

**RCUK Digital Economy Theme
NEMODE – 3K Small Grants Call 2013**

New business models and the value of open data: definitions, challenges and opportunities

Dr Carla M Bonina

Research Fellow

London School of Economics and Political Science

Department of Management

Houghton Street, WC2A 2AE, London, UK

Email: c.m.bonina@lse.ac.uk

<http://blogs.lse.ac.uk/nef/>

**For submission to completion of the Small Grants Scheme 2013.

Please do not distribute without author's permission**

First version: 30 August 2013. This (revised) version: 27 September 2013

This work is licensed under the Creative Commons Attribution 3.0 Unported License.

To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>.



Contents

1	Introduction	3
2	Definitions: open data, business models and value	5
2.1	Open data	5
2.2	Open data landscape: users, actors, practices.....	6
2.3	Business models and open data.....	9
2.4	Value and ethics in open data.....	10
3	Economic value of open data: business models perspective.....	12
3.1	Value capture and open data	12
3.2	Canvas model and open data	14
3.3	Empirical studies of open data.....	16
4	Value and ethics in open data.....	16
4.1	Industry, productivity, creativity	17
4.2	Accountability, transparency, autonomy.....	17
4.3	Voluntarism, sociability, cooperation, civic engagement	18
5	Looking ahead: challenges, issues and trade-offs.....	19
5.1	Personal data and open data: privacy and politics.....	19
5.2	Data divide: Effective use of data and equality issues	20
5.3	Costs of collecting, producing and releasing open (government) data	22
5.4	Coordinating a diversity of interests and changes in behaviour.....	23
6	Concluding remarks: the future of open data.....	23
	Acknowledgments.....	25
	References	26
	Further resources in open data (UK and Europe).....	29

1 Introduction

There is increasing excitement about the potential economic and social benefits of using newly released public and private data in the digital economy. “Big data” and “open data” have received notable attention in the UK from different fronts: those anxious about new economic opportunities emerging from data-driven businesses, those inspired by the social benefits of transparency and public service monitoring, and those concerned with issues of about data ownership and privacy. In this project, I focus on *open data*—that is, data released in open access format, generally coming from the government but also other public and non-public bodies.

Open data is making its way into the UK and also globally. One of the key reasons why open data has been in the UK and European public discourse relies on its acclaimed potential to generate economic growth (European Commission 2011, UK Cabinet Office 2012). A recent study commissioned by the European Commission estimates that the economic value of opening up and re-using public sector information are about €40 billion per year in the EU alone (Vickery 2011). Similarly, in the UK Deloitte estimates that the direct economic benefits are around £1.8bn per year in the UK. As a phenomenon, open data is quite recent—as is the open access movement and the new technologies enabling many of these new developments based on open data to happen. New products and companies emerge as businesses, governments and end users use or re-use open data, and we still know little about its potential effects. The Open Knowledge Foundation, a non-profit organization that promotes open knowledge, suggests a plethora of new possibilities with potential for value generation (Open Knowledge Foundation 2010):

While there are instances of the ways in which open data is already creating both social and economic value, we don’t yet know what new things will become possible. New combinations of data can create new knowledge and insights, which can lead to whole new fields of application.

Despite the economic estimates and predictions, it is far easier to measure the costs of producing and releasing open data than its economic and social impacts. Models for evaluating wider impacts

of open data need to be developed, as it was noted in the independent review of public sector information commissioned by the UK government in 2013 (Shakespeare 2013).

The purpose of this paper is to understand what new possibilities emerge in the open data landscape, and to explore its implications and challenges for the UK economy and society. In particular, I seek to address three main conceptual questions:

- What types of business models can be identified in open data?
- What constellations of value can be mapped?
- What are the important trade-offs, challenges and consequences emerging from open data for business and society?

To answer these questions, I analyse extant research based on extensive collection of materials in three areas: recent developments on open data (including example of cases), literature on business models, and research on value and ethical theory in the social sciences. I also draw on anecdotal material from informal interviews, workshops and seminars that I have been involved with in recent months that are relevant for the discussion. The ultimate objective is to build a conceptual base that can guide future work on the area.

This report is structured as follows. In section two I present the main definitions and concepts that I use in the report. That is: open data, business models and value. Section three reviews current advances on archetypes of business models proposed in the literature. Section four advances other clusters of value that emerge as relevant when considering open data and its impacts for society. Section five presents some important challenges and issues that we need to take into consideration from a policy perspective. In the last section, I present concluding remarks and avenues for future research on open data. An appendix at the end of the report compiles a list of open data materials relevant for the UK, such as datasets, open data tools and case studies for future reference.

2 Definitions: open data, business models and value

2.1 Open data

There is notably a lack of clarity about key terms used in the literature and in public debates. The origins of open data can be traced to the open source software community, although the connections between both communities of research have not materialised in great extent yet (Willinsky 2005, Lindman *et al.* 2013). For the purposes of this project, I draw upon two main sources to build a coherent definition. The first one addresses the concept of openness, which defines open data¹ as follows:

A piece of content or data is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike. (<http://opendefinition.org/>)

The second one complements and expands on it, and it is based on recent work commissioned by the Royal Society and published in the report entitled “*Science as an open enterprise*”. In order to be “open data”, data must be *accessible, assessable, intelligible* and *useable*. The following table presents these concepts in detail.

Open data features	Definition
accessible	Data must be located in such a manner that it can readily be found and in a form that can be used.
assessable	In a state in which judgments can be made as to the data or information’s reliability. Data must provide an account that is intelligible to those wishing to understand or scrutinise them.
intelligible	Comprehensive for those who wish to scrutinise something. Audiences need to be able to make some judgment or assessment of what is communicated.
useable	In a format where others can use the data or information. Data should be able to be reused, often for different purposes, and therefore will require proper background information and metadata. The usability of data will also depend on those who wish to use them.

Table 1: Intelligent openness term, adapted from the Royal Society (2012, p. 12).

¹ As per data itself, I follow the definition provided in the Royal Society report (2012, p. 12), that is: “qualitative or quantitative statements or numbers that are (or assumed to be) factual. Data may be raw or primary data (i.e. direct from measurement), or derivative of primary data, but are not yet the product of analysis or interpretation other than calculation”.

Within the open data community, in practice, the general understanding is that a piece of data is open if it is accessible without limitations on the user's identity or intent, is provided in digital, machine readable format capable of being linked with other data; and is provided free of restriction on use, re-use or redistribution according to its actual licensing conditions (UK Cabinet Office 2012, The Open Data Institute 2013a).

Open data can therefore be small, medium or big data—the latter usually defined as very large datasets that require massive computing power to be processed (i.e. terabytes of sensor data or billions of Wikipedia entries). Open data can also come from government data, but open data does not restrict its scope to that. Open data can also be from private firms—and that is one of the core missions of the newly created Open Data Institute². However, in terms of scope, the focus of this report will mainly draw upon open government data given its relative importance for the open data community, since the government is the largest collector of data regarding aspects of economic and social life.

Open government data is also different from public sector information. The former is simply a subset of the latter—that is, public sector information that the government releases following the principles of openness described above. There are reasons why not all government data shall be open, such as those commonly expressed as exclusions in the Freedom of Information Acts (i.e. privacy, sensibility, third party ownership).

2.2 Open data landscape: users, actors, practices

As briefly suggested in the introduction, the open data landscape includes different actors/users with diverse roles and opportunities. For the government, open data offers enormous potential to improve public service delivery, transparency and efficiency of internal operations. Open data brings promising opportunities to generate innovation and economic benefits for a wide range of actors in the economy: established companies that may find new revenues from providing analysed data,

² The Open Data Institute is an independent, non-profit organisation founded in 2012, whose mission is to catalyse an open data culture in diverse spheres of economic and social activities.

individual developers and start-ups producing software applications for mobile and internet-based devices, customers benefiting from better services, journalists, educators and researchers that may use open data on their work.

In the private sector, open data is not just a matter of small and entrepreneurial firms. A recent survey conducted to 318 C-level executives by the Economist Intelligent Unit (2013) reveals that:

- 70% of the firms analysed collect open data (from government datasets but also from other sources);
- 54% of high-growth firms *collect* and *analyse* open data; in the case of low-growth firms, the percentage is 36%;
- Open data is the second most important type of data that executives see as the most valuable for deciding on changes of strategy or developing new strategies (48%); the first one are syndicated data from third-party data providers (i.e. market data, weather, etc.) (52%)—see the figure below.

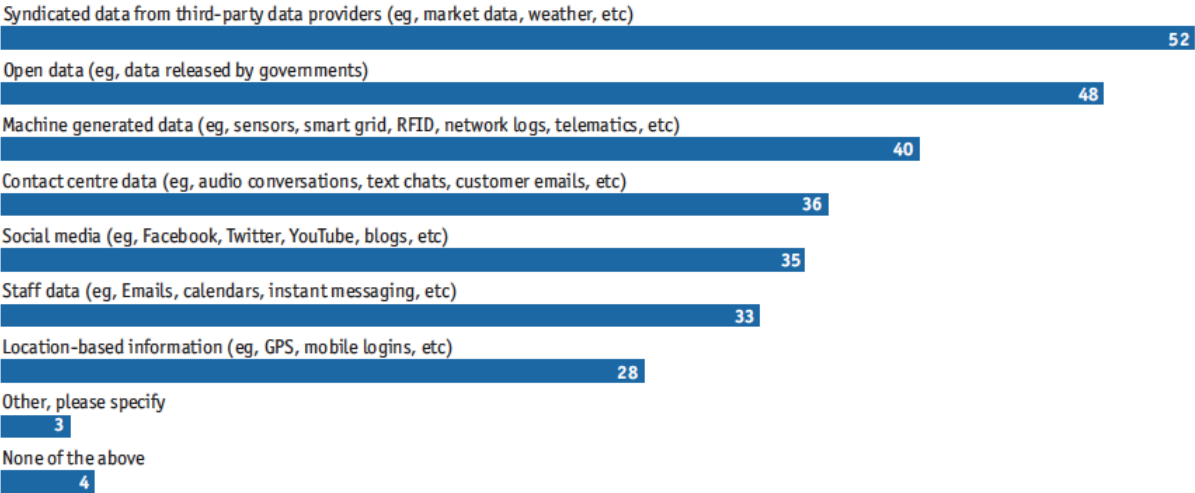


Figure 1: % respondents to the question: “Which of the following types of data would your company benefit from the most in deciding on changes of strategy or developing new strategies?” Source: The Economist Intelligent Unit (2013), p. 41.

Diverse stakeholders interact in the open data landscape. As in the cases of open innovation (Chesbrough 2006, Boudreau and Lakhani 2013) and open source software development (Fitzgerald 2006), collaboration and co-creation are important mechanisms to create, develop and sustain a diverse range of services. For example, a common one is crowdsourcing—that is, using competitions such as *hackathons* and award-driven challenges to create new solutions and products.³ Local and central governments have also invested in the co-creation of services (Agranoff and McGuire 2003, Agranoff 2012). These are practices that local governments use by bringing in users into participatory forums (i.e. London Data Store, or the Open Data User Group as part of data.gov.uk). In sum, the open data landscape entails several stakeholders with different opportunities for value creation and appropriation.

Example: The London Datastore and co-creation of services

The Greater London Authority (GLA) has been a pioneer in releasing data in open format to generate new services and software applications for citizens. A milestone was the creation of the London Datastore in January 2010, whose main goal is stated below:

“[we see the Datastore] as an innovation towards freeing London’s data. We want citizens to be able [to] access the data that the GLA and other public sector organisations hold, and to use that data however they see fit – free of charge. The GLA is committed to influencing and cajoling other public sector organisations into releasing their data here too. Releasing data though is just half the battle. Raw data often doesn’t tell you anything until it has been presented in a meaningful way. We want to encourage the masses of technical talent that we have in London to transform rows of text and numbers into apps, websites or mobile products which people can actually find useful.”⁴

In simple terms, the London Datastore offers a repository of public data that the GLA and other public sector information that diverse government agencies hold about the city. The data are in open format—that is, machine readable, accessible to everyone, and provided for free. In practice, there are big disparities in terms of data reliability, transparency and accessibility. For example, while some data are vastly used (i.e. real-time data from transport for London), other datasets are not transparent or accessible (i.e. public procurement opportunities with GLA and contracts awarded by

³ See Boudreau and Lakhani (2013) for examples on crowdsourcing and crowd contests to foster innovation.

⁴ Source: Greater London Authority (2013). Available at <http://data.london.gov.uk/>

GLA). As per the practices described earlier, common examples of co-creation and third party developments include the mobile transport apps that flourished in London soon after the GLA released Transport for London's real-time transportation and commuting data in 2010.

Despite these examples and the successful cases in transportation, the question on how open data may help generating value has been at core of recent public debates (i.e. UK Cabinet Office 2012). The notion of business models can help to elucidate the extent to which releasing open data will allow different types of companies to build new products and services that can feed back into the economy. I shall explore these issues in the next sub-section. But the value of open data may be found in broader social, democratic or even environmental set of values, as I shall discuss afterwards.

2.3 Business models and open data

The literature on business models offers a starting point to address how open data can create and capture value, although it is fragmented and has led to different understandings. Scholars have referred to business models as representations, as conceptual tools, as a method or even as a pattern (Amit and Zott 2001, Osterwalder *et al.* 2005, Shafer *et al.* 2005, Baden-Fuller and Morgan 2010, Teece 2010, Zott *et al.* 2011). In addition, traditional models based on value chain may not necessarily apply or be relevant in the new digital economy, where services, data and providers are mixed in the picture. For example, as explained before, open data may be more than government data—the private sector is a big generator of data itself. Still, many organisations that generate large amounts of data do not necessarily value the benefits of releasing it as open data.

In a recent review, Zott *et al.* (2011) examine extant literature on the concept of business model. Although they share the same concerns about the lack of coherence or any unified view on the subject, their study reveals some common research trends. These can be summarised as follows: (i) the business model is emerging as a new unit of analysis; (ii) business models suggest a systemic/holistic approach that is used to explain what firms do and how they do conduct business; (iii) different activities in the firm are important features of how business models have been

conceptualised; and (iv) business models are about explaining how value is created as to how value is captured. In practice, and despite the many approaches that the literature has provided in the last twenty years, the main components of a business model can be grouped in three categories (Chesbrough and Rosenbloom 2002, Liebenau *et al.* 2012): areas of value creation (i.e. value propositions for customers and vendors), revenues (how organisations earn revenues) and logistics (issues relating to value chain). I shall use these to review the literature on business models and open data.

Even though the concepts of value creation and value capture expand on previous narrow definitions, they rely on the practices of business behaviour and its effects on economic outcomes. I suggest that conceptualising value in open data needs to extend beyond the economic-centric approach.

2.4 Value and ethics in open data

Within information systems research, there has been a long tradition of literature on the opportunities and value that new technologies can ascribe to transforming business models, organisations and industries (Brynjolfsson and Hitt 2000, Kohli and Devaraj 2003, Kohli and Grover 2008). These studies tend to focus on a narrow societal set of values: economic impacts of information technology and its manifestations within the boundaries of the firm or networks of firms (Kohli and Grover 2008, p. 24). Within management studies there is an emerging ‘turn to values’ and ethical theory (Walsh *et al.* 2003, Maurer *et al.* 2011, West and Davis 2011, Ravasi *et al.* 2012) that is in part a critical response to the spread of market fundamentalism in social life (Somers and Block 2005).

In public management studies, this trend has been echoed in the increasing attention to public values and public value creation in activities, outputs and outcomes of the interaction between citizens, governments and the private sector (Moore 1995, Bozeman 2007, Benington 2011). The idea of public value gained a dominant place after the seminal work of Mark Moore, *Creating Public Value*. Moore’s work, initially published in 1995 in the USA, offers a (re)freshing perspective on the role of the state (Moore 1995, Benington and Moore 2011). Moore’s initial piece (1995) was mainly

concerned with creating a normative theory of what public managers *should do* to create public value given the particular circumstances they are immersed in. The concept “public value” relates to that of a public philosophy of governance that a society’s collective aspirations determine where public value lies (Moore 1995, p. 52, Barzelay 2000, p. 246). Although Moore’s work is more about managerial aspects of the public sector than about providing a stable concept of public value, an implied corollary that is relevant for our work relies on the importance of considering broader democratic values (political, social, economic) in policy interventions.

A related stream refers to public values (in plural) and public interest. This research builds on the work of Bozeman (1987) and colleagues on *publicness*, and more precisely, on its normative part. In his book entitled *Public Value and Public Interest*, Bozeman defines public values as follows (2007, p. 132, emphasis added):

A society’s ‘public values’ are those providing normative consensus about (a) the rights, benefits, and prerogatives to which citizens should (and should not) be entitled; (b) the obligations of citizens to society, the state, and one another; and (c) *the principles on which governments and policies should be based*.

For the purpose of this work, the third part of the definition is particularly relevant and I shall adopt that. From this view, public values can be considered as those basic *building blocks* of the public sector. Within this third stream of work, public values have been studied as clusters or inventories in both public and private organisations (Beck Jørgensen and Bozeman 2007, Bozeman 2007, van der Wal *et al.* 2008).

A third piece of work that has influenced my work is Benkler and Nissenbaum’s article on the ethical dimension of commons-based peer production⁵ (Benkler and Nissenbaum 2006). Drawing on ethical theory, the authors identify four clusters of virtues that can be associated with the characteristics and principles of commons-based peer production. The four clusters are summarised in the table that follows.

⁵ Commons-based peer production describes a socio-economic system of production in which large numbers of people cooperate into projects, mostly without the coordination practices of traditional hierarchical organization. Benkler first coined the term in its seminal article “Coase’s penguin, or Linux and the nature of the firm” in 2002. Common examples of commons-based peer production are Wikipedia, and some open source software developments.

Cluster	Virtues
Cluster I	Autonomy, independence, liberation
Cluster II	Creativity, productivity, industry
Cluster III	Benevolence, charity, generosity, altruism
Cluster IV	Sociability, camaraderie, friendship, cooperation, civic virtue

Table 2: four clusters of virtue in commons-based peer production

These collections of works—whether public values or ethical theory applied to peer-production—offer an attractive perspective to analyse open data from a normative angle—that is, to emphasise the ethical dimensions of open data that are as important as its functional ones.

3 Economic value of open data: business models perspective

The business models that may help extracting the potential value of open data are not well understood. For example, open data is more than government data—the focus on openness stresses accessibility, reuse and open standards, and the private sector is a big generator of data. Still, many organisations that generate large amounts of data do not necessarily understand or value the benefits of releasing it as open data. The equation gets even more complicated when we try to understand how these models apply to different users and actors involved in the process. In the following subsections, I discuss three recent developments in the literature, from the perspective of businesses (usually start-ups but also more established or larger firms). These are: (i) common trends for value capture in open data (revenue stream), (ii) applications of the ‘canvas’ model for open data and public sector information re-use, and (iii) recent empirical studies that use a business model approach to map relevant activities and actors in the open data services landscape.

3.1 Value capture and open data

Making businesses and creating competitive advantages from open resources is not straightforward and represent a key challenge when it comes to turn open data into a profitable business. There are

at least three general models on how to monetize services based on open data (The Open Data Institute 2013b):

- **Freemium:** companies usually provide a basic service for free, and then charge for a “value added service” or data product; in other words, this means giving users free access to basic data/information, and then charge for certain analytics combined with other data for strategic purposes.
- **Cross subsidy:** this model is based on the idea that you can charge higher prices to one group of users to subsidize lower prices for another group; this model implies using revenues from current data services to develop new services, or charge existing consumers to cover the costs of reaching a wider set of consumers/potential clients.
- **Network effects:** by means of collaborating with other partners and/or organisations, the costs of storing and maintaining data can be reduced; value gain comes from the extended use of data or the extended reach of data for a broader community (i.e. open maps examples).

Example 1: Placr

Placr is a UK-based data service company, founded in 2010, whose core business is to provide a single UK source of transport information through the unification of timetables, live departure and disruption information for rail, metro, bus and ferry services. CEO Jonathan Raper explains how they got started:

We were looking at various business ideas that would best utilise our skills at the same time as the open data issue was gaining momentum in Autumn 2009. The Greater London Authority had a developer day to discuss open data and I got involved at that stage. When Transport for London started to release enormous amounts of data in 2010 it became obvious that there was going to be business there somewhere. It’s actually taken quite a long time for it all to become sufficiently mature to build a business on, but that is how it got started. (<http://www.theodi.org/case-study/placr>)

Placr provides an example of business model differentiation; their unique approach lies in the fact that few people have brought such different data sources together in a consistent way to improve

the experience of transport. Regarding their revenues, they use a freemium model, and combine two key services model—that is, ‘cash and carry’ and a ‘retail’ model:⁶

- The API service (‘cash and carry’) is an aggregation and analytics service for UK public transport that includes timetables, routes, and live running unified into a single API format that offers end users, developers and operators access to transport data-as-a-service;
- The retail model: they have developed a “placr.mobi” consumer web application that provides travel information for the whole country; local authorities buy the app and customise it as needed⁷.

Example 2: Mastodon C

Mastodon C is a data-expert company founded in April 2012, and the winner of London’s Green Hackathon in January 2012. The company began with their success in the competition, and thereafter creating a dashboard using open weather and power grid data to help software developers choose technical infrastructure with minimal carbon emissions. They currently have two offerings:

- “Kixi”: they run a self-managing, auto-scaling big data platform on zero carbon cloud infrastructure to process big data analytics, built on open source technologies;
- Consultancy services: they provide services to help organisation to evaluate their use of data.

Similarly to Placr, they rely on complementary services (Kuk and Davies 2011, Morgan *et al.* 2013) in order to exploit value capture in their business model.

3.2 Canvas model and open data

Ferro and Osella (2012, 2013), based on the canvas model proposed by Osterwalder and colleagues, suggest a three-dimension framework to assess archetypes of business models emerging from public sector information according.⁸ These three dimensions are:

⁶ Source: <http://www.theodi.org/case-study/placr>

⁷ Source: <https://developer.transportapi.com/plans>

- Types of data elaboration: describe the types of activities, uses and re-uses of open data (i.e. from raw data, to data analytics, validation and visualisation);
- The role of data in the value proposition: describes how open data (or public sector information) is viewed in terms of value creation, its market and value for customers;
- Price mechanisms: define the revenue mechanisms and pricing schemes for the business.

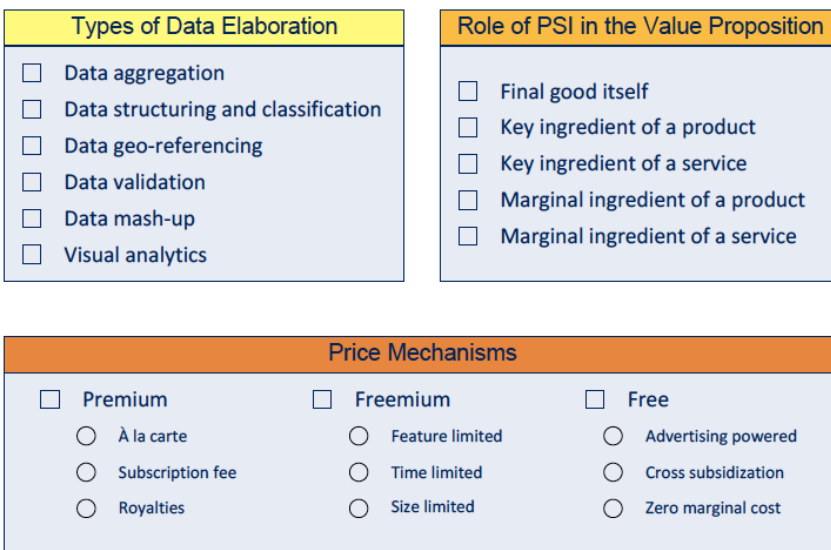


Figure 2: business models for public sector information re-use. Source: Ferro and Osella (2012, p. 4).

Business cases could be therefore analysed based on a combination of each of these elements/dimensions. For example, in the case of Placr mentioned earlier, the company is a core re-user of data, and offers end users an intrinsic value from data aggregation, classification, geolocation and referencing. The role of open data is, for Placr, a key ingredient of the services they provide (rather than a final good in itself), and the price mechanisms are a combination of freemium, and premium models as defined by Ferro and Osella. Although cases may not fit into a ‘neat’ type of business model in this framework, it provides a useful case with potentials for rich empirical analysis.

⁸ Although the authors refer to public sector information use and re-use, these models can be applied to open data too despite their conceptual distinctions explained in section two of this document.

3.3 Empirical studies of open data

These efforts are good starting points to develop business models for open data further. In order to build archetypes, one ought to provide many organisational examples, which imposes difficulties to research. In addition to the work of Ferro and Osella (2013), a research programme in Finland has also provided interesting results. Building on a research agenda for open data services (Lindman *et al.* 2013), Kinnari *et al.* (2013) studied 14 start-ups in Finland that provide business services based on open data. The authors focus on a modified version of the canvas model to study the companies' practices regarding their offerings, revenue model, usage of resources, relationships and management practices. With a focus on the value network emerging in open data services, the authors find five participant profiles: (i) commercial open data publisher, (ii) extract & transform, (iii) data analyser, (iv) user experience provider, and (v) support services & consultancy. A key finding from a policy and economic perspective is that while there is untapped potential for business value from open data, most open data services rely on ad-hoc development without expectations on long-term economic profit.

Overall, there is enormous potential to build more empirical work using a business model approach like those suggested in this section, and these initial efforts are by all means valuable. Building taxonomies of business models for open data services can be very useful for both, improving strategic decision making and enhancing policy analysis. Let me now turn into other constellations of value in open data.

4 Value and ethics in open data

The purpose of this section is to reflect on the clusters or constellations of value that emerge in open data. I found at least three preliminary clusters that are based on the theoretical developments discussed in section 2.3. While these families of values may be analytically separable, they should not be understood as neat separations. The purpose of the following sub-sections is to present what could become key categories of values when analysing the effects of open data use in diverse spheres of society.

4.1 Industry, productivity, creativity

In the previous section, I showed one important dimension of value and ethics: new businesses emerge in the open data landscape, with enthusiastic entrepreneurs that find new and creative applications in open and shared resources. As in the case of commons-based peer production, working with open data opens up new possibilities for creative and productive practices (Benkler and Nissenbaum, 2006). In the *hackathons*—common events where communities of developers gather to come up with new ideas for products or services—people engage with creativity and productivity in an active way. According to Benkler and Nissenbaum (2006), these characteristics highlight the virtues of productivity and creativity that are rather uncommon in most industrially organised structures of work.

4.2 Accountability, transparency, autonomy

The emergence of open data initiatives driven by governments (notably the UK and also the USA), came about as a result of lack of confidence in the political system and pressures for encouraging broader transparency in democratic states. Accountability, in turn, is best regarded as one of those golden concepts that no one can be against because it conveys an image of trustworthiness and transparency (Bovens 2007). Open data is seen as particularly beneficial to increase transparency, fight corruption and encourage democratic values. The quest for transparency in the public sector has indeed been the driver for many open government data initiatives in the UK. Through transparency, and appropriate contextual information, open government data contributes to accountability, democratic values and civil liberties.

- Example: Legislation.gov.uk brings together legislation from across the UK in one place, made available as open data and maintained through collaboration across the public and private sector.

Besides the common examples of opening public sector accounts, a less studied one with accountability impacts is open science. The publication of findings as well as the data they have relied upon—though not always “open”—has been at centre of the development of science (The Royal Society, 2012). Knowledge creation and accumulation relies on a community of scientist that

may support or reject theories based on their publicity. In other words, scientific production and the advance of knowledge is the result of peer-validation and production. The ‘openness’ of results allow the community to identify errors or contribute with further knowledge based on previous research outputs. Engagement and promotion of open science may translate in benefits for many spheres of cultural and social life, such as health, environment, arts and so on.

4.3 Voluntarism, sociability, cooperation, civic engagement

This is perhaps the cluster with more similarities to those in the commons based peer production case. The values in this cluster imply giving to “a commons, a community, a public, a mission or a fellowship of which the giver is part, and the giving dimension might be only one aspect of it” (Benkler and Nissenbaum 2006, p. 408). Some open data projects are indeed cases of commons-based peer production projects, where participants cooperate, contribute time, share expertise and knowledge to achieve a social or public interest good. Many open data projects, even when they seek some profit, have a civic or social goal to achieve. In the case of Mastodon C, for example, the ultimate interest is to contribute with a more sustainable environment.

Open data cases that are built upon voluntarism, cooperation and civic engagement abound. Some include:

- **OpenStreetMap:** a collaborative project to create a free editable map of the world.
<http://www.openstreetmap.org/>
- **Where Does My Money Go?** promotes transparency and citizen engagement through the analysis and visualisation of information about UK public spending.
<http://wheredoesmymoneygo.org/>
- **Open Spending:** “the core of our project is a community of transparency enthusiasts, public finance experts and data wranglers. We’re discussing and developing new ways to monitor and explain budgets and government spending through the use of technology.”
<http://openspending.org/>
- **Fix my Street:** a web application in the UK for people to report, view, or discuss local problems with their city council. Built by MySociety, a not-for-profit organisation together

with The Young Foundation, it has now spread internationally.
<http://www.fixmystreet.com/>

5 Looking ahead: challenges, issues and trade-offs

In this section, I address a number of important challenges and trade-offs that emerge from open data. The subsections are meant to be separate for easiness of reading, but overlaps do exist. In addition, I cover here important challenges that have been noted in the literature and public debates, as well as those that may be particularly relevant in terms of economic value and virtues discussed in the previous sections.

5.1 Personal data and open data: privacy and politics

Perhaps one of the most controversial issues with open data is in relation to personal data, and the protection of identities, privacy and regulation. The Information Commissioner's Office, and independent authority that upholds information rights in the UK, provides a succinct definition of personal data. Personal data are data that relate to a living individual who can be identified from either of the following⁹:

- (a) from those data, or
- (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller,
- (c) and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.

Protecting citizens identities may also be seen as a public interest, specially, when disclosing identities may be the cause of harm, discrimination or unfair treatment by public or private organisations, individuals or communities. The main trade-offs occur when trying to use open data for social good that may affect the rights of privacy of individuals.

⁹ Source: http://www.ico.org.uk/for_organisations/data_protection/the_guide/key_definitions

A good example of tensions between open data and personal data is the ongoing debate on medical trials (Edwards 2013). Having access to unbiased medical evidence is critical for improving public health. This means that, for many drugs, public access of patient level data is required. As Edwards (2013) comments on the editorial of the Journal of Research Ethics about Tamiflu and the open data campaign:

We could ask whether it is factually true that providing patient level data is unnecessary for the scientific review of a new product. A full scientific or peer review may require at least the ability to repeat the analysis undertaken, oftentimes by an in-house statistician, to corroborate the aggregate result. Indeed, the BMJ will not now accept papers for publication without a commitment to make patient level data available for independent analysis. However, the legal case rests on public access of patient level data and thus includes providing data for lay consumption.

In the case of patient level data (as in other cases of personal data more generally), the main challenge for open data comes from an accumulated body of knowledge that suggests anonymisation cannot resolve patient confidentiality successfully (Edwards 2013, Gymrek *et al.* 2013). On the other hand, issues with confidentiality may be used by certain organisations that are unwilling to open up data to protect their interest. Therefore, finding ways to balance the benefits from open data without harming individual's privacy needs policy and regulatory consideration (Martin and Bonina in press).

5.2 Data divide: Effective use of data and equality issues

The power of open data relies on its use. One acclaimed social value from open data is its potential to improve the lives of end users, and in particular, those poor and marginalised. However, open data may reinforce the more empowered actors in society (i.e. better educated, well-off economically, younger). On this respect, Gurstein (2011) warns that, unless end users are given the opportunity to use open data “effectively”, they will face similar disadvantages as those raised a decade ago in the realm of the digital divide and digital inclusion debates (Hargittai 2002, Warschauer 2002, 2004).

Gurstein (2011) suggests seven elements to maximise a meaningful and supportive effect of open data on the poor and less empowered people in society:

1. *Internet* —having an available, affordable and accessible internet service infrastructure to support making the data available to all users.
2. *Computers and software* — having access to computers and related application software to access and process available data; this includes powerful computers and infrastructure capable to comply with data analytics;
3. *Computer/software skills* — having sufficient knowledge, training and skills to use, re-use and analyse the data.
4. *Content and formatting* — having the data available in a format (i.e. language, coding for display, appropriate geocoding) to allow for effective use at various levels of literacy.
5. *Interpretation/Sense making* — having sufficient knowledge and skills to manipulate, interpret and create data analysis with added value to people’s lives.
6. *Advocacy* — having supportive resources, such as community infrastructures, training, individual resources to support local interventions based on open data.
7. *Governance* — having the necessary financial, regulatory or policy regime to enable the use of open data.

In the UK, there are still notable divides when it comes several of these elements. For example, the Office for National Statistics (ONS) suggests that 20% of all households in the UK do not have internet access (Office for National Statistics 2012)—this is about 10 million people in the country, of which a 40% are amongst the most socially and economically disadvantaged ones in the country. The ONS report, “Internet access: Households and Individuals” published in August 2012, reveals that the most important reason for not being online is lack of need. But also, at least 22% claims it is because of lack of skills, and an additional 29% accuses equipment and access costs (ONS 2012). In addition, a project led by the Royal Geographical Society with the Institute of British Geographers published in 2011 the following alarming figures (Royal Geographical Society 2013)¹⁰:

- 49% of people without access are in the lowest socio-economic groups;
- 70% of people who live in social housing are not online;
- 38% people not online are also unemployed;

¹⁰ Figures available on the 21st Century Challenges series, and are available here: <http://www.21stcenturychallenges.org/60-seconds/what-is-the-digital-divide/> Accessed on 12 August 2013.

- 15% of people living in deprived areas have used a government online service or website in the last year, compared to 55% nationally.

Therefore, access to the internet and computers, as well as lack of skills, supportive regulatory systems and local resources are not solved in the UK. This represents a challenge for a wider positive effect of open data for an inclusive society, despite the general public discourse. Open data may need direct intervention in these areas in order to avoid creating further divides.

5.3 Costs of collecting, producing and releasing open (government) data

In part, the costs of open data are related to its technical data challenges. In order to be valuable, data has to be consistent, reliable, accessible and useable. Solving issues with format, storage, analysis and accessibility may require serious investments, including human resources and material resources such as computing power, digital infrastructure and so on. Issues with data reliability are important, as new businesses depend on them. In other words, problems with datasets that are the engine for open data products may simply disappear if the conditions change.

Data reliability and continuity is listed among the key concerns for third party developers. For example, Chromaroma is a game launched in 2010 that uses the time stamped swipes in/out of London Underground stations, available through the GLA London Datastore. On the risks about working with Transport for London and relying on their data, Chromaroma's creator Toby Barnes comments: "[Transport for London]'s an organisation based in Victorian times, with its feet fixed in steam travel. So this is a whole new era for them and they are really trying to get their heads around what it means to be a data company. But it's a challenge – they're not just letting us do what we want."¹¹ As per the benefits, how do you flatten demand is one question (when users look at the app they can figure out which time is the best to use the underground). The challenge is to get efficiency from the views of different stakeholders.

Although these technical challenges broadly apply to the management of databases and data mining more generally, whether investing or not on these issues using public funds is at the centre of the debate. For example, advocates of open data suggest that the wider social and economic of opening up

¹¹ Interview in the Guardian (Kiss 2010).

government data outpaces the costs of producing and releasing them, and public sector information should be therefore free. However, other views suggest that the government should monetise the valuable datasets. This has been the case recently with the NHS patient records, which the government plans to sell to private firms such as BUPA and other private companies willing to do medical research. Besides the challenges with data protection, there is no clear answer to whether all taxpayers should pay for datasets that can be therefore profitable to a small group of companies.

Other related challenges entail the following: dealing with lack of funding (including public spending cuts), lack of ownership for system interfaces, resolving organisational silos and integrating old legacy systems.

5.4 Coordinating a diversity of interests and changes in behaviour

As noted several times, the open data landscape includes a diversity of stakeholders with different interests. Therefore, challenges of coordination exist as well as on how to build policies based on collective agreement (as the proposals of public value would suggest). Other challenges in this front include: coping with the diversity of stakeholders and differences in how they perceive what is valuable, fair or just; and the need for behavioural change among citizens and businesses. The founders of Placr, for example, addressed that a key challenge for the company has been to persuade governments to open up their data in other key public services. That is the case of the Ordnance Survey, Companies House and the Land Registry, which still appear quite closed. Such data could be used to generate better planning tools and strategic development with commercial purposes, adding value to different spheres of society.

6 Concluding remarks: the future of open data

There are growing expectations about the ways in which open data will affect innovation, economic growth and well-being in societies and communities. I am convinced that open data will shape key areas of economy and society in the next decade. Given the newness of the topic, many of the challenges associated to open data as well as the opportunities remain underexplored and we need more cumulative research on these matters.

In particular, the challenges that open data presents are not easily solvable. First, from an economic perspective, it is not clear how to monetise or create economic value from an open, public good type of resource, that will be sustainable over time. Other challenges from a social, cultural and political perspective include issues of equality and social justice (i.e. who pays for open government data, who benefits from them), tensions between personal and open data as well as how to protect anonymisation without losing the benefits of transparency and accountability (i.e. health trial example).

Therefore, a general conclusion is that we need to develop and test frameworks for understanding opportunities and challenges emerging from open data, with the capacity to inform practice, strategy and policy. These frameworks shall be multidimensional; understanding the economic value and new business opportunities is by all means important and we have some initial and valuable efforts on this front (i.e. those discussed on section three). But we also need to develop frameworks that take into consideration broader values and ethical concerns. As in the case of commons-based peer production, open data has the potential to develop cultural, social and political virtues in benefit to society (Benkler and Nissenbaum 2006). Despite the challenges discussed (and other that are not covered in these document), the potentials of open data to generate virtuous practices and value generation may be huge.

In terms of avenues for future research, I suggest to focus on specific sectors of society to make the scope and findings of research significant. On these, I suggest at least two cases for further exploration: financial services, and healthcare. I suggest a focus on these sectors given their relative importance in society and their relevance for open data services. The financial services are an exemplar given its historical use of data-driven models. Despite so, open data may challenge the established landscape with new offerings based on cutting-edge technologies, such as new equity platforms based on ‘crowdfunding’ and peer-to-peer lending. Questions to ask could be: how is open data transforming the financial services sector? What new business models emerge with open data? How does open data changes the value chain in niche sectors of the financial services? What is the relation of open data and job creation in the sector? Healthcare, in turn, offers an ideal setting to study the ethical challenges of open data vis-à-vis the social benefits of getting more efficient and effective health coverage. Relevant questions may include the following: in what ways open data can

improve hospital productivity and patient treatment? What ethical challenges emerge from open data in health care? How do we minimise them?

A final remark entails what is happening with the big companies. Open data is not just a story of start-ups and small firms, even though these are the ones that have been in the radar because of their capacity to innovate quickly, with interesting value propositions and offerings that go beyond profit-driven motivations. As shown in the survey to C-suite executives, collecting and analysing open data is seen as the second most important source of data to drive strategic decision-making. We could draw parallels with what has happened in the open source software sector. The use of open source software is not just confined to a community of programmers that produce software for ‘good’. Instead, it has permeated large corporations that see the value of engaging in these projects—because of its bigger versatility, security, cost-effectiveness and ease of innovation. The Future of Open Source Survey, an annual survey conducted by North Bridge Venture Partners and Black Duck Software, suggests that more than 50% of software acquired in the next five years will be open source¹². We can expect to see open data taking the centre of the scene in the next five years as well.

Acknowledgments

I would like to thank the RCUK New Economic Models in the Digital Economy (NEMODE) programme for the financial support received to complete this pilot report under the “Small Grants Scheme”. I am also grateful to Jonathan Liebenau, Silvia Elaluf-Calderwood and Ulrich Atz for constructive feedback and insightful conversations.

¹² (Sixth Annual Future Of Open Source Survey Results 2012, King and Hickins 2012).

References

- Agranoff, R., 2012. *Collaborating to Manage: A Primer for the Public Sector*. Georgetown University Press.
- Agranoff, R. and McGuire, M., 2003. *Collaborative public management: new strategies for local governments*. Washington, D.C.: Georgetown University Press.
- Amit, R. and Zott, C., 2001. Value creation in E-business. *Strategic Management Journal*, 22 (6-7), 493–520.
- Baden-Fuller, C. and Morgan, M.S., 2010. Business Models as Models. *Long Range Planning*, 43 (2–3), 156–171.
- Barzelay, M., 2000. The New Public Management: a bibliographical essay for Latin American (and other) scholars. *International Public Management Journal*, 3 (2), 229–265.
- Beck Jørgensen, T. and Bozeman, B., 2007. Public values - An inventory. *Administration & Society*, 39 (3), 354–381.
- Benington, J., 2011. From Private Choice to Public Value? In: J. Benington and M.H. Moore, eds. *Public value: theory and practice*. Basingstoke: Palgrave Macmillan, 31–51.
- Benington, J. and Moore, M.H., eds., 2011. *Public value: theory and practice*. Basingstoke: Palgrave Macmillan.
- Benkler, Y. and Nissenbaum, H., 2006. Commons-based Peer Production and Virtue*. *Journal of Political Philosophy*, 14 (4), 394–419.
- Boudreau, K.J. and Lakhani, K.R., 2013. Using the Crowd as an Innovation Partner. *Harvard Business Review*, 91 (4), 61–69.
- Bozeman, B., 1987. *All organizations are public: bridging public and private organizational theories*. 1st ed. San Francisco: Jossey-Bass.
- Bozeman, B., 2007. *Public values and public interest: Counterbalancing economic individualism*. Washington DC: Georgetown University Press.
- Brynjolfsson, E. and Hitt, L., 2000. Beyond Computation: Information Technology, Organizational Transformation and Business Performance. *Journal of Economic Perspectives*, 14 (4), 23–48.
- Chesbrough, H. and Rosenbloom, R.S., 2002. The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies. *Industrial and Corporate Change*, 11 (3), 529–555.
- Chesbrough, H.W., 2006. *Open business models: how to thrive in the new innovation landscape*. Harvard Business School Press.
- Edwards, S.J., 2013. Editorial: Tamiflu and the open data campaign. *Research Ethics*, 9 (3), 94–96.
- European Commission, 2011. *Open data - An engine for innovation, growth and transparent governance*. Brussels: European Commission, No. COM(2011) 882.

- Ferro, E. and Osella, M., 2012. Business Models for PSI Re-Use: A Multidimensional Framework. Presented at the Using Open Data: Policy Modeling, Citizen Empowerment, Data Journalism Workshop, European Commission, Brussels.
- Ferro, E. and Osella, M., 2013. *Eight Business Model Archetypes for PSI Re-Use*.
- Fitzgerald, B., 2006. The Transformation of Open Source Software. *Management Information Systems Quarterly*, 30 (3).
- Greater London Authority (GLA), 2013. Welcome to the London Datastore [online]. *London Datastore*. Available from: <http://data.london.gov.uk/datastore> [Accessed 16 Jul 2013].
- Gymrek, M., McGuire, A.L., Golan, D., Halperin, E., and Erlich, Y., 2013. Identifying Personal Genomes by Surname Inference. *Science*, 339 (6117), 321–324.
- Hargittai, E., 2002. Second-level digital divide: Differences in people's online skills. *First monday*, 7 (4).
- King, R. and Hickins, M., 2012. Big Companies Open Up. *Wall Street Journal*, 6 Sep.
- Kinnari, T., Lindman, J., and Rossi, M., 2013. *Long-term Business from Open Data*. Helsinki, Finland: Aalto University School of Business.
- Kiss, J., 2010. Chromaroma: the makeover London commuting has been waiting for. *PDA-The Guardian*.
- Kohli, R. and Devaraj, S., 2003. Measuring Information Technology Payoff: A Meta-Analysis of Structural Variables in Firm Level Empirical Research. *Information Systems Research*, 14 (2), 127–145.
- Kohli, R. and Grover, V., 2008. Business Value of IT: An Essay for Expanding Research Directions to Keep Up With the Times. *Journal of the Association for Information Systems*, 9 (1), 23–39.
- Kuk, G. and Davies, T., 2011. The Roles of Agency and Artifacts in Assembling Open Data Complementarities. In: *ICIS 2011*. Presented at the International Conference on Information Systems.
- Liebenau, J., Elaluf-Calderwood, S.M., and Karrberg, P., 2012. Strategic Challenges for the European Telecom Sector: The Consequences of Imbalances in Internet Traffic. *Journal of Information Policy*, 2 (0).
- Lindman, J., Rossi, M., and Tuunainen, V.K., 2013. Open Data Services: Research Agenda. In: *2013 46th Hawaii International Conference on System Sciences (HICSS)*. Presented at the 2013 46th Hawaii International Conference on System Sciences (HICSS), 1239–1246.
- Martin, A.K. and Bonina, C., in press. Open Government and Citizen Identities: Promise, Peril, and Policy. In: M.L. Smith and C. Reilly, eds. *Open Development: Technological, Organizational, and Social Innovation in International Development*. Cambridge, MA: MIT Press.
- Maurer, C.C., Bansal, P., and Crossan, M.M., 2011. Creating Economic Value Through Social Values: Introducing a Culturally Informed Resource-Based View. *Organization Science*, 22 (2), 432–448.
- Moore, M., 1995. *Creating Public Value: Strategic Management in Government*. Cambridge, MA: Harvard University Press.

- Morgan, L., Feller, J., and Finnegan, P., 2013. Exploring value networks: theorising the creation and capture of value with open source software. *European Journal of Information Systems*, 22 (5), 569–588.
- Office for National Statistics, 2012. Internet Access - Households and Individuals, 2012. *ONS Statistical bulletin*.
- Open Knowledge Foundation, 2010. Why Open Data? [online]. *Open Data Handbook*. Available from: <http://opendatahandbook.org/en/why-open-data/> [Accessed 20 Jul 2013].
- Osterwalder, A., Pigneur, Y., and Tucci, C.L., 2005. Clarifying Business Models: Origins, Present, and Future of the Concept. *Communications of the Association for Information Systems*, 16 (1).
- Ravasi, D., Rindova, V., and Dalpiaz, E., 2012. The cultural side of value creation. *Strategic Organization*, 10 (3), 231–239.
- Royal Geographical Society, 2013. Facts and figures of the digital divide in UK [online]. *21st Century Challenges - Royal Geographical Society with IBG*. Available from: <http://www.21stcenturychallenges.org/60-seconds/what-is-the-digital-divide/> [Accessed 29 Aug 2013].
- Shafer, S.M., Smith, H.J., and Linder, J.C., 2005. The power of business models. *Business Horizons*, 48 (3), 199–207.
- Shakespeare, S., 2013. *Shakespeare review of public sector information*. London: UK Department for Business, Innovation and Skills, No. BIS/13/744.
- Sixth Annual Future Of Open Source Survey Results [online], 2012. Available from: <http://www.blackducksoftware.com/news/releases/2012-05-21> [Accessed 2 Aug 2013].
- Somers, M.R. and Block, F., 2005. From Poverty to Perversity: Ideas, Markets, and Institutions over 200 Years of Welfare Debate. *American Sociological Review*, 70 (2), 260–287.
- Teece, D.J., 2010. Business Models, Business Strategy and Innovation. *Long Range Planning*, 43 (2–3), 172–194.
- The Economist Intelligent Unit, 2013. *The Data Directive. How data is driving corporate strategy—and what still lies ahead*. London: The Economist.
- The Open Data Institute, 2013a. What is open data? [online]. Available from: <http://www.theodi.org/guide/what-open-data> [Accessed 16 Jul 2013].
- The Open Data Institute, 2013b. How to make a business case for open data [online]. Available from: <http://www.theodi.org/guide/how-make-business-case-open-data> [Accessed 26 Apr 2013].
- UK Cabinet Office, 2012. *Open Data White Paper: Unleashing the Potential*. London: The Stationery Office, White Paper.
- Vickery, G., 2011. *Review of Recent PSI Re-Use Studies*. Brussels: European Commission Directorate General/Information Society.
- Van der Wal, Z., de Graaf, G., and Lasthuizen, K., 2008. What's valued most? Similarities and differences between the organizational values of the public and private sector. *Public Administration*, 86 (2), 465–482.

- Walsh, J.P., Weber, K., and Margolis, J.D., 2003. Social Issues and Management: Our Lost Cause Found. *Journal of Management*, 29 (6), 859 –881.
- Warschauer, M., 2002. Reconceptualizing the digital divide. *First Monday*, 7 (7).
- Warschauer, M., 2004. *Technology and social inclusion: Rethinking the digital divide*. Cambridge, MA: MIT Press.
- West, K. and Davis, P., 2011. What is the public value of government action? Towards a (new) pragmatic approach to values questions in public endeavours. *Public Administration*, 89 (2), 226–241.
- Willinsky, J., 2005. The unacknowledged convergence of open source, open access, and open science. *First Monday*, 10 (8).
- Zott, C., Amit, R., and Massa, L., 2011. The Business Model: Recent Developments and Future Research. *Journal of Management*, 37 (4), 1019–1042.

Further resources in open data

DATASETS

Data.gov.uk

Open data portal of the UK central government with over 9,000 datasets available—mainly from central government departments but also other public sector bodies and local authorities.

<http://data.gov.uk/>

Aberdeen City Council Open Data

Open data portal from Aberdeen City Council. Provides open and linked data and a series of new applications based on APIs.

http://www.aberdeencity.gov.uk/open_data/open_data_home.asp

London Datastore

Open data portal of the Greater London Authority, including datasets from the Mayor of London and other public sector bodies, such as the popular datasets from Transport for London.

<http://data.london.gov.uk/>

Datacatalogs.org

A comprehensive list of open data catalogues in the world that is curated by a group of experts from local, regional and national governments in the world. <http://datacatalogs.org/>

PublicData.eu

A European data portal that provides access to open and freely reusable datasets from local, regional and national governments across Europe. <http://publicdata.eu/>

OPEN DATA PLATFORMS

Comprehensive Knowledge Archive Network (CKAN)

An open-source data portal platform that provides open source tools to streamline publishing, sharing, finding and using data in open access format. CKAN is administered by the Open Knowledge Foundation. <http://ckan.org/>

Junar

A cloud-based open data platform that enables businesses, governments, and other organizations to free their data to drive new opportunities, collaboration, and transparency. <http://www.junar.com/>

Open data Kit

A free and open-source set of tools that help organise, build, manage and analyse open data, created and maintained by researchers in the University of Washington's Department of Computer Science and Engineering. <http://opendatakit.org/>

CASE STUDIES

The Open Data Institute Library

A collection of guides, case studies and consultation responses on open data. <http://www.theodi.org/library>

Data.gov.uk – open data case studies (over 70)

The UK central open data portal provides a list, summary and links to open data case studies (currently up to 70) in the UK. Case studies available here: <http://data.gov.uk/blog/open-data-case-studies>

Open Data User Group (ODUG)

The ODUG provides updates about the requirements of people who are using, or could use, the datasets that the government collects. <http://data.gov.uk/odug>