web to facilitate transparency. The multitude of sources has resulted in heterogeneous structures and formats as well as varying quality of such data. The authors report on a project dubbed GovWild (Government Web Data Integration for Linked Data) that integrates and cleanses open government data at a large scale. Also, the authors point to the unified and clean integration result, published as Linked Open Data at govwild.hpi-web.de, and feature their web application to showcase the usability of the created dataset.


The aim of the article is to investigate what characterizes the information constructs that the archival discourse and the open data discourse communicate in text, and what their similarities and differences are. This article proposes that it is possible to see the open data initiative and modern archival practice as two discourses that have used different terminology to express and communicate their messages in the literature. In this article, the authors have applied a hypothesis-like assumption that the information constructs used in open data are actually nothing other than records, as they are in the archival discourse.


To design the first steps, the City of Berlin performed an Open Data Online Voting. More than 1,300 participants voted for three interesting categories out of a list of 20. A list from “Administration” via “Education” and “Health” to “Waste” was given. Five categories covered nearly 50 percent of the responders’ interest. The statistical results of the Survey were compared to a SAS survey about the general interest in Open Data, Transparency and Access from 2010. The Citizens were also asked, if they would read published Open Data and work with Open Data, to create Apps and Visualisation of combined Data Sets. This could be used as a blueprint for other cities.

This article discusses the current status of open government data in Brazil and summarizes the lessons learned from publishing Brazilian government data as linked data.


The use of Government Open Data to deliver services and promote transparency is a current goal of many governments. The paper briefly describes two contest winner applications with focus on these goals, and presents a discussion about the difficulties of integrating Government Open Data to develop them.


The enhanced development of information and communication technologies in government has created new opportunities for agencies to collect, share and re-use data. At the same time, the commercial worth of governmental data sets and value-added information products/services have increased. Government agencies are finding that data they have routinely collected to fulfil their statutory and business functions can now more easily be re-used for commercial purposes. The prospect of increasing revenue through the commercial re-use of public sector information (PSI) is clearly appealing for governments and their agencies. Examples of PSI that have been re-used commercially include residential property transaction details, land title information, ordinance survey data and street address registers. This article examines those shifting boundaries, particularly in light of information privacy and national security concerns.
arising from governmental commercialization of PSI that includes personal information. The author seeks to show that information privacy laws can be subverted in the face of overt commercialization by government agencies; that commercialization can have a negative impact on individuals; and can also give rise to societal concerns relating to privacy, security, and the open governance of the information society.


Open Government Data often contain information that, in more or less detail, regard private citizens. For this reason, before publishing them, public authorities manipulate data to remove any sensitive information while trying to preserve their reliability. This paper addresses the lack of tools aimed at measuring the reliability of these data. The authors present two procedures for the assessment of the Open Government Data reliability, one based on a comparison between open and closed data, and the other based on analysis of open data only. The authors evaluate the procedures over data from the data.police.uk website and from the Hampshire Police Constabulary in the United Kingdom. The procedures effectively allow estimating the reliability of open data and, actually, their reliability is high even though they are aggregated and smoothed.


Public authorities are increasingly sharing sets of open data. These data are often preprocessed (e.g. smoothened, aggregated) to avoid to expose sensible data, while trying to preserve their reliability. The authors present two procedures for tackling the
lack of methods for measuring the open data reliability. The first procedure is based on a comparison between open and closed data, and the second derives reliability estimates from the analysis of open data only. The authors evaluate these two procedures over data from the data.police.uk website and from the Hampshire Police Constabulary in the UK. With the first procedure the authors show that the open data reliability is high despite preprocessing, while with the second one the authors show how it is possible to achieve interesting results concerning the open data reliability estimation when analysing open data alone.


Web 2.0 has changed the way that information is presented to people. Public administrations have an important role to play in this new era. Global institutions, central, regional and local governments gather and produce a wide variety of information that is potentially reusable by citizens and the digital content industry. Gov 2.0 follows the approach of giving access to open public data to citizens. Open Government Data (OGD) establishes the principles for providing public data to the public. The development of innovative applications by companies or individuals from these public records, will meet the demand of information from citizens, as well as developing the basic principles of transparency, publication and reutilization. The authors present a local administration experience showing the steps to implement the OGD strategy. Benefits and responsibilities of those involved in the full process are presented. The authors have developed a real case of publicly available data from the population census of the municipality.

Recently, a second generation of advanced open government data (OGD) infrastructures has emerged, influenced by the principles of the Web 2.0 paradigm, and oriented towards the elimination of the clear distinction between providers and consumers of such data, and the support of data 'pro-sumers'. This paper presents and validates a methodology for evaluating these advanced second generation of ODG infrastructures, which is based on the estimation of value models of them from users' ratings. This value model includes assessments of the various types of value generated by such an infrastructure, and also of the relations among them as well. This enables a deeper understanding of the whole value generation mechanism and a rational definition of improvement priorities. The proposed methodology has been used for the evaluation of an advanced second generation ODG e-Infrastructure developed in the European project ENGAGE.


Considerable investments are made to develop numerous e-infrastructures for the reuse of open government data for scientific, commercial and political purposes. This necessitates a deeper understanding and assessment of the value these infrastructures generate. For this purpose, the paper presents a framework for evaluating open government data infrastructures, both ‘traditional’ ones following the web 1.0 paradigm and also advanced ones influenced by the web 2.0 paradigm. The evaluation framework is based on findings of previous research on the evaluation of public projects, information systems and e-services, and also on technology acceptance and IS success models. The proposed evaluation framework consists of an evaluation model with measurable evaluation dimensions and criteria, as well as a comprehensive evaluation procedure for using this evaluation model, which enables both higher level and detailed evaluation. It includes quantitative as well as qualitative methods in order to provide comprehensive and deep insights. Finally, the authors describe an application of the proposed framework (both the model and the procedure) for the evaluation of a European e-infrastructure for opening government data. this first application has provided some first evidence concerning the applicability and usefulness of the
proposed evaluation framework, and at the same time useful directions and ideas for the improvement of the above-mentioned e-infrastructure.


Quién te financia is a campaign that advocates for transparency within the Chilean political process. The campaign’s website allows users to request details of financial donations from politicians, and allows them to add their name toward a petition urging more transparency within Chilean campaign finance. The site also allows users to utilize social media to demonstrate their support for the campaign, such as the ability to use Twitter to assert support for specific financial transparency platforms. The campaign uses data from the Smart Citizen Foundation, as well as data supplied by users, academics, and politicians.


Transparency initiatives are increasingly present at all levels of government, making volume and variety of data grow as governmental institutions wish to make their administration more open. In this context, transparency portals appear as large silos of documents with heterogeneous and often unstructured data, in which the information provided is mainly through PDF, HTML, Excel spreadsheets or other print-like format. Although they allow comfortable reading to humans, these documents fail to provide a deep, integrated and multidimensional data analysis, which would allow citizens the effective use of transparency. The authors hence present a work in progress to structure this information and to allocate them into repositories that comply with Open Government Data principles to provide citizens with the power of machine analysis, indiscriminate and independent technology data access. The authors describe a collaborative framework that engages users in capturing and transforming unstructured information into machine-readable datasets.

Since the launch of the Open Government Partnership, several countries have acceded to this multilateral agreement to develop and to implement ambitious reforms to make their governments more open. Brazil, as one of the eight founding countries, has implemented a series of actions to open its government. One of these key actions is its Access to Information Law. The Brazilian law established a legal framework of guidelines for opening data from all levels of government in the country, in addition to considering internet and transparency portals key elements for the consolidation of open government. In this scenario, local governments built websites dedicated to transparency without observing the law requirements and, consequently, Open Government Data (OGD) principles. This paper shows a comprehensive assessment of transparency websites through a survey with 20 municipalities. The Brazilian law requirements are used as evaluation criteria because they somehow cover OGD principles. The results show a gap between local transparency portals and the effective implementation of the OGD principles. This gap leads to a misconception that transparency portals are dissociated from the open government initiative, which is not true.


The Open Data Barometer aims to uncover the true prevalence and impact of open data initiatives around the world. It analyses global trends, and provides comparative data on countries and regions via an in-depth methodology combining contextual data, technical assessments and secondary indicators to explore multiple dimensions of open data readiness, implementation and impact.
Under the UK government’s open government data (OGD) initiative, departments have been encouraged to make nonpersonal government held datasets available online under open licenses and in standard formats, enabling the re-use of data to support transparency and accountability, improved public services and innovation and economic growth. This policy covers a wide range of government datasets, from core reference data and regularly collected performance indicators, to one-off research commissioned to support policy making. In arguments for open data, government datasets are commonly treated as if they are pre-existing artefacts waiting to be transferred from their current locations locked away on government hard drives, to public availability on websites and data-portals. However, in practice, many datasets are constructed in the process of being opened: whether as combinations of source material, or as derivative extracts of internal data systems. In this short paper the authors present a brief case study of one instance of open data release, focusing on a dataset related to the ‘Digital Landscape Research’ published in 2012 alongside a new Government Digital Strategy. The authors explore different factors influencing how the data came to be published, and question whether a simple call for ‘raw data’ best serves the cause of promoting data re-use.


This editorial essay presents an overview of the contributions in the Special Issue. It also raises critical questions for further research and thinking about Open Government Data.

http://dx.doi.org/10.2139/ssrn.2403069
With the advent of open data initiatives, raw data has been staged as a crucial element of government transparency. If the consequences of such data-driven transparency have already been discussed, we still don’t know much about its back rooms. What does it mean for an administration to open its data? Following information infrastructure studies, this communication aims to question the modes of existence of raw data in administrations. Drawing on an ethnography of open government data projects in several French administrations, it shows that data are not ready-at-hand resources. Indeed, three kinds of operations are conducted that progressively instantiate open data. The first one is exploration. Where are, and what are, the data within the institution are tough questions, the response to which entails organizational and technical inquiries. The second one is extraction. Data are encapsulated in databases and its release implies a sometimes complex disarticulation process. The third kind of operations is ‘rawification’. It consists in a series of tasks that transforms what used to be indexical professional data into raw data. To become opened, data are (re)formatted, cleaned, ungrounded. Though largely invisible, these operations foreground specific ‘frictions’ that emerge during the sociotechnical shaping of transparency, even before data publication and reuses.


Governments around the world have been releasing raw data to their citizens at an increased pace. The mixing and linking of these datasets by a community of users enhances their value and makes new insights possible. The use of *mashups* — digital works in which data from one or more sources is combined and presented in innovative ways — is a great way to expose this value. Mashups enable end users to explore data that has a real tangible meaning in their lives. Although there are many approaches to publishing and using data to create mashups, the authors believe Linked Data and Semantic Web technologies solve many of the true challenges in open government data and can lower the cost and complexity of developing these applications. In this chapter the authors discuss why Linked Data is a better model and how it can be used to build useful mashups.

The International Open Government Dataset Search (IOGDS) team discusses what they've learned about international government data publication trends and tendencies through the application of data analytics and visualization to metadata.


Government organizations produce and disseminate a large quantity of information every day. Open government data movement have made these data available for reuse and accessibility. However, to merely publish data retrieved from legacy systems is not enough for reuse and integration. Despite the approaches proposed to publish government data, using technologies like XML, RDF e OWL, these are not suitable for representing real intended meaning of database conceptual schemas belonging to legacy systems. Some approaches propose using top-level ontologies. This paper aims to raise the difficulties to establish the ontological commitment of legacy schemas towards top-level ontologies. It illustrates the difficulties by describing a small case study on the Legal domain.


GotToVote initially started as an attempt to make an otherwise hard-to-find document released by the Kenyan government regarding voter registration centres available widely by circulating the document online. However, the project soon grew to use mapping data in conjunction with the document, allowing users to find their nearest voting centres. GotToVote now operates within Kenya, Ghana, Malawi, Zimbabwe and
provides election results within these countries and allows users to check that they are registered to vote and that their details are correct. The data used by the GotToVote sites is available on the respective government websites, but is used by GotToVote in a way that makes it clear and accessible for users.


The article discusses problems encountered in accessing information about public toilets in Chennai, despite the Indian Right to Information (RTI) legislation. Attempts to access the data revealed corruption at a local level in the allocation of public funds.


In this paper, the authors conceptualize Open Data ecosystems by analysing the major stakeholders in the UK. The conceptualization is based on a review of popular Open Data definitions and business ecosystem theories, which the authors applied to empirical data using a timeline analysis. The work is informed by a combination of discourse analysis and in-depth interviews, undertaken during the summer of 2013. Drawing on the UK as a best practice example, the authors identify a set of structural business ecosystem properties: circular flow of resources, sustainability, demand that encourages supply, and dependence developing between suppliers, intermediaries, and users. However, significant gaps and shortcomings are found to remain. Most prominently, demand is not yet fully encouraging supply and actors have yet to experience fully mutual interdependence.
In this paper, we conceptualize Open Data ecosystems by analysing the major stakeholders in the UK. The conceptualization is based on a review of popular Open Data definitions and business ecosystem theories, which we applied to empirical data using a timeline analysis. Our work is informed by a combination of discourse analysis and in-depth interviews, undertaken during the summer of 2013. Drawing on the UK as a best practice example, we identify a set of structural business ecosystem properties: circular flow of resources, sustainability, demand that encourages supply, and dependence developing between suppliers, intermediaries, and users. However, significant gaps and shortcomings are found to remain. Most prominently, demand is not yet fully encouraging supply and actors have yet to experience fully mutual interdependence.

The aim of this paper is to describe a local government effort to realise an open government agenda. This is done using a storytelling approach. The empirical data are based on a case study. The authors participated in, as well as followed, the process of realising an open government agenda on a local level, where citizens were invited to use open public data as the basis for developing apps and external Web solutions. Based on an interpretative tradition, they chose storytelling as a way to scrutinise the competition process. In this paper, they present a story about the competition process using the story elements put forward by Kendall and Kendall (2012). The research builds on existing research by proposing the myth that the “public” wants to make use
The authors provide empirical insights into the challenge of gaining benefits from open public data. In particular, they illustrate the difficulties in getting citizens interested in using open public data. Their case shows that people seem to like the idea of open public data, but do not necessarily participate actively in the data reuse process. The results are based on one empirical study. Further research is, therefore, needed. The authors would especially welcome more studies that focus on citizens’ interest and willingness to reuse open public data. This study illustrates the difficulties of promoting the reuse of open public data. Public organisations that want to pursue an open government agenda can use these findings as empirical insights. This paper answers the call for more empirical studies on public open data. Furthermore, it problematises the “myth” of public interest in the reuse of open public data.


The aim of this paper is to explore the concept of public sector information (PSI), what it is, its history and evolution, what constitutes its corpus of documents and the issues and challenges it presents to society, its institutions and to those who use and manage it. The paper, by examining the literatures of the law, political science, civil society, economics and information and library science explores the inherent tensions of access to and use of PSI—pragmatism vs. idealism; openness vs. secrecy; commerce vs. altruism; property vs. commons; public good vs. private good. It focusses on open government data (OGD)—a subset of what is popularly referred to as ‘big data’—its background and development since much of the current debate of its use concerns its commercial value for both the private sector and the public sector itself. In particular it looks at the information itself which, driven by technologies of networks, data mining and visualisation gives value in industrial and economic terms, and in its ability to enable new ideas and knowledge.

This paper looks at the role of the European Directive on re-use of public sector information in the current trend towards opening up government data. After discussing the PSI directive, it gives an overview of current policies and practices with regard to open government data in the Member States. It is argued that the success of the open government data movement in some Member States can be related to the confusion or ignorance about the relationship between traditional freedom of information legislation and the re-use of public sector data. If future information policies decide to follow this trend, they should always ensure that existing rights on freedom of information are not harmed.


Although a significant number of public organizations have embraced the idea of open data, many are still reluctant to do this. One root cause is that the publicizing of data represents a shift from a closed to an open system of governance, which has a significant impact upon the relationships between public agencies and the users of open data. Yet no systematic research is available which compares the benefits of an open data with the barriers to its adoption. Based on interviews and a workshop, the benefits and adoption barriers for open data have been derived. The findings show that a gap
exists between the promised benefits and barriers. They furthermore suggest that a conceptually simplistic view is often adopted with regard to open data, one which automatically correlates the publicizing of data with use and benefits. Five ‘myths’ are formulated promoting the use of open data and placing the expectations within a realistic perspective. Further, the recommendation is given to take a user’s view and to actively govern the relationship between government and its users.


The exponentially growing production of data and the social trend towards openness and sharing are powerful forces that are changing the global economy and society. Governments around the world have become active participants in this evolution, opening up their data for access and re-use by public and private agents alike. The phenomenon of Open Government Data has spread around the world in the last four years, driven by the widely held belief that use of Open Government Data has the ability to generate both economic and social value. However, a cursory review of the popular press, as well as an investigation of academic research and empirical data, reveals the need to further understand the relationship between Open Government Data and value. In this paper, the authors focus on how use of Open Government Data can bring about new innovative solutions that can generate social and economic value. The authors apply a critical realist approach to a case study analysis to uncover the mechanisms that can explain how data is transformed to value. They explore the case of Opower, a pioneer in using and transforming data to induce a behavioural change that has resulted in a considerable reduction in energy use over the last six years.

In the last years, several research endeavours were launched aiming at involving popular social media platforms in electronic participation. These early endeavours seem to present some essential limitations related mainly to scalability and uptake. In order to avoid these limitations, the authors introduce a two-phased approach for supporting participatory decision-making based on the integration and analysis of social and government open data. The proposed approach is based on the literature related to the analysis of massive amounts of social data for future events prediction. In this paper the authors also present a Web data driven architecture for the implementation of the proposed approach. The architecture is based on the use of linked data paradigm as a layer that will enable integration of data from different sources. The authors anticipate that the proposed approach will (i) allow decision makers to understand and predict public opinion and reaction about specific decisions; and (ii) enable citizens to inadvertently contribute in decision-making.


Open government data (OGD) refers to making public sector information freely available in open formats and ways that enable public access and facilitate exploitation. Lately, a large number of OGD initiatives launched worldwide aiming to implement one-stop portals acting as single points of access to governmental data. At the same time, the so-called linked data technologies emerged aiming at publishing structured data on the web in such a way that enables semantically enriching data, uniform access to data, and linking of data. In this paper, the authors first propose a classification scheme for OGD initiatives based on the relevant literature. The authors thereafter review and analyse OGD initiatives based on the proposed scheme. The authors finally present an architecture and prototype implementation for the most advanced OGD class in the scheme, which enables linking decentralised data.

Public sector information constitutes a valuable primary material for added-value services and products, which however remains unexploited. Recently, Open Government Data (OGD) initiatives emerged worldwide aiming to make public data freely available to everyone, without limiting restrictions. Despite its potential however there is currently a lack of roadmaps, guidelines and benchmarking frameworks to drive and measure OGD progress. This is particularly true as proposed stage models for measuring eGovernment progress focus on services and do not sufficiently consider data. In this paper, the authors capitalize on literature on eGovernment stage models and OGD initiatives to propose a stage model for OGD. The proposed model has two main dimensions, namely organizational & technological complexity and added value for data consumers. The authors anticipate the proposed model will open up a scientific discussion on OGD stage models and will be used by practitioners for constructing roadmaps and for benchmarking just like the European Union stage model is currently used for measuring public service online sophistication.


Public sector bodies maintain a large amount of data from various domains. This data represents a potential resource that organizations and individuals can use to enhance their own datasets or which can be used to develop new and innovative products and services. In order to foster the reuse of the data held by the public sector bodies a number of countries around the world has started to publish its data according to the Open Data principles. In this paper the authors present a set of benefits that can be achieved by publishing Open Government Data (OGD) and a set of risks that should be assessed when a dataset is considered for opening up. Benefits and risks presented in this paper were mostly identified during two OGD activities.

As Open Data becomes commonplace, methods are needed to integrate disparate data from a variety of sources. Although Linked Data design has promise for integrating world wide data, integrators often struggle to provide appropriate transparency for their sources and transformations. Without this transparency, cautious consumers are unlikely to find enough information to allow them to trust third party content. While capturing provenance in RPI’s Linking Open Government Data project, the authors were faced with the common problem that only a portion of provenance that is captured is effectively used. Using the water quality portal’s use case as an example, the authors argue that one key to enabling provenance use is a better treatment of provenance granularity. To address this challenge, the authors have designed an approach that supports deriving abstracted provenance from granular provenance in an open environment. The authors describe the approach, show how it addresses the naturally occurring unmet provenance needs in a family of applications, and describe how the approach addresses similar problems in open provenance and open data environments.


This paper aims to explore the role of records management in supporting the effective use of information visualisation and visual analytics (VA) to meet the challenges associated with the analysis of Big Data. This exploratory research entailed conducting and analysing interviews with a convenience sample of visual analysts and VA tool developers, affiliated with a major VA institute, to gain a deeper understanding of data-related issues that constrain or prevent effective visual analysis of large data sets or the use of VA tools, and analysing key emergent themes related to data challenges to map them to records management controls that may be used to address them. The authors identify key data-related issues that constrain or prevent effective visual analysis of large data sets or the use of VA tools, and identify records management controls that may be used to address these data-related issues. This paper discusses a relatively new field, VA, which has emerged in response to meeting the challenge of analysing big,
open data. It contributes a small exploratory research study aimed at helping records professionals understand the data challenges faced by visual analysts and, by extension, data scientists for the analysis of large and heterogeneous data sets. It further aims to help records professionals identify how records management controls may be used to address data issues in the context of VA.


This chapter describes issues with data quality, value and traceability and questions the use of open data as a mechanism for government accountability. It describes the Norwegian approach to opening government information as an alternative that respects the need for information integrity.


An introduction is presented which discusses reports in the issue including one on opening government information to the underserved, one on Dutch open government policies, and one on web sites providing news and other local information.


This article, which is one of the products of an international collaborative research initiative called iTrust, aims to explore these questions and offer suggestions concerning how the issues they raise can be addressed. The article describes the results of the first stage in a multi-stage research project leading to methods for developing retention and disposition specifications and formal schedules for open data and big data initiatives. A fictitious organization is used to describe the characteristics of open data and big data initiatives, the gap between current approaches to setting retention and disposition specifications and schedules and what is required and how that gap can be
closed. The landscape described as a result of this stage in the research will be tested in case studies established in the second stage of the project. The argument is made that the business processes supporting open data and big data initiatives could serve as the basis for developing enhanced standards and procedures that are relevant to the characteristics of these two kinds of initiatives. The point is also made, however, that addressing the retention and disposition issues requires knowledge and leadership, both of which are in short supply in many organizations. The characteristics, the issues and the approaches will be tested through case studies and consultations with those involved with managing and administering big data and open data initiatives. There is very little, if any, current literature that addresses the impact of big data and open data on the development and application of retention schedules. The outcome of the research will benefit those who are seeking to establish processes leading to formally approved retention and disposition specifications, as well as an instrument the approved retention and disposal schedule designed to ensure the ongoing integrity of the records and data associated with big data and open data initiatives.


There is a growing body of research considering the opportunities and challenges associated with open government data. To date research considering the potential medium to long-term (say 2-20 years) impact of open government data has focussed on individual potential pathways. The open government data agenda seeks to drive a transition in a complex socio-technical system of data providers, intermediaries and users. In this paper the authors outline how socio-technical systems transition theory could be applied in a UK open government data context and sketch three transition pathways that might emerge. By doing so the authors hope to foster a collaborative effort, across the web science research community, to explore and develop pathways for open government data.

A loose coalition of advocates for Open Government Data (OGD) argue that such data has the potential to have a transformative impact by catalysing innovation across sectors of the economy and by fostering democratic participation and engagement. However, there is limited evidence to date of the OGD agenda having such a transformative impact. This article applies the Multi-level Perspective (MLP); an approach more typically applied to study transitions to a sustainable society, to explore the nature of the barriers currently faced by the OGD agenda. It argues that such barriers exist in two forms: implementation barriers and barriers to use. The empirical results presented include survey responses measuring the perceptions of U.K. OGD community members of 33 barriers to the OGD data agenda. These results are analysed to identify implications for OGD policy and practice. The article concludes with a start at establishing a theoretical grounding for the study of barriers to the OGD agenda.


This paper describes the results of the LOD2 Open Government Data Stakeholder Survey 2010 (OGD Stakeholder Survey). The objective of the survey was to involve as many relevant stakeholders as possible in the 27 European Union countries in an online questionnaire and ask them about their needs and requirements in the area of open data as well as for the publicdata.eu portal. The main areas of the survey have been questions about Open Government Data itself, questions about data, about the usage of data, questions about the requirements for a centralised data catalogue as well as questions about the participants themselves. The goal of the OGD Stakeholder Survey has been to reach a broad audience of the main stakeholders of open data: citizens, public administration, politics and industry. In the course of the survey that was open for 5 weeks from November 2010 to December 2010 in total 329 participants completed the survey. The results have been published in April 2011 in the form of HTML and PDF, the raw data in CSV. In addition to these publication formats (HTML, PDF, CSV) the authors published the data also as Linked Data using various vocabularies and tools.
The article argues that data consumers must be able to trust the data's provenance, and that a descriptive model enables consumers to make informed choices about data sources.


While Open Data initiatives are diverse, they aim to create and contribute to public value. Yet several potential contradictions exist between public values, such as trust, transparency, privacy, and security, and Open Data policies. To bridge these contradictions, the authors present the notion of precommitment as a restriction of one’s choices. Conceptualized as a policy instrument, precommitment can be applied by an organization to restrict the extent to which an Open Data policy might conflict with public values. To illustrate the use of precommitment, the authors present two case studies at two public sector organizations, where precommitment is applied during a data request procedure to reconcile conflicting values. In this procedure, precommitment is operationalized in three phases. In the first phase, restrictions are defined on the type and the content of the data that might be requested. The second phase involves the preparation of the data to be delivered according to legal requirements and the decisions taken in phase 1. Data preparation includes amongst others the deletion of privacy sensitive or other problematic attributes. Finally, phase 3 pertains to the establishment of the conditions of reuse of the data, limiting the use to restricted user groups or opening the data for everyone.


The article aims to investigate how key e-Governance dimensions related to openness, such as transparency and accountability, which are a necessary condition for reaching a high maturity of e-Government, may not be sufficient for open government. For this purpose, an interpretative framework to identify country attitudes towards Open Government is proposed and it is applied to two cases drawn from different legal, cultural and organisational backgrounds. Among the key findings of the article, the 'attitudes mapping' resulting from the application of the interpretative framework to the case studies points out the key role of different governance traditions in the path towards open government.


"Legal interoperability" could be defined as the possibility of legally mixing data coming from different sources (including governmental data, data generated by online communities and data held by private parties). Legal interoperability is similar to technical interoperability, since it is a prerequisite for mixing data and create new knowledge or services. But it also has its own peculiarities, for instance because it could be achieved simply choosing the appropriate licensing scheme, but also because self-help mechanisms which could - at a certain price - guarantee technical interoperability to third parties cannot (lawfully) solve legal interoperability issues. In the mid/long run, legal interoperability could be achieved through the evolution of legal frameworks in order to harmonize the landscape of Government Data. In the short term, the shortcomings generated by diversified legal frameworks may be alleviated through the careful choice of copyright licenses. The presentation will focus on the latter aspects, discussing existing public licenses (such as the Creative Commons and Open Data Commons ones), representing a de facto standard in this domain, and the main open data licenses developed by European governments (e.g. the Open Government Licenses in the UK, the French License Ouverte or the Italian Open Data License).
https://www.fixmystreet.com/.

The UK’s first nationwide resource for reporting local problems (such as fly-tipping or broken street paving), Fix My Street allows users to report problems either through entering the postcode of the area they encountered the issue, matched using data from the Ordinance Survey, or by allowing the website and/or application to have access to their location through geographical mapping data. A report, which users can augment with photographs, is passed to the responsible body for resolution, and published online. The postcode mechanism also allows for users to see all the problems that have been reported within a specific area, allowing for communal awareness and support. If users opt to create an account on the website, they are then given the option to view a resource of all reports that they have filed. Fix My Street, which is run by the charity mySociety, also offers a bespoke service for local government, which integrates the tool with existing websites. Currently, several UK councils have partnered with Fix My Street, including Stevenage, Bromley, Barnet, Oxfordshire, Hart, East Sussex and Greenwich.


Run by the charity mySociety since 2006, They Work For You uses data from the UK Parliament and presents it digitally in an understandable and accessible manner. Users supply a postcode (matched using data from the Ordinance Survey), and are taken to the record for their Member of Parliament, Member of Scottish Parliament or Member of the Legislative Assembly (Northern Ireland). Each parliamentarian page contained information regarding votes, recent appearances, a parliamentarian profile and statistics, as well as a database of speeches. Users can also create alerts for specific members, or for particular issues, and will be emailed when either the individual or issue are mentioned within Parliament.
A partnership between the National Health Service (NHS) in England, the Tavistock Institute, Involve, the Democratic Society and Public-i, NHS Citizen is a national programme, which aims to get citizens of the UK involved with the NHS as owners, enabling them to influence NHS decision making. The platform allows users to post ideas on an online forum, ‘NHS Citizen Gather,’ or using the initiative’s Facebook and Twitter pages. From here, 10 ideas are taken forward to go before a Citizen’s Jury, consisting of 15 members of the public. Five ideas will then be taken forward to NHS Assembly Meetings, to be held twice yearly. NHS Citizen cites the importance of citizens as its primary data source, as asserted in the project’s method statement, where citizen’s engagement and ideas are posited as ‘evidence’ for refinement.


Open Government Data (OGD) is a government policy that is being increasingly adopted by several countries around the world. Despite its growing acceptance, there are still several difficulties both in making OGD available and ensuring it is used effectively by society. A growing body of knowledge has been produced in the last few years in the e-government scientific community to address these problems. This paper conducts a systematic literature re-view on the question of OGD, which is aimed at outlining a knowledge base of the existing scientific literature on OGD produced in the period 2007-2012 that is available in the on-line databases of relevant scientific societies (ACM, AIS, IEEE). In doing so, the paper provides an overview of this knowledge base, and maps out the different research approaches that are being adopted to study OGD, its geographical scope and the application domain they refer to. The results indicate the need of broadening the geographic coverage of OGD studies and also of deepening the discussion of quality criteria for assessing open data.
This paper analyses the implementation of the Open Government Data strategy and portal of the City of Vienna. This evaluation is based on qualitative interviews and online polls after the strategy was implemented. Two groups of users were involved in the evaluation: internal target groups (employees and heads of department in the City of Vienna's public administration departments) and external stakeholders (citizens, business representatives, science and research, journalists). Analysed aspects included the present organizational processes, the benefits (to business and society), and requirements for future Open Government Data initiatives. This evaluation reveals success factors which accompanied the implementation: the clear definition of responsibilities and the implementation along a process model, the integration of the Open Government Data platform into existing Content Management Systems, the evaluation of the Open Government Data initiative very shortly after its inception. Based on the theoretical and empirical findings, recommendations for future Open Government Data strategies are made which target the local authority and would require action on the federal level such as Creative Commons Attribution License as the default for subsidy funds or public relation measures carried out directly by the data providing departments.

Open Government Data initiatives and particularly Open Government Data portals have proliferated since the late 2000’s. A comprehensive analysis of the capabilities and potential of these initiatives is currently missing from the recent research literature. In order to address this gap, the paper at hand aims towards analysing the landscape of Open Governmental Data in the European Union from a functional, semantic and technical perspective. The research focused on the collection and categorization of an
indicative number of public data sources for each of the 27 European Union country-members through investigating their services and characteristics. By modelling and classifying the data sources according to their key attributes, the authors were able to proceed to their statistical analysis and assessment in terms of their content, licensing, multilingual support, acquisition, ease of access, provision and data format. The results portray the current quality of Public Sector Information infrastructures and highlight what still needs to be done in order to make public data truly open and readily available for researchers, citizens, companies and innovation in general.


This article describes trends of open data development and a new discipline, which was formed largely due to the fact that the data have become available and open on the Internet. The authors provide a brief overview of the main directions in the development of open data and data journalism: educational projects, interaction with the community of developers using data management platforms, development of business community on open data basis. The article also discusses Russian educational projects dealing with open data and data journalism.


In this poster, the authors present the International Open Government Dataset Search (IOGDS). IOGDS is a faceted browsing interface for searching over more than one million open government datasets from around the world. The authors present ongoing research and development towards the improved discovery and access to open government data. IOGDS has been designed around S2S, a framework for constructing user interfaces over Semantic Web services. The authors discuss how the Web service abstraction layer in the S2S framework can be extended to support the use of external
vocabularies to aid in the discovery of open government datasets. The authors also discuss how the S2S framework can provide localization and context for non-expert users to improve the accessibility of open government data.


This paper presents a preliminary exploration on the status of open government data worldwide as well as in-depth evaluation of selected open government data portals. Using web content analysis of the open government data portals from 35 countries, this study outlines the progress of open government data efforts at the national government level. This paper also conducted in-depth evaluation of selected cases to justify the application of a proposed framework for understanding the status of open government data initiatives. This paper suggest that findings of this exploration offer a new-level of understanding of the depth, breath, and impact of current open government data efforts. The review results also point to the different stages of open government data portal development in term of data content, data manipulation capability and participatory and engagement capability. This finding suggests that development of open government portal follows an incremental approach similar to those of e-government development stages in general. Subsequently, this paper offers several observations in terms of policy and practical implication of open government data portal development drawn from the application of the proposed framework.


This paper aims to share the contribution of records managers to the Open Government in the City Council of Girona (Catalonia), based on the strategy of participation in the Open Data project. The contribution of the Records Management Department is to facilitate locating data sets and to be responsible for the data selection processes. Records Management allows a first global identification of the functions, activities and
producers, using the Records Classification Scheme and the General Register of Case Files. In addition, it can obtain a greater detail of information using the Application Catalogue and the Register of Personal Data Files. Records Management can contribute to data selection in the Open Data projects by the appraisal of data, also taking into consideration functions and records of the organization. At the same time, Open Data projects could reinforce the strategies of records preservation in databases, because the same extraction processes could be shared. In addition, Open Data projects could mean an opportunity for Historical Archives if they take charge of the maintenance of the Open Data Archive. This could be relevant in the future for the transformation of data from information systems into documentary heritage. The opening of the city’s Open Government portal is scheduled for the beginning of 2014. Thus, it is necessary to consider this work as a work in progress. This work is a specific case study of the contribution of Records Management to the Open Data projects; its proposals and reflections have been made from a very practical point of view.


Government data is powerful, plentiful, and relevant to citizens’ concerns. Making it open supports transparency, crowdsourced data enhancement, and innovative service development. The authors review the state of linked open government data, in the context of the potential for the publishing organizations and the Linked Data Web itself, as well as the administrative and political issues raised.


In this article a maturity model is proposed, named OD-MM (Open Data Maturity Model) to assess the commitment and capabilities of public agencies in pursuing the principles and practices of open data. The OD-MM model has a three level hierarchical structure, called domains, sub-domains and critical variables. Four capacity levels are
defined for each of the 33 critical variables distributed in nine sub-domains in order to determine the organization maturity level. The model is a very valuable diagnosis tool for public services, given it shows all weaknesses and the way (a roadmap) to progress in the implementation of open data.


The Open Data movement has assumed that accurate data is available to demonstrate government expenditure and actions. However, in many countries, the records of government policies, activities and transactions that should provide the basis for a large proportion of government data are incomplete, inaccurate or inaccessible. This will result in inaccurate data. This paper suggests that alongside the enthusiasm for Open Data, there needs to be parallel attention to ensuring that government records are managed as the evidence base for Open Government. The paper explores key records issues that will affect Open Data and draws attention to the value of international records management standards.


This set of three case studies explores the intersection of openness, digital governance, and high quality information in Estonia, Finland, and Norway with the aim of identifying lessons that will support the same objectives in lower resource countries. Openness, a key aspect of the international agenda for increasing transparency and accountability, for reducing public sector corruption, and for strengthening economic performance, rests on the principle that citizens have a right to know what their governments are doing and to benefit from using government information. Goals for open, accountable, and inclusive governance rest on the assumption that trustworthy information is available and can be shared meaningfully through strategies for digital governance. This assumption needs to be examined. Does reliable and complete information exist across lower resource countries? Can it be accessed readily? Will it survive through time?
In order to increase transparency and civic participation, governments around the world sought ways to open their data and allow both to itself as to companies and the civil population a greater share in the maintenance, surveillance and optimization of the services provided. To this end, using a technology called linked data (LD), the data released by governments became easier to be understood and properly used by humans and machines alike, thus creating what today is called linked open government data (LOGD). The purpose of this article is to present the state of art of the research in LOGD through bibliometric research, ultimately presenting a feedback on the matter.


Transparent Chennai marries data created for governmental purposes and user generated content to advocate for issues facing citizens of Chennai, and in particular, issues facing the poor. Available in both English and Tamil, Transparent Chennai creates maps and datasets, with the aim of enabling the area’s poor to have a greater voice within planning and city governance. The site uses mapping data, provided through Google, which can be overlaid with a variety of datasets, from user-generated data (such as the location of slum eviction sites) to governmental data, such as the location of ward boundaries.

Open Government Data (OGD) initiatives, and in particular the development of OGD portals, have proliferated since the mid-2000s both at central and local government levels in OECD and non OECD countries. Understanding the preconditions that enable the efficient and effective implementation of these initiatives is essential for achieving their overall objectives. This is especially true in terms of the role played by OGD in relation to Open Government policies in general. This paper highlights the main principles, concepts and criteria framing open government data initiatives and the issues challenging their implementation. It underlines the opportunities that OGD and data analytics may offer policy makers, while providing a note of caution on the challenges this agenda poses for the public sector. Finally, the overall analysis of key concepts and issues aims to pave the way for an empirical analysis of OGD initiatives. So far, little has been done to analyse and prove the impact and accrued value of these initiatives. The paper suggests a methodology comprising an analytical framework for OGD initiatives (to be applied to ex post and ex ante analysis of initiatives) and a related set of data to be collected across OECD countries. The application of the analytical framework and the collection of data would enable the acquisition of a solid body of evidence that could ultimately lead to mapping initiatives across OECD countries (i.e. a typography of initiatives) and developing a common set of metrics to consistently assess impact and value creation within and across countries.


This paper presents a benchmark proposal for the Open Government and its application from the open data perspective using data available on the U.S. government's open data portal (data.gov). The benchmark is developed over the adopted Open Government conceptual model, which describes Open Government through data openness, transparency, participation and collaboration. Resulting in two measures, that is, one known as the e-government openness index (eGovOI) and the other Maturity, the benchmark indicates the progress of government over time, the efficiency of recognizing and implementing new concepts and the willingness of the government to recognize and embrace innovative ideas.
The right to information springs from the right to expression under the notion that in order for an individual to be able to freely express ideas, opinions and thoughts, it should be able to freely formulate them, hence to be very well informed. The Right to Information (RTI) movement focuses on the fact that government information is being collected for the benefit of the public, and not for the benefit of the organisations themselves, and it constitutes the basis of a democratic regime. On the other hand, Open Government Data (OGD) refers to data produced or commissioned by government or government controlled entities, which can be freely used, reused and redistributed by anyone. The core of OGD movement lays into the proactive dissemination of unstructured raw data aiming firstly to innovation and economic growth by exploiting the given data to produce new products with added value, and secondly to accountability and transparency by providing access to data that bears political value on the way governments and administrations work. This paper attempts to present the underlined ideology of these two movements, which have fashioned the access to government information and to substantiate that the way to the future is a one-way street towards the greatest possible dissemination of government information as this can be established by the synergy of RTI and OGD in order to preserve and promote the primary human and political right to receive information so as to be informed and able to freely express opinions, ideas and thoughts.


Open government data (OGD) is growing steadily, and it has been a vital communication channel between government and citizens. Government data published by linked data technology can promote transparency and even improve reuse rate of data. The paper describes the related concepts, publishing principles, and advantages of linked government data. It then deeply analyses technology architecture which includes expression layer, compiling layer, interlinking, and browse/query layer on the basis of
related concepts. Following this, the paper summarizes applications of linked government data in the developed countries such as the United States and the United Kingdom. Finally, it gets a view of applications and challenges of linked government data in China smart city construction.


Governments create large amounts of data. However, the publication of open data is often cumbersome and there are no standard procedures and processes for opening data. This blocks the easy publication of government data. The purpose of this paper is to derive design principles for improving the open data publishing process of public organizations. Action Design Research (ADR) was employed to derive design principles. The literature was used as a foundation, and discussion sessions with civil servants were used to evaluate the usefulness of the principles. Barriers preventing easy and low-cost publication of open data were identified and connected to design principles, which can be used to guide the design of an open data publishing process.

Five new principles are 1) start thinking about the opening of data at the beginning of the process, 2) develop guidelines, especially about privacy and policy sensitivity of data, 3) provide decision support by integrating insight in the activities of other actors involved in the publishing process, 4) make data publication an integral, well-defined and standardized part of daily procedures and routines, 5) monitor how the published data are reused. The principles are derived using ADR in a single case. A next step can be to investigate multiple comparative case studies and detail the principles further. The authors recommend using these principles to develop a reference architecture. The design principles can be used by public organizations to improve their open data publishing processes. The design principles are derived from practice and discussed with practitioners. The discussions showed that the principles could improve the publication process. Decreasing the barriers for publishing open government data could result in the publication of more open data. These open data can then be used and stimulate various public values, such as transparency, accountability, innovation,
economic growth and informed decision and policy-making. Publishing data by public organizations is a complex and ill-understood activity. The lack of suitable business processes and the unclear division of responsibilities blocks publication of open data. This paper contributes to the literature by presenting design principles which can be used to improve the open data publishing process of public sector organizations.


Reports and research appears to assume that the benefits of open data dominate open data’s negative consequences. Moreover, much of the existing research discusses benefits and disadvantages on a high level without providing much detailed insight in the underlying processes. Yet many governments are reluctant to open their data, as they are afraid of possible negative consequences of opening data. The objective of this policy paper is to better understand the aspects of the dark side of open data and contributes to the literature by providing a more realistic perspective on open data. The authors conducted nineteen in depth interviews with public sector officials and data archivists and identified sixteen categories of negative effects. For the dark side inherent to open data efforts the research suggests that a context and dataset dependent decision-making model needs to be made weighing the benefits of open data on the one hand (e.g. creating transparency, the possibility to strengthen economic growth), and the risks and disadvantages of open data (e.g. violating privacy and possible misuse and misinterpretation of data) on the other hand.


In developing open data policies, governments aim to stimulate and guide the publication of government data and to gain advantages from its use. Currently there is a multiplicity of open data policies at various levels of government, whereas very little systematic and structured research has been done on the issues that are covered by open
data policies, their intent and actual impact. Furthermore, no suitable framework for comparing open data policies is available, as open data is a recent phenomenon and is thus in an early stage of development. In order to help bring about a better understanding of the common and differentiating elements in the policies and to identify the factors affecting the variation in policies, this paper develops a framework for comparing open data policies. The framework includes the factors of environment and context, policy content, performance indicators and public values. Using this framework, seven Dutch governmental policies at different government levels are compared. The comparison shows both similarities and differences among open data policies, providing opportunities to learn from each other's policies. The findings suggest that current policies are rather inward looking, open data policies can be improved by collaborating with other organizations, focusing on the impact of the policy, stimulating the use of open data and looking at the need to create a culture in which publicizing data is incorporated in daily working processes. The findings could contribute to the development of new open data policies and the improvement of existing open data policies.
A work of fundamental significance to the current project is Novais, De Albuquerque, and Craveiro's (2013) literature review of open government data from 2007 to 2012. The present literature review predominantly concerns work from 2012 to 2015, not least because it is primarily concerned with records and archives management contributions to open government data, or areas of open government data work to which the records / archives management profession could make useful inputs, which has arisen as a subject of enquiry only since 2012.

The present literature review found that the number of resources available for the study of open government data is growing rapidly, but this rapid growth is not necessarily reflected in the traditional means of scholarly communication, such as peer-reviewed journals. Instead, much of the discourse about open government data is spread across various media, platforms and formats, and takes place between various actors including governments, academics, non-government organisations and activists, both as individuals and participants in networks of varying degrees of formality and coherence. This dynamic environment presents challenges to the study of open government data, as the exploration of specific ideas might unfold conversationally, across various platforms with differing levels of permanence or persistence. For example, there has been a great deal of discussion in the United Kingdom about government registers of company ownership data. This discussion has taken place over social media platforms (particularly Twitter), civil society listservs and online fora (either public or membership-based), face to face meetings and teleconferences that may or may not be reported via NGO websites and blogs, and opinion pieces published in the mainstream media. However, this energetic discussion has not yet been captured in a journal article, so that, to judge from a traditional literature review company data does not appear to be a significant issue.

The particular interest of this literature review is the intersection between open government data and records and archives management. The review has found that this is a vastly under-researched area that warrants much more attention than it is currently receiving either from the open data or the records and archives management communities.
Following Thurston’s 2012 piece “Trustworthy Records and Open Data” in The Journal of Community Informatics (Thurston 2012), the records / archives management field has begun to engage with open data. 2014 seems to have been a particular turning point, with the publication of Borglund and Engvall’s (2014) “Open Data?: Data, Information, Document or Record?”, Lemieux, Gormly, and Rowledge's (2014) “Meeting Big Data Challenges with Visual Analytics: The Role of Records Management”, Lowry's (2014) “Opening Government: Open Data and Access to Information”, McDonald and Léveillé's (2014) “Whither the Retention Schedule in the Era of Big Data and Open Data?” , and Serra's (2014) “The Mapping, Selecting and Opening of Data: The Records Management Contribution to the Open Data Project in Girona City Council”, which reports what seems to represent the first attempt (the city of Helsinki’s early work on uniting open data and records management has not been written about in English) to extract data from a records management system as a basis for open data derived from a source that enables reliable contextual metadata to be provided, which helps to underpin the integrity of the data.

The attention to open data from the records management community continued into 2015 with work from Léveillé and Timms (2015) and Thurston (2015), the latter the result of an in-depth study of openness and information integrity in Estonia, Finland and Norway under the auspices of the World Bank. 2015 also saw the further development of records management guidance in the Open Government Guide, an online resource developed by international civil society organisations to support governments in developing commitments for Open Government Partnership national action plans (“Records Management” 2015). Much of this work is concerned with introducing controls to ensure the trustworthiness of data, which had appeared infrequently in the open data literature (two rare examples are Ceolin et al.’s (2013) “Reliability Analyses of Open Government Data” and Ceolin et al.’s (2014) “Two Procedures for Analyzing the Reliability of Open Government Data”).

Scholars in the area of e-government are approaching open government data from their own perspective. Examples include Ngwenya, Lubbe, and Klopper's (2012) thinking about openness and e-government in the developing world, Milić, Veljković, and Stoimenov's (2012) framework for data mining, and Misuraca and Viscusi's (2014) questions about the usefulness of raw data, which obviously has synergies with questions coming out of the records and archives management discipline. Given the work that went on in the late 1990s and 2000s to align records management with e-government, it is easy to see that this work
from the e-government community could be useful context for records / archives interventions into the open government data discourse, particularly regarding the value of contextual information for evidence of activity in the digital environment.

An important finding of this review is that open government data is an area of convergence between disparate disciplines, each with its own vocabulary. Terms such as ‘quality’, ‘value’, and ‘persistence’ are used differently by different communities. For instance, provenance has a particular meaning in archival science, but it is also used in data management, for instance in Lebo et al.'s (2012) “Towards Unified Provenance Granularities” in Groth and Frew’s (eds.) Provenance and Annotation of Data and Processes. It is easy to envisage a large-scale interdisciplinary research project that attempts to map and harmonise these professional vocabularies with a view to facilitating work between disciplines. The InterPARES terminology database could serve as a point of departure for this work.

The primary interest of the open data community has been securing access to government information and distributing it, with increasing attention on the means of delivering information (the raw data or representations of it) to end users, such as civil society. Relatively little attention was given to issues of data quality in the early enthusiasm for data openness. However, we can see, from this literature review, signs of a recognition from the open data community that data quality is a significant issue in determining the value and use of open data. This is demonstrated in, for example, Both's (2012) article, where value to end users is considered. Even a rudimentary correlation between data value and appraisal, as a concept in archival science, suggests possible lines of enquiry that could yield results beneficial to data retention practices and user satisfaction.

Ideas of data quality in relation to data handling – particularly that there is ‘no such thing’ as raw data - seem to have started to emerge in the literature in 2013, with Davies and Frank's (2013) article on the ‘socio-technical life of a government dataset’. Denis and Goëta (2014) took this a step further with “Exploration, Extraction and ‘Rawification’. The Shaping of Transparency in the Back Rooms of Open Data.” This work provides an avenue into the conversation about open data for records managers, whose concern with trustworthiness and authenticity has led to the creation of a body of technical knowledge that could be presented to what is essentially a new user community: open data proponents, analysts and users.
Links between open government data and policy-making, or governance more broadly, are fairly ‘old’ and well-established in the literature. In 2010, Napoli and Karaganis explored the use of ‘publicly available data’ in US communications policy. In 2011, Kalampos, Hausenblas, and Tarabanis explored the use of open data in participatory decision-making, and in 2012, Rajshree and Srivastava looked at the use of open government data in fighting corruption, which has long been a focus for thinking about the value of records as evidence. More recently, a number of resources have considered the socio-political implications of open government data. This is epitomised in the work of Jo Bates, whose work has raised questions about the UK government’s openness agenda and the appropriateness of its tactics in involving civil society in policy making (see, for instance, the 2012 article “‘This Is What Modern Deregulation Looks Like’: Co-Optation and Contestation in the Shaping of the UK’s Open Government Data Initiative.”). Echoing the observations that Darch and Underwood (2010) made about the correlations between Freedom of Information and neoliberal ideas, Bates has recently asked questions of the political ramifications of open data in the UK (see Bates “The Strategic Importance of Information Policy for the Contemporary Neoliberal State: The Case of Open Government Data in the United Kingdom”, 2014). In view of some of the current thinking about records and archives in relation to political power, this work could provide a useful means of extending the socio-political line of enquiry between the fields of open data and records and archives.

At a more practical level, there is a body of technical guidance on opening government data, much of it emanating from civil society or from government-funded organisations such as the Open Data Institute, which has produced the following guidance:

- “Benchmarking Open Data Automatically | Guides” 2015
- “Finding Creative Use for Public Data” 2015
- “How to Make a Business Case for Open Data | Guides” 2015
- “How to Plan and Budget an Open Data Initiative | Guides” 2015
- “How to Prioritise Open Data to Drive Global Development | Guides” 2015
- “Open Data Maturity Model | Guides” 2015

A preliminary review of the documents suggests an absence of records / archives technical knowledge. For instance, there is nothing in this subset about data preservation over time, and
the work that has been done on web archiving and web crawling is not presented to data users in guides of this kind.

As the scholarly writing on open government data becomes more diverse, we see broader issues appearing in the literature. For instance, people are now thinking about legacy issues: Façanha and Cavalcanti (2014) have written on bringing government legacy system data into the public domain.

Studies of data visualisation, though often found in various fields, such as computer science, data management, and statistics, are still rare in the scholarly literature on open government data. A notable example is by Graves and Hendler (2014). It is interesting to note that the other two principal examples are Kukimoto's (2014) work “Open Government Data Visualization System to Facilitate Evidence-Based Debate Using a Large-Scale Interactive Display”, which introduces the language of ‘evidence’ into data visualisation, and Lemieux, Gormly, and Rowledge's (2014) work, “Meeting Big Data Challenges with Visual Analytics: The Role of Records Management”, which appears to be the only work to date that links data visualisation with records management.

Another emerging trend in the open government data literature is the attempt to deal with privacy and data anonymisation. Examples include Meijer, Conradie, and Choenni's (2014) “Reconciling Contradictions of Open Data Regarding Transparency, Privacy, Security and Trust” and the Open Data Institute’s (2015) “Save the Titanic: Hands-on Anonymisation and Risk Control of Publishing Open Data | Guides.” This is clearly an area where records and archives professionals could be making contributions.

There is a great deal of interest in the open data community in how open data can facilitate innovation, but this is not reflected in the scholarly literature, where Jetzek, Avital, and Bjorn-Andersen's (2014) article is one of the few examples of research that considers the value of open data for innovation. It is possible that a useful line of research may be to explore the impact on innovation of open data that is not derived from reliable records sources.

The literature review revealed only one example of research into linkages between open data and social media (Alexopoulos et al. 2014). Given InterPARES Trust’s concern for
information integrity in the online, cloud and social media environments, it seems that some of the emerging work on social media could be fed into this.

There is a discrete body of online resources that have gone unexamined by the records and archives professional community. These resources – civic technologies – take government data and present them to users in visual or interactive ways for the purposes of effecting social change. Broadly defined, civic technology is technology that intersects with public life, enabling engagement or participation of the public with the government for a variety of purposes, from making government more transparent and accountable, to enhancing civic communities, and impacting policy decisions. The Knight Foundation (2015) has identified two specific strands of civic technology: open government and community action. From this, civic technology can be further divided into six streams: public data access and transparency; social causes and civic engagement; place based networks and community forums; funding for projects enhancing public services and spaces; and peer to peer sharing of resident owned goods and services (Knight Foundation, 2015). Civic technology, therefore, can be embodied through a variety of initiatives: an application that enables residents to share goods and skills and a website allowing resident feedback on environmental issues posited by local government are both examples of civic technology. However, a common thread is a foregrounding of the user-citizen, with a focus on their empowerment and engagement within the community, and therefore the political process, through technological means.

This literature review looked at a number of civic technologies, and analysed the following three examples in the annotated bibliography (section 2 of this report):

- GotToVote: http://gottovote.cc/
- Fix my Street: https://www.fixmystreet.com
- They Work For You: http://www.theyworkforyou.com
- Transparent Chennai: http://www.transparentchennai.com/
- NHS Citizen: https://www.nhscitizen.org.uk
- Quién te financia: http://quientefinancia.cl/

It is clear from the literature that civic technologies provide opportunities for case studies of the data management / handling practices into which records management could feed. For
example, Gurstein's (2012) article on the use of data for addressing problems with public toilets in India, an often-cited example of the positive potential social impact of open data, and Hayakawa, Imi, and Ito's (2014) work on the OpenStreetMap community in Japan following the Tohoku earthquake, invite critical review from the records management community. For instance, what is the impact on citizens and data users when data is not drawn from records? How can data from records work with or against crowdsourced data to produce an accurate picture of events or circumstances?

Moving from the subjects of research to the approaches to research: A number of national case studies of open government data exist, including case studies for Albania (Hoxha, Brahaj, and Vrandečić 2011), Brazil (Breitman et al. 2012, Matheus, Ribeiro, and Vaz 2012, Albano and Reinhard 2014, Brito et al. 2014, Corrêa, and da Silva 2014), Germany (“BMI: Studie „Open Government Data Deutschland“” 2012), Greece (Alexopoulos, Spiliotopoulou, and Charalabidis 2013, Galiotou and Fragkou 2013, Theocharis and Tsihirintzis 2013), Italy (Viscusi et al. 2014), Latvia (Bojārs and Liepiņš 2014), Mexico (González et al. 2014), Saudi Arabia (Al-Khalifa 2013), and Taiwan (Yang et al. 2013). International, comparative studies are far less common (Murillo (2015) has looked at online data availability in sixteen Latin American countries). These two approaches – national and comparative / international studies – present two parameters that could be used in examining records issues in relation to open government data in various jurisdictions. These case studies also suggest various scopes of study that might be useful in research in the records and archives field; some of the case studies are broad overviews of national ‘data landscapes’ (i.e. Bojārs and Liepiņš 2014) while others are quite specific, looking at issues as particular as sources of data (i.e. Alexopoulos, Spiliotopoulou, and Charalabidis 2013).

Automatically”. It would be worthwhile to analyse these models and guides more closely to see to what extent they recognise or incorporate records issues.


This literature review has presented a great number of avenues for further research for the records / archives community, but the overarching characteristic of all of these is the need for collaboration and engagement between disciplines and fields of practice, since there are significant gaps between communities that could be closed to the benefit of open data projects, the quality of the data being released, and therefore, ultimately, the users of the information. Many new fields of enquiry are opened up to us, but many will require us to work collaboratively with people outside our discipline. It is not so much about starting new research as seeking new partnerships to work with in closing the gaps.

References


