Being Open About Data

Analysis of the UK open data policies and
applicability of open data

ANTTI HALONEN

REPORTS

The Finnish Institute is a London-based private trust. Our mission is to identify emerging issues relevant to contemporary society and to act as catalyst for positive social change through partnerships.

We work with artists, researchers, experts and policy makers in the United Kingdom, Finland and the Republic of Ireland to promote strong networks in the fields of culture and society. We encourage new and unexpected collaborations and support artistic interventions, research, the creative industries, foresight and social innovation in new, socially central areas.

The Reports of the Finnish Institute in London is a series of publications, which publishes research, studies and results of collaborative projects carried out by the institute. The reports provide evidence and ideas for policy-makers and civic society organisations dealing with contemporary social and cultural challenges.

35–36 Eagle Street
London WC1R 4AQ
United Kingdom

T +44(0)20 7404 3309
www.finnish-institute.org.uk

Table of contents

Executive Summary

This paper presents an analysis of the recent UK open-data policies and draws an argument on how governments can sustainably promote the development and use of open data. Moreover, research contributes to the ongoing discussion on the normative values of openness by presenting a conceptual analysis of open data as an integral part of the freedom-of-information continuum.

The key findings can be listed as follows:

- Key to benefits is the quality of user engagement
- Open data and its objectives should be addressed as a part of the freedom-of-information continuum
- The decision to emphasise the release of expenditure data was not ideal: governments do not know best what kind of data people want to have and should aim at releasing it all
- Leadership, trust and IT knowledge are crucial, not only political leadership but within organisations too
- The social and democratic impacts of open data are still unclear and in future there is a need for sector-specific research

Research was conducted from September 2011 to January 2012, mostly by semi-structured interviews of key experts and analysis of the government policies. Starting point for the research was to primarily address the applicability of open data: how data is being used and what kind of benefits is it possible to identify from the data use. During the research process it became evident, however, that open data as a concept is so diverse that a mere analysis of data use would be insufficient in order to gain a deeper understanding of the phenomenon.

Open data is applied in various ways with lots of small-scale success stories available, mostly in the form of mobile-phone or web applications. These apps and websites – as innovative and useful as they are – are yet not the key issue when addressing the overall value of open data. These services make everyday life of citizens a tiny bit easier, and when accumulated they may result in significant economic benefits. However, the open-data community has also been vocal about the potential positive impacts on democracy. These impacts are significantly harder to identify and need

much more research in order to produce comprehensive and reliable results. In addition, we must realise the difference between transparency and democracy-oriented goals that are usually associated with the freedom-of-information movement and the technology and innovation-oriented goals of the open-data movement.

Key to further benefits, whether economic or democratic, is more education and improved user engagement – of both citizens and public officials. The level of knowledge and understanding of open data is currently rather low, and most data producers don't yet see the potential benefits that lie in open data. Equally, individual citizens are not necessarily capable of using datasets as the threshold for accessing and using raw datasets effectively is, at times, quite high. The best examples, in fact, are those where the data-portal interfaces are made as simple and easy to use as possible by providing relevant context to data and equally where data users are already engaged in public participation, be it within the public sector itself or some organised civic-society movement. Consequently, it is only the data user herself who knows what kind of data would be most useful. A certain service-design approach would be desirable.

There are already examples available where companies have benefited commercially from data and where public-sector organisations have gained efficiency benefits. In the future, it is more important to focus on the normative side of open data and on its potential impacts on democracy. There is a risk of creating a hollow mantra of open data improving the level of democracy without any evidence provided. However, the potential for great improvement in democratic accountability is there, and in some cases there is already sufficient evidence.

All these benefits require the threshold for accessing, understanding and using the data to be as low as possible. In order to achieve this, the data producers must possess a certain level of ICT knowledge to implement the system so that it is both simple enough to use and sophisticated enough to be able to manage information flow comprehensively. In many cases, the ICT and data-management infrastructure is not sufficient, and organisations lack the human resources to renew it so that it matches the requirements of openness. This should not be an excuse for not to release data, however, but a wake-up call for both data providers and open-data community alike.

Finally, it should be made clear that open data is not apolitical initiative. There is a strong political side to it, which dates all the way back to the long development process of governmental transparency in the UK. The initial focus on the release of expenditure data is claimed to be driven by political motives, and in terms of development of sustainable and productive use of data it was not necessarily the right decision to make. The discussion of open data was sidetracked when the focus was on the rift between local and central government and not on how public-sector organisations and civic communities could benefit from data. For many local-government data producers, the whole open-data initiative is equivalent to the £500 expenditure-data agenda and hence they don't necessarily see the wider context and potential benefits that might lie in open data. Bearing this in mind, the open-data community should be wary of arguing too eagerly in favour of open data improving the general access to information. Open data at its current form is mostly a target-driven policy without the reactive pull-factor that is essential for a political right that is freedom of information.

2. Open data – What and why?

2.1. What is open data?

The term "open data" usually refers to non-personal data that is accessible to all and can be freely used, re-used and distributed by anyone. Re-use of data is made possible by releasing data in machine-readable format and under such a licence that typically allows both commercial and non-commercial use. Typical examples are datasets that are created by public authorities, but open-data principles may be applied to private datasets, as well.

There are several widely used definitions for open data. The Open Knowledge Foundation, an influential British NGO advocating governmental transparency in the digital age, sums up the definition of open data as follows:

> "Open data is data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and share-alike."[17]

The UK government Public Sector Transparency Board, however, has given the following definition for public data:

> "Public data is the objective, factual, non-personal data on which public services run and are assessed, and on which policy decisions are based, or which is collected or generated in the course of public service delivery."[18]

Despite these definitions, conceptually the open-data movement derives from a wider epistemological background. At minimum, three different aspects of openness can be listed, as follows:

1) Technological openness
2) Non-proprietary openness
3) Legal openness

---

17 See http://opendefinition.org
18 New Public Sector Transparency Board and Public Data Transparency Principles,
http://data.gov.uk/blog/new-public-sector-transparency-board-and-public-data-transparency-principles

Firstly, in technical terms, open-data discussion is typically taking place around concepts of machine-readability, the semantic web and linked open data, which further emphasises the connectivity of data and possibilities to create so-called smart applications. The essential issue in terms of technological openness is thus the question of machine-readability.

The basic principles of linked open data are typically listed in a five-star model as follows:

* Data is available on the web (in whatever format), but with an open licence
** Data is available as machine-readable structured data (e.g. in Excel, instead of an image scan of a table)
*** As in two stars plus non-proprietary format (e.g. CSV instead of Excel)
**** All the above plus use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at things created by others
***** All the above, plus linking your data to other people's data to provide context [19]

Non-proprietary openness, or openness as an ideology, derives from the demand for interoperability and inclusivity, which also draws from the creative value of sharing and utilising common resources. Eric Raymond argues in his influential essay *Cathedral and bazaar* that the creative process of hackers and software developers mainly derives from the combination of gaining a personal reputation by creating "cool" new features and the virtually seamless cooperation that takes place between developers.[20] An intimate connection between the open-data movement and the earlier open-access movement is clearly visible.

In terms of government data, the argument goes that data that is created by public funds should be free for all to use and re-use. Data is thus seen as a common resource that does not suffer from scarcity. It is seen as morally wrong to restrict the use of data to just a certain group of people. The international "right to data" initiative derives from this setup and can be seen as a tail of the long-standing freedom-of-information movement.

---

19 Five-star model adapted from http://lab.linkeddata.deri.ie/2010/star-scheme-by-example/
20 Raymond, Eric: *The Cathedral and the Bazaar*, 2000, http://catb.org/~esr/writings/homesteading/cathedral-bazaar/cathedral-bazaar.ps

The idea of legal openness is linked to the licensing of data. In Chapter 4, I will put forward an argument on the necessity of the institutional basis of open data, in which licensing is a major determinant. In short, data must be licensed under such a licence that recognises the user's right to exploit data in a variety of ways, including commercially.

All these aspects of openness are effectively compiled in an eight-point list, which has been adopted by open-data advocates and governments alike[21]:

1. Data must be **complete**: all public data is made available. Public data is data that is not subject to valid privacy, security or privilege limitations.
2. Data must be **primary**: data is collected at source, with the finest possible level of granularity, not in aggregate or modified norms.
3. Data must be **timely**: data is made available as quickly as possible to preserve the value of the data.
4. Data must be **accessible**: data is available to the widest range of users for the widest range of purposes.
5. Data must be **machine-processable**: data is reasonably structured to allow automated processing.
6. Access must be **non-discriminatory**: data is available to anyone, with no requirement of registration.
7. Data formats must be **non-proprietary**: data is available in a format over which no entity has exclusive control.
8. Data must be **licence-free**: data is not subject to any copyright, patent, trademark or trade-secret regulation. Reasonable privacy, security and privilege restrictions may be allowed.

Paul Clarke has identified four broad types of datasets that are published by public bodies[22]:

21 Open Government Data: eight principles of open government data, http://www.opengovdata.org/home/8principles; explaining remarks adapted from Schellong, Alexander & Stepanets, Ekaterina: *Unchartered waters – the State of Open Data in Europe*, CSC Public Sector Study Series 1/2011, http://assets1.csc.com/de/downloads/CSC_policy_paper_series_01_2011_unchartered_waters_state_of_open_data_europe_English_2.pdf
22 Clarke, Paul: *There's data, and there's data*, http://paulclarke.com/honestlyreal/2010/06/theres-data-and-theres-data/

1. **Historical data**: information on what has happened in the past
2. **Planning data**: what is projected to happen
3. **Infrastructural data**: static information on society: postcodes, opening hours, organisation structures, etc.
4. **Operational data**: public-transport timetables and other real-time information

These four types of definition are useful in terms of analysing the applicability of data – for whom different types of data are most useful and why. It also helps in further categorisation of datasets. One of my main arguments is the importance of conducting sector-specific research, for which these types of definitions would probably be useful.

There is an ever-increasing list of applications and web services that are created by applying open data: from live tube timetables to services that indicate the level of crime in the area where a user is walking. These apps make the user's everyday life more convenient and are increasingly important in their financial value. More information on different apps and services can easily be found online, and so this report does not concentrate on these examples.

2.2. Why open data?

Open data is intrinsically a combination of various different things and thus is associated with different objectives and benefits for different groups of people. There is not any single pattern of goals but various interrelated application areas, which together form a compilation of objectives and potential benefits.

Moreover, as Rufus Pollock, for instance, argues, open data is primarily a means to an end, not an end in itself.[23] The objectives of open data are thus related to goals set – be it in the field of economics, democracy or public services – but data openness *per se* does not necessarily have significant value. The question of the intrinsic value of openness is a substantial philosophical issue, and it will not be tackled here in detail.

At the moment, it is difficult to predict which particular application area will rise into an outstanding position in the future. It seems that new potential application areas are popping up rapidly. Nevertheless, it is still very early days, and it looks like the research so far has been mostly a combination of evidence and expectations, as Jonathan Gray, for instance, has pointed out.[24] Moreover, history shows that predictions on the applicability of various social and technological innovations have often been rather inaccurate.

Research literature, however, suggests that there are a few emerging areas where open data could potentially be most beneficial. These objectives are expected to release both internal and external benefits.

Internal objectives are those that would enhance an organisation's own work in some particular way or form. These benefits include better efficiency, increased internal understanding of the organisation's work and objectives and releasing of resources into more productive tasks.

On the other hand, the open-data movement is also expected to provide value for society at large. These *external* objectives arise mostly from supporting entrepreneurship, empowering citizens, and democratic accountability and participation. In short, open data is seen as a tool for a more democratic society, an improved economy and empowered citizens.

The joint feature of both sets of objectives is the perception that open data has provided opportunities for the general public to use public information in a way that was not possible before.[25] In comparison to reactive freedom of information, the open-data movement – or two-way online transparency, as it is sometimes called – is seen as enabling a participative *writing society* instead of a *reading society*, where citizens are theoretically able to receive information but not to re-use it in creative ways.

---

24 Gray, Jonathan: http://jwyg.okfn.org/2011/12/06/interview-for-university-of-southhampton-open-data-study/#comments
25 See, for instance: Davies, Alysia & Lithwick, Dara: *Government 2.0 and Access to Information: Recent Developments in Proactive Disclosure and Open Data in the United States and Other Countries*, Library of Parliament of Canada, 15 April 2010, http://www2.parl.gc.ca/Content/LOP/ResearchPublications/2010-15-e.pdf

In the UK, the central government has particularly emphasised two key areas that open data could benefit: accountability and entrepreneurship.[26] In addition, there are reports that suggest wide-scale benefits that open data could provide in public service provision.[27] These objectives play a key role in the recent Cabinet Office white paper on the future of public services.[28] More on the current government discourse is analysed in Chapters 3 and 4.

The emphasis of the UK central government is in accordance with the findings of Noor Huijboom and Tijs Van den Broek, who have compared open-data policies of five different countries where open-data strategies have already been implemented: the United Kingdom, United States, Spain, Denmark and Australia. They conclude that the following three primary motivations can be pointed out for publishing government data (Figure 1):

1) Increasing democratic control and participation
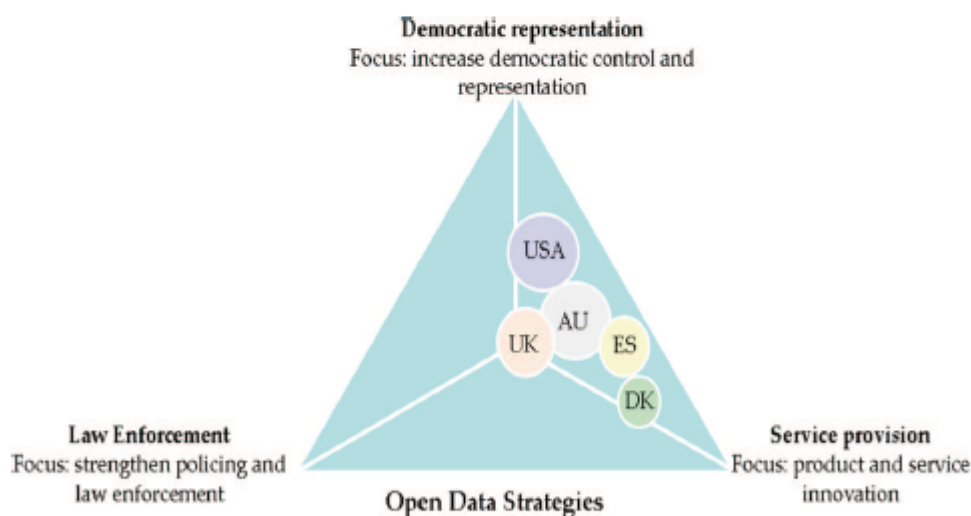2) Fostering service and product innovation
3) Strengthening law enforcement[29]

Figure 1: overview of open data strategies (Huijboom & Van den Broek, p.5)

---

26 See, for instance: Transparency Board Minutes, 11 January 2011, http://data.gov.uk/blog/transparency-board-minutes-11th-january-2011-0
27 2020 Public Services Trust: *Online or In-line: The future of information and communication technology in public services*, March 2010, http://clients.squareeye.net/uploads/2020/documents/online_or_inline.pdf
28 Cabinet Office White Paper: *Open Public Services*, July 2011,
http://www.cabinetoffice.gov.uk/sites/default/files/resources/open-public-services-white-paper.pdf
29 Huijboom, Noor & Van den Broek, Tijs: *Open data: an international comparison of strategies*, European Journal of ePractice, No. 12, March/April 2011,
http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf

According to Huijboom and Van den Broek, all five countries clearly prioritise either the participative or service-innovation aspect.[30]

The next few pages are dedicated to identifying the core objectives that can be singled out from the research literature and open-data discussion.

2.2.1 Efficiency

A common expectation throughout the open-data community is that by opening up their datasets to public scrutiny, organisations can expect to gain significant efficiency savings.[31] The reasoning behind this argument derives from the idea that through scrutiny, enabled by transparency, it is easier for vigilant citizens to identify wasteful behaviour in public-sector organisations. Another argument states that by opening up expenditure information, an organisation's employees themselves gain better understanding of the organisation's financial situation and can thus act accordingly.

The UK government has relished the emergence of "armchair auditors", that is a group of interested citizens who can easily hold public officials to account by scrutinising datasets of public spending. Community Secretary Eric Pickles (Con) has stated the following:

> "The public should be able to see where their money goes and what it delivers. The swift and simple changes we are calling for today will unleash an army of armchair auditors and quite rightly make those charged with doling out the pennies stop and think twice about whether they are getting value for money." [32]

There is already some research to back up the expectations of improved efficiency. Publishing of expenditure data in the state of California, for example, has reportedly

30 Ibid.
31 See e.g. PM David Cameron's letter to Government departments, 31 May 2010, http://www.number10.gov.uk/news/letter-to-government-departments-on-opening-up-data/
32 Secretary of Communities and Local Government Eric Pickles: New era of transparency will bring about a revolution in town hall openness and accountability, 4 June 2010, http://www.communities.gov.uk/news/corporate/1606882

led to efficiency savings of $20m.[33] In the UK, there are reports of data transparency enhancing the performance of those NHS organisations that have opened up their data to public scrutiny.[34] In Canada, open data is claimed to have helped in revealing one of the biggest tax frauds in the country's history.[35]

Another line in the efficiency argument is related to the possible savings that could be made by managing freedom-of-information requests more efficiently. According to David Eaves, the costs of complying with freedom-of-information requests in Canada alone have increased rapidly during the last few years, and the number is constantly growing.[36] The number of requests has, in fact, been growing throughout the English-speaking post-industrialised world, and the average cost per request is estimated to vary between £200 and £800, depending on the country.[37] With proactive publishing of open data, however, citizens are expected to find the information they want without the need to submit a freedom-of-information request.

2.2.2 Democratic accountability

In a wider sense, I regard democratic accountability as including both transparency *per se* and government accountability for using public funds and making political decisions as representatives of the wider demos. However, later in this report, I have separated transparency and accountability when addressing the benefits achieved.

Literature suggests that there are varying views on whether technologically more sophisticated ICT-environment and e-government initiatives have led to a more transparent government. Cory Armstrong has linked the development of e-government and online transparency in the USA to the level of professionalism and ease of information accessibility on websites. Not only is the type of information placed online important in terms of determining the level of transparency, but also where and how it is placed online. She also argues that the level of online

---

33 UN e-Government Survey 2010, http://unpan1.un.org/intradoc/groups/public/documents/UN-DPADM/UNPAN038853.pdf
34 See, for instance, http://www.ucl.ac.uk/constitution-unit/events/tim-kelsey-presentation
35 Eaves, David: Case study: *How open data saved Canada $3.2 billion*, http://eaves.ca/2010/04/14/case-study-open-data-and-the-public-purse/
36 Eaves, David: *Access to Information is Fatally Broken... You Just Don't Know it Yet*, http://eaves.ca/2011/03/30/access-to-information-is-fatally-broken-you-just-dont-know-it-yet/
37 Colquhoun, Anna: *The Cost of Freedom of Information*, The Constitution Unit, December 2010, http://www.ucl.ac.uk/constitution-unit/research/foi/countries/cost-of-foi.pdf

transparency was greater in communities with a higher proportion of Republicans.[38] In terms of the political discourse that takes place around transparency and other political implications, it might be interesting to assess whether similar trends occur in a UK context, but unfortunately it is outside the scope of this study.

Armstrong's findings that professionalism and accessibility enhance online transparency are countered by Vicente Pina, Lourdes Torres and Sonia Royo, who argue that at the level of the European Union, ICT initiatives have not yet had any dramatic impacts on government transparency.[39] In a sense, Pina et al are members of the same continuum as Darrell West, who initially argued that e-government has not met its main objectives of transforming service delivery and public trust in government[40], and perhaps even Frank Bannister and Regina Connolly, whose conclusion is that expectations of e-government-powered transparency might have been too high so far.[41]

I argue that open-data initiatives are intrinsically related to this wider context of e-government and online transparency, and therefore the basis for this research has consisted of a literature review of relevant studies on e-government. In Chapter 4, I will consider further the historical and conceptual development of open data and transparency.

The accountability arguments usually state that while accountability is one of the most important aspects of democracy, it is very difficult to have it without true openness in government.[42] Open data is seen as a means for maximising citizens' potential to scrutinise government and spot wasteful spending, hence improving accountability.[43]

---

38 Armstrong, Cory L.: *Providing a clearer view: An examination of transparency on local government websites*, Government Information Quarterly, Volume 28, Issue 1, January 2011, pp. 11-16.

39 Pina, Vicente; Torres, Lourdes; Royo, Sonia : *Is e-government leading to more accountable and transparent local governments?* An overall view, Financial Accountability & Management, Volume 26, Issue 1, February 2010, pp. 3-20.

40 West, Darrell M.: *E-Government and the Transformation of Service Delivery and Citizen Attitudes*, Public Administration Review, Volume 64, Issue 1, February 2004, pp.15-27.

41 Bannister, Frank & Connolly, Regina: *The Trouble with Transparency: A Critical Review of Openness in e-Government*, Policy & Internet, Volume 3, Issue 1, Article 8, 2011,
http://www.psocommons.org/policyandinternet/vol3/iss1/art8/

42 See, for example, Chapman, Richard; Hunt, Michael: *Open Government in a Theoretical and Practical Context*, Ashgate Publishing, Aldershot, 2006, p. 140.

43 See, for example, Pickles, 2010.

Justin Longo argues that accountability has, in fact, been the leading objective of the UK open-data initiative and that this has its roots in the expenses' scandal and the use of crowdsourcing to "unearth previously hidden examples of misappropriation of funds".[44] Moreover, accountability is hoped to increase public trust in government, which has been one of the key objectives of reactive freedom-of-information laws.[45]

2.2.3 Empowering and public participation

Enhancing public participation has been one of the most widely used objectives of both freedom of information (FOI) and e-Government initiatives, although it has not necessarily been completely achieved.[46] The share of citizens of the total population who have made FOI requests so far has been relatively low in the UK, and any significant connection between FOI and increasing participation has been extremely difficult to prove.[47] In terms of e-Government, the connection is not much clearer, with several research findings indicating conflicting results.[48] Raymond La Raja, for one, concludes that increased transparency has resulted in a decreasing level of participation, since people are not as willing to sign petitions if they believe that the petition and the list of people who have signed it will appear online.[49]

However, open data, with its more interactive nature, is hoped to enable the creative re-use of information, which would enhance the opportunities for meaningful public participation. Governments who open up public data hope that enthusiastic citizens will create innovative platforms and services for public use. It is hoped that the threshold for joining the public discussion will be effectively demolished.

In the UK, citizens are encouraged to join in establishing public services and, in this development, it is argued that access to public data is crucial.[50] This kind of

---

44 Longo, Justin: *#OpenData: Digital-Era Governance Thoroughbred or New Public Management Trojan Horse?*, Public Policy and Governance Review, Volume 2, No. 2, Spring 2011, pp. 38-52.
45 See, for example, Hazell, Robert; Worthy, Ben; Glover, Mark: *The Impact of the Freedom of Information Act on Central Government in the UK – Does FOI work?*, Palgrave MacMillan, 2010
46 Hazell et al., 2010.
47 Ibid.
48 Andersen, KN; Henriksen, H; Medaglia, R; Danziger, JN; Sannarnes, MK; Møyfrid K; and, Enemærke, M. 2010. *'Fads and Facts of E-Government: A Review of Impacts of E-Government (2003-2009)'*. International Journal of Public Administration Vol. 33 (11) 564-579
49 La Raja, Raymond: *Does Transparency of Political Activity Have a Chilling Effect on Participation?* Meetings of the Midwest Political Science Association, Chicago, March 31 – April 3, 2011.
50 Cabinet Office: Open Public Services White Paper, July 2011,
http://www.cabinetoffice.gov.uk/sites/default/files/resources/open-public-services-white-paper.pdf

empowering is also claimed to increase public trust and help citizens to value and understand public services better when they are involved in the actual provision phase.[51]

2.2.4 Economy and entrepreneurship

Rufus Pollock has estimated that the opening up of government datasets would create an economy with an annual worth exceeding £6bn in the UK alone.[52] Other calculations estimate the value of open data on a global level to be up to £20bn annually[53] or even €28bn in the EU27 market, with rapid growth. [54]

These calculations are based mainly on the expectations of a network of flourishing enterprises, which could exploit datasets commercially by creating new products and services built on public sector information. Also, enterprises and public-sector organisations are expected to reduce costs in accessing and using public information.[55]

The entrepreneurial approach to open data has been particularly popular among the current UK government, whose Chancellor, George Osborne, has described the movement as an opportunity to "maximise our business opportunities at hand". It is a logical part of the continuum, where the possible market potential of public-sector information has been emphasised since the 1990's.[56]

In terms of data use for entrepreneurial purposes, it is argued that crucial issues are charging, licensing and data format. In order to exploit data for entrepreneurial benefits, a licence, for example, must allow the use of data for commercial purposes. In the case of UK government data, this is mostly the case. Open licences are considered further in Chapter 4.

---

51 Ibid.
[52] Pollock, Rufus: *The Economics of Public Sector Information*, University of Cambridge, 2008, http://econpapers.repec.org/paper/camcamdae/0920.htm
[53] Speech by the Chancellor of the Exchequer George Osborne at Google Zeitgeist 16 May 2011, http://www.hm-treasury.gov.uk/press_48_11.htm
54 Vickery, 2011
55 ibid.
56 More in Chapter 3.

According to Pollock, the benefits of open data derive mainly from the marginal-cost release of data.[57] Marginal-cost release enables the small and medium-sized enterprises to exploit data, which they couldn't necessarily do if there were charges associated with the data. The loss of sales profit would then be replaced with government subsidies. Moreover, Koski has argued that the charging regime could smother small and medium sized enterprises and give a disproportionate edge to bigger companies that could afford to purchase data regardless of the fees involved. [58]

Despite all the aforementioned areas of society where open data could potentially prove beneficial, arguably the most intriguing aspect of data is the fact that no-one really knows all the possible applications for the data. There would be a myriad of opportunities if all relevant data were in an appropriate open and linked format and released under open licence. The best we can really do is to expect the unexpected. Some of the different application areas are presented in this paper as case studies, in order to illustrate the diversity of data-powered civic action.