Open and Linked Data

In: The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences

By: Rob Kitchin

Pub. Date: 2014
Access Date: April 15, 2020
Publishing Company: SAGE Publications Ltd
City: London
Print ISBN: 9781446287484
Online ISBN: 9781473909472
DOI: https://dx.doi.org/10.4135/9781473909472
Print pages: 48-66

Open and Linked Data

Given the expense and resources required to produce datasets and their value in revealing information about the world, access to them has generally been restricted in some way, for example limiting access to approved users, or requiring a fee, or circumscribing how the data can be used through licensing or policy. Even when datasets have been relatively open and available, they have required specialist equipment and tools, such as computers and software, skills such as statistics and mapping know-how, and contextual knowledge concerning a field or topic, to make sense of them, much of which is beyond the capabilities of the general population. As a consequence, data and the information and knowledge derived from them have traditionally been largely closed in nature – locked inside an institution or archive. Indeed, gaining access to datasets that can help answer particular questions has been a centuries-old frustration of researchers, analysts, journalists and civil society organisations.

The open data movement seeks to radically transform this situation, both opening up data for wider reuse, but also providing easy-to-use research tools that negate the need for specialist analytic skills. The movement is built on three principles: openness, participation and collaboration (White House 2009); that through transparency, sharing and working together the value of data for society can be realised. Its aim is to democratise the ability to produce information and knowledge, rather than confining the power of data to its producers and those in a position to pay for access. In particular, attention has been focused on opening data that has been produced by state agencies (often termed public sector information – PSI) or publicly funded research (see Chapter 2), given that these have been funded by the public purse for the public's benefit, with a more limited focus on opening up data generated by private industry, which might have more proprietary value to its creators.

The open data movement has been developing for a couple of decades, in tandem with, but largely separate from, the right to information (RTI) movement concerned with making transparent key decision and policymaking by agencies (also known as freedom of information), and the open source and open science movements concerned with reconfiguring software and publishing into open and transparent forms with respect to licensing, copyright and intellectual property rights. Since the late 2000s the movement has noticeably gained prominence and traction, initially with the *Guardian* newspaper's campaign in the UK to ‘Free Our Data’ (http://www.theguardian.com/technology/free-our-data), the Organization for Economic Cooperation and Development (OECD)'s call for member governments to open up their data in 2008, the launch in 2009 by the US government of http://data.gov, a website designed to provide access to non-sensitive and historical datasets held by US state and federal agencies, and the development of linked data and the promotion of the ‘Semantic Web’ as a standard element of future Internet technologies, in which open and linked data are often discursively conjoined (Berners-Lee 2009). Since 2010 dozens of countries and international organisations (e.g., the European Union [EU] and the United Nations Development Programme [UNDP]) have followed suit, making thousands of previously restricted datasets open in nature for non-commercial and commercial use (see DataRemixed 2013). Such a shift in position has been facilitated by

influential international and national lobby groups such as the Open Knowledge Foundation and the Sunlight Foundation, accompanied by the lobbying of knowledge-economy industry groups and companies, as well senior civil servants convinced by the arguments used, and dozens of local groups seeking to leverage municipal data.

While the arguments of the open data movement are presented in a commonsensical manner, using tropes such as transparency, accountability, participation, innovation and economic growth, the rapid opening up of government and scientific data has not been universally welcomed. Indeed, many of the social, political and economic consequences of opening data are presently being revealed, critiqued and debated. Moreover, somewhat paradoxically, the scope and duration of intellectual property rights has been strengthened over the past half century, meaning that the growth of the open data movement is being paralleled by a rise in proprietary rights (Pollock 2006). Resistance to providing open access to scientific data was discussed in the previous chapter, and critiques of open data are examined in the final section of this chapter. First, however, the chapter sets out the characteristics of open and linked data, the various ways in which the case is being made for opening data, and the economics associated with making data open. Although the opening up of government data is still partial, there is little doubt that a significant shift is occurring regarding how data in general are viewed and shared.

Open Data

At one level, the term open data is relatively straightforward to define. For example, Pollock (2006) delineates it thus: ‘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’. This definition, however, hints at the diverse nature of what can be meant by ‘open’ and how it can mean different things to different agencies in the context of intellectual property rights. Openness might refer to use or reuse, reworking, redistribution, or reselling, and might have terms and conditions with respect to each. For example, a user might be able to freely use a dataset under licence, but not to rework it for profit or to resell it, and any use might require attribution. In other words, access to the dataset is open, but not necessarily what one can do with the data accessed. In other cases, organisations retain and control access to the data themselves, but make the associated metadata freely available, or an organisation might allow some users access to the data, but these can only be re-disseminated if there has been some value added to them and there is no access to the underlying primary data.

Nevertheless, a number of organisations have sought to set out the ideal characteristics of open data. Open Definition, for example, contends that a work is open if its manner of distribution satisfies the conditions set out in Table 3.1, which provide few restrictions on access, use, reworking and redistribution, and actively promote for-profit use of open data with no financial compensation for the original creator of the data. Similarly, OpenGovData sets out nine principles of open data that extend beyond the mode of distribution to also consider the nature of the data themselves (see Table 3.2). In addition to these two sets of conditions/

principles, other agencies such as the OECD (2008) and the Australian government (Fitzgerald 2010) have proposed that open data needs to be accompanied by asset lists and mechanisms of data discovery, to issue indications of data quality and reliability, use open data formats and standards that enhance interoperability, and provide easy-to-use data infrastructures that facilitate regular publication and promotion of new datasets, access to these datasets, as well as suites of basic and specialist tools that enhance use and analysis. Ensuring interoperability is particularly important because it allows for different datasets to be conjoined, thus building new, more complex datasets and revealing new insights (Open Knowledge Foundation 2012). There are no restrictions with regards to the focus of open data – it could concern any type of socio-economic, business, cultural (media, libraries, heritage), environmental, or scientific phenomena – but, in general, the emphasis to date has been on opening up data that has high public policy and commercial reuse value such as economic, transport and spatial data.

Table 3.1 Open Definition's ideal characteristics of open data

Access	The work shall be available as a whole and at no more than a reasonable reproduction cost and must be available in a convenient and modifiable form.
Redistribution	The licence shall not restrict any party from selling or giving away the work either on its own or as part of a package made from works from many different sources. The licence shall not require a royalty or other fee for such sale or distribution.
Reuse	The licence must allow for modifications and derivative works and must allow them to be distributed under the terms of the original work.
Absence of technological restriction	The work must be provided in such a form that there are no technological obstacles to use or redistribution.
Attribution	The licence may require as a condition for redistribution and reuse the attribution of the contributors and creators to the work. If this condition is imposed it must not be onerous.
Integrity	The licence may require as a condition that modified versions being redistributed carry a different name from the original work.
No discrimination against persons or groups	The licence must not discriminate against any person or group of persons, ensuring access for all.
No discrimination against fields of endeavour	The licence must not restrict anyone from making use of the work in a specific field of endeavour. For example, it may not restrict the work from being used in a business.
Distribution of licence	The rights attached to the work must apply to all to whom it is redistributed without the need for execution of an additional licence by those parties.
Licence must not be specific to a package	The rights attached to the work must not depend on the work being part of a particular package.
Licence must not restrict the	The licence must not place restrictions on other works that are distributed along with the licensed work. For example, the licence must not insist that all derived works are open.

distribution of other works	

Table 3.2 OpenGovData's principles of open data

Data must be complete	All data are made available, subject to statutes of privacy, security or privilege limitations.
Data must be primary	Data are published as collected at the source, with the finest possible level of granularity, not in aggregate or modified forms.
Data must be timely	Data are made available as quickly as necessary to preserve the value of the data.
Data must be accessible	Data are available to the widest range of users for the widest range of purposes.
Data must be machine-processable	Data are reasonably structured to allow automated processing of them.
Access must be non-discriminatory	Data are available to anyone, with no requirement of registration.
Data formats must be non-proprietary	Data are available in a format over which no entity has exclusive control.
Data must be licence-free	Data are not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed as governed by other statutes.
Compliance must be reviewable.	A contact person must be designated to respond to people trying to use the data or complaints about violations of the principles and another body must have the jurisdiction to determine if the principles have been applied appropriately.

These idealised characteristics largely regard open data as a product, rather than as a service. In contrast, Gurstein (2013) has argued that open data need to be rethought as a service process – as an interaction and relationship between data supplier and end-user. For him, it is not enough that data are simply made available by organisations for reuse. Instead, the opening of data should be more service orientated, taking into account the needs and expectation of end-users. Such a reconceptualisation suggests a different approach to data management and end-users is required by those who produce open data, one that potentially has resourcing implications. For him, this kind of approach would require end-users to be involved in the ongoing planning, development and management of open data projects, as well as the use of metrics to evaluate the success of such projects in enhancing the public good. This would necessitate capacity building programmes aimed at up-skilling users to be able to manage, process and analyse data appropriately and effectively. Given that by their nature open data generate no or little income to fund such service arrangements, nor indeed the costs of opening data, while it is easy to agree that open data should be delivered as a service, in practice it might

be an aspiration unless effective funding models are developed (as discussed more fully below).

Linked Data

The idea of linked data is to transform the Internet from a 'web of documents' to a 'web of data' through the creation of a semantic web (Berners-Lee 2009; P. Miller, 2010), or what Goddard and Byrne (2010) term a 'machine-readable web'. Such a vision recognises that all of the information shared on the Web contains a rich diversity of data – names, addresses, product details, facts, figures, and so on. However, these data are not necessarily formally identified as such, nor are they formally structured in such a way as to be easily harvested and used. Indeed, most Web documents are largely unstructured in nature. By encoding and structuring documents using unique identifiers and a mark-up language it is possible to make visible the data they contain, enabling others to automatically incorporate, process and understand them and to link them with other related data (P. Miller, 2010).

In order to make the semantic web work as linked data, Berners-Lee (2009) argues that documents published on the Internet need to conform to four 'expectations of behaviour'. First, objects and resources need to be identified within each document using a mark-up language such as XML (Extensible Markup Language) and unambiguously named using URIs (Uniform Resource Identifiers). Second, the inherent structure of the Web should be used to connect documents and information (e.g., using HTTP – Hypertext Transfer Protocol – and URIs so that names can be looked up). Third, discovering information about a named object or resource should be made easy (e.g., when someone looks up a URI, useful information should be presented by using RDF standards). The Resource Description Framework (RDF) is a document that details and describes the nature of URIs within a domain and each URI should resolve to this document (Goddard and Byrne 2010). Using these standards ensures a common framework for harvesting data across the Internet, rather than a plethora of incompatible approaches being adopted (Dietrich 2012). Fourth, links should be provided to related objects and resources, if known, so that users can discover more things.

These 'expectations of behaviour' enable documents to be machine-read and processed so that the nature of entities can be identified (e.g., United States is a country, Barack Obama is a person), concepts can be disambiguated (e.g., Mars the chocolate bar, Mars the planet; or Barack Obama as the author of a document vs. Barack Obama as the subject of a document), and to establish associations and relationships within the data (e.g., Barack Obama is the President of the United States) (Goddard and Byrne 2010). In such a way, it becomes possible for Web documents to be machine-read so that software does not just understand the content but can also derive new information and knowledge by reasoning about that content (Goddard and Byrne 2010).

When documents are published in this way, information on the Internet can be rendered and repackaged as data and can be linked in an infinite number of ways depending on purpose. However, as P. Miller (2010) notes, 'linked data may be open, and open data may be linked, but it is equally possible for linked data to carry licensing or other restrictions that prevent it being considered open' or for open data to be made available in

ways that do not easily enable linking. In general, any linked documents that are not on an intranet or behind a pay wall are also open in nature. For Berners-Lee (2009), open and linked data should ideally be synonymous and he sets out five levels of such data, each with progressively more utility and value (see Table 3.3). His aspiration is for what he terms five-star (level five) data – a fully operational semantic Web. It should be noted that for many organisations, just getting to levels above one is considered an achievement, with level five some way off without skilled labour, additional resources and new tools.

Table 3.3 Five levels of open and linked data

Level	Form	Benefits	Costs
1	Non-machine-readable	Data are available.	Data are locked in document and is difficult to release.
2	Machine-readable but using proprietary format (e.g., Excel)	Data can be analysed with proprietary software; data can be exported in other formats.	Depends on propriety software to access and use.
3	Machine-readable using non-proprietary format (eg., CSV)	Data can be analysed in any software package.	Are data on the Web, not data in the Web, and are not linked in nature and so exist in isolation.
4	Machine-readable, using non-proprietary format and URIs and RDF	Data can be accessed from anywhere on Web, be easily linked to and combined with other data, and plugged into existing tools and libraries.	Can increase data preparation time and data management and curation.
5	Machine-readable, using non-proprietary format and URIs and RDF, and linking to other data and metadata	As level 4, but data become more discoverable and users have full access to data schema/ontology.	Needs active data management to maintain inward and outward links.

Source: Adapted from http://5stardata.info/ by Michael Hausenblas.

The Case for Open Data

While some countries, such as the United States, have a long tradition of making some public sector data freely available, especially high-utility data such as census, mapping and weather data, in others access has generally been mediated through a cost recovery model that releases data for a fee and under licence. In the UK, for example, government documents are under Crown copyright and high-utility data are controlled by trading funds (e.g., map data within Ordnance Survey, weather data within the Met Office) which act as monopolies, controlling access to key datasets (Arthur and Cross 2006; Pollock 2006). Over the past couple of decades a range of interconnected arguments has been assembled to make the case for recasting the role of such trading funds and to open up public sector data. In broad terms these contentions can be divided into five forms, all of which are seen as being of benefit to both the data holder and end-user.

First, since much data held within public and non-governmental bodies concerns the operations of those bodies, they provide a means through which to measure the success of their various programmes and activities. Opening these data to public scrutiny thus makes the workings and decision-making of an organisation transparent and can be used to assess accountability and value for money (Janssen 2012; Gordon 2013). Such transparency and accountability have become increasingly important in public discourse in an age of austerity and limited resources, and in the context of ongoing debates about procurement, wastage and fecklessness within public services.

Second, enabling end-users to access an organisation's data, it is argued, allows them to become more informed about an issue, which facilitates choice and decision-making with respect to public services, and encourages active and informed participation in the public realm (Janssen 2012; Yiu 2012). Moreover, it promotes active citizenship and political involvement in shaping how local governance is performed and organised (Huijboom and Van der Broek 2011). These, in turn, foster social innovation, enhance community relations, and elevate the standard of public debate (Yiu 2012). Consequently, open data enable and promote participatory democracy.

Third, opening up data about an organisation and its performance encourages the body to utilise such data to improve operational efficiencies and productivity through evidence-informed monitoring and decision-making. Moreover, units within the institution gain access to and can utilise data from across the entire organisation, leading to new insight and knowledge and greater joined-up thinking and efficiencies (Northcutt 2012; Verwayen et al. 2011). Further, they can gain valuable feedback and advice from external agencies utilising, analysing and interpreting the data, increasing the quality, fidelity and utility of the data. These internal and external analyses lead to enhanced organisational governance and the governance of society more broadly, for example tackling fraud and other crimes by creating more effective services (Huijboom and Van der Broek 2011). They also enable the ongoing monitoring and assessment of the impact of new policies and programmes (Open Knowledge Foundation 2012).

Fourth, providing open data enables brand enrichment. Making data freely available raises the profile of an organisation, marks it as innovative, entrepreneurial and serving its public mission, increases connections and interactions with customers and end-users, and drives traffic to an organisation's services (Verwayen et al. 2011). In turn, this builds trust and reputation, creating a virtuous loop.

Fifth, while some publicly created and held datasets have marginal economic value, for example that related to cultural heritage, much has great commercial value. Publicly generated data can be used to add value to existing business data, create new applications and services and thus new markets, and improve business knowledge and decision-making (Janssen 2012; Yiu 2012). Not unsurprisingly, therefore, industry interests have been keen advocates of opening up publicly generated data for commercial reuse, especially data that have long been known to hold and create value such as that administered by public sector trading funds (e.g., map data). Restricting access to public data, which is generated in the public interest, it is argued, stifles innovation and creates a dead weight loss to the economy; that is, it limits use to those who can afford it, pricing others out of the market (Pollock 2006; Yiu 2012). Thus, public investment in generating data is being underutilised and many opportunities to add value to private sector activities are being lost (Yiu 2012). Further, the general lifting effect on the wider economy of opening up the value of public data is being missed (Northcutt 2012).

The balance of how these five arguments are mobilised across jurisdictions and context varies; however, collectively they constitute a powerful discursive regime that has been remarkably successful in persuading organisations to open up their data, not least because they dovetail with general processes of neoliberalism that promote the marketisation of public services (Bates 2012). Indeed, the arguments for opening data are mostly driven by political and business interests aimed at holding public institutions and non-governmental organisations to account and gaining access to a valuable commodity, rather than seeking to challenge the notion of intellectual property per se (Verwayen et al. 2011).

In contrast, and unsurprisingly in this context, the arguments and calls for businesses to make their data freely available for wider reuse are much less well developed and promoted. However, many of the same arguments for opening public sector data hold for business data, or at least selective elements of such data. Opening data has the potential to enable companies to benefit from the analysis and insights generated by other end-users, combine data with other datasets that enrich them in productive ways, foster collaboration with partners and suppliers that creates efficiencies in logistics and new products, and build a relationship with customers that helps expand the customer base and enhances a company's profile and reputation (Deloitte 2012). Moreover, it may aid the development of public-private partnerships. Here, an open innovation approach, which has been successful with regard to open software, is adopted, working on the principle that sharing resources will generate more commercial value than jealously hoarding and guarding them. In cases where private companies have received public monies to aid research and development, a case can be made that any data produced from such endeavours should be either fully or selectively made open as a means of producing a public good for such investment.

The Economics of Open Data

There are a number of factors that are acting as brake on the opening up of data. Some of these relate to concerns with respect to warranty and liability issues, concerns over privacy and security, potential

embarrassment over data quality or extent, and technical proficiency to build suitable Application Programming Interfaces (APIs) and manage infrastructure. By far the most significant concern, however, is financial. Data might be non-rivalrous in nature, meaning that it can be distributed for marginal cost, in theory at least, but the initial copy needs to be paid for along with ongoing data management and customer service (Pollock 2006). As such, open data might well be a free resource for end-users, but its production and curation is certainly not without significant cost (especially with respect to appropriate technologies and skilled staffing). In many cases, such data has also been a major source of revenue for organisations and, in the case of companies, competitive advantage. A key question, therefore, centres on how open data projects are funded sustainably in the absence of a direct revenue stream. There are two sides to this funding conundrum. On the one hand there is the funding required by the state to produce data and to make it open. On the other hand there is the funding to keep citizen-led initiatives going, which are reliant on volunteer labour and grants, and business models that will enable companies using open data to flourish.

The funding of government data services varies between countries and agencies. In many cases, taxes pay for the generation and processing of data. In some jurisdictions and cases, however, data services are complicated by four factors. First, they have been contracted out to third parties to manage and run on behalf of the state, where the third party adds proprietary value or makes the data available at a fee. This has recently happened with the forthcoming Irish postcodes that are going to be managed by a company on behalf of the state and will fund the operation by selling/licensing the data. Second, third-party resellers are actively lobbying to stop data being made open as it destroys their business model. Third, some state agencies operate as trading funds. They do not receive all of their funding from tax revenue, but raise a substantial portion of their income from the sale of data. Ordnance Survey Ireland, for example, operates in this way with less than half of its income coming directly from the state in the form of a subvention. Admittedly some of the payments it receives come from other state agencies, but they also come from private enterprise and individual purchases. Making all of its data available for free undermines its ability to operate and fund ongoing services. Fourth, making data open is not simply a case of publishing them in the form held by the state. Much of the data needs to be repurposed and curated to enable them to be made open (e.g., anonymised, aggregated) and new systems put in place to enable this to happen. This is not a trivial exercise and in a time of austerity and cutbacks it means reallocating funding to pay for this work, which is also needed for essential services. These four complications mean that declarations that all state data has already been paid for via taxes and should be freely available are often not as simple as desired. With respect to citizen-led initiatives, these too have costs with respect to staffing, equipment and services if they are to be sustainable as long-term endeavours. How to secure such resources beyond voluntary labour and gifts is a difficult challenge and, at present, such organisations are largely reliant on philanthropic donations and state funding, neither of which holds guarantees of renewal. And yet, the consequences of reductions or fluctuations in the financial base of open data services are likely to be a decline in data quality, responsiveness, innovation, and general performance (Pollock 2009).

It is generally argued by open data advocates that securing a stable financial base for open data within and outside the state is best achieved by direct government subvention of the costs. Proponents of this

approach argue that the increased public expenditure is offset in four ways. First, enabling direct access to the data can reduce some of the producers' transaction costs, such as staffing required for marketing, sales, communicating with customers, and monitoring compliance with licence arrangements (Pollock 2006). Second, the open model can leverage free additional labour and innovation from the crowd of users that adds significant value to the dataset and for the organisation in terms of data quality, analysis and derived knowledge, new products and innovations, and new relationships and partnerships (deVries et al. 2011; Houghton 2011). Third, open data will produce diverse consumer surplus value, generating significant public goods which are worth the investment of public expenditure (Pollock 2009). Fourth, open data will lead to new innovative products that will create new markets, which in turn will produce additional corporate revenue and tax receipts. These tax receipts will be in excess of additional government costs of opening the data. In other words, zero or marginal cost approaches are more advantageous over the long term than cost-recovery strategies (European Commission 2012).

Indeed, it is believed that opening data has the potential to create several tens of billions of euros annually in the EU alone (Open Knowledge Foundation 2012). The examples used to support such an argument are the US's decision to make publicly generated GPS and weather data freely available (Pollock 2006; de Vries et al. 2011). Both underpin multi-billion-dollar industries that generate sizeable tax revenues and employ large numbers of workers. *The Economist* (2013) estimates that 3 million jobs in the US depend on GPS. If one compares the use of public weather data in US and Europe, there is a vast difference in their use and the generation of additional value, with Pollock (2006) noting that while the 'two economies are of roughly the same size, the US commercial weather industry is over ten times larger than Europe's while the nascent weather risk management industry is over a hundred times larger'. In other words, where public data has been made openly available for free or marginal cost, it has been utilised in ways that produce significant additional value. Markets for open data are both high-end (a small base of customers paying substantial fees for high additional value; e.g., highly targeted weather forecasts) and low-end (large volumes of users for low additional value where the service is free, supported by advertising revenue; e.g., traffic apps) (de Vries et al. 2011). The potential effect of opening public sector data in Europe is illustrated by a study of 21 open data projects in 10 countries which found that moving to zero or marginal cost operations increased the number of users by between 1,000 and 10,000 per cent, attracting many new types of users, in particular SMEs (deVries et al. 2011). That said, while there will be some large wins, Eaves (2013) notes that the majority of finance created will be from a long tail of savings and efficiencies and modest increased turnover.

Despite such arguments, the jury is still out on whether opening up all public sector data is economically viable and sustainable, especially in the short term. It is certainly the case that open data that can be commercialised, leading to high-and low-end products that can generate income streams and employment. However, de Vries et al. (2011) reported that the average apps developer made only $3,000 per year from apps sales, with 80 per cent of paid Android apps being downloaded fewer than 100 times. In addition, they noted that even successful apps, such as MyCityWay which has been downloaded 40 million times, do not yet generate profits. Instead, venture capitalists are investing in projects with potential while a sustainable business model is sought. It may well be that it will take time for new innovations and markets to develop; for

example, industries underpinned by GPS took many years to blossom and mature after the decision to make the data openly available was taken in 1984. It might also be the case that some data has much wider value but weak economic value, and will always have to be subvented if it is to remain open in nature.

In the absence of the state underwriting all the associated costs of opening its data, other potential funding solutions are being examined, notably various business models. Ferro and Osella (2013) detail eight different models that are, or might be, used to fund open data initiatives (see Table 3.4). These different models are not mutually exclusive. Other semi-open models are to make the data free for non-commercial reuse, but to charge for-profit reusers, or to enter into public-private partnerships with the public sector providing the data and private companies providing finance and value-added services for access and reuse rights (OECD 2008). Another model is to build a consortium that collectively owns the data, pools labour, resources and tools, and facilitates capacity-building, but charges a membership fee to consortium members to cover shared value-added services. Which model, or combination of models, is adopted depends on which group is designated to pick up the cost for the production and maintenance of the data – users, updaters/resellers or government – and whether an organisation is seeking to recover full costs or marginal costs, or even make more than full costs which can be reinvested back into the service (Pollock 2009; Ferro and Osella 2013).

Table 3.4 Models of open data funding

Model	Description
Premium product/ service	Offers end-users a high-end product or a service that adds value to data (e.g., derived data, tools or analysis) for payment, either as fixed payment, recurrent fees or pay-per-use, without using monopoly rights. This enables the data producer to gain first-mover advantages in the marketing and the sale of complementary goods.
Freemium product/ service	Offers end-users a graded set of options, including a free of charge option that includes basic elements (e.g., limited features or sampled dataset), with more advanced, value-adding options being charged a fee. Opens up the product/service to a wider, low-end market and more casual use, while retaining paid, high-end product/service for more specialised users.
Open source	Offers end-users data products/services for free, cross-subsidised by other core funding or other products/ services.
Infrastructural razor and blades	An initial inexpensive or free trial is offered for products/services (razor) that encourages take-up and continued paid use (blades). It might be that access is free through APIs, but that computational usage is charged on a pay-as-you-go model, with the latter cross-subsidising the former.
Demand-oriented platform	Offers value-added services by cleaning, refining, standardising and linking data, providing a standard set of analytic tools, and making accessible through a one-stop shop with users charged on a freemium or premium pricing (essentially scales up options 1 and 2 to gain efficiencies).
Supply-oriented platform	Developers are given free access to data to create services and products which they can sell to the public sector body, who can then make them available to others through freemium model.
Free with advertising	Products/services are provided for free, but users receive advertising when using the product/service (revenue-generating) or the products/services are provided by different companies and branded as such to

	encourage use of their other products/services (cross-subsidisation).
White-label development	A customised product/service is created for a client and branded for their use, with that client paying a one-off fee or subscription that includes maintenance and update costs.

Concerns with Respect to Opening Data

With the exception of how to sustainably fund open data initiatives, and the possible losses for companies which have built business models on reselling public sector data (de Vries et al. 2011), there seem at first consideration to be few downsides to opening up public data for reuse, and many upsides. However, a number of counter-critiques have recently started to emerge that contend that the open data movement is not politically or economically benign, that some elements are disingenuous in their aims and, moreover, that there are a number of pernicious effects that can result from opening data that can disenfranchise some citizens. These critiques do not suggest abandoning the move towards opening data, but contend that open data initiatives need to be much more mindful of how data are made available, how they are being used, and how they are being funded. Critiques can be divided into three broad classes: open data facilitates the neoliberalisation and marketisation of public services; it promotes a politics of the benign and empowers the empowered; and it lacks sustainability, utility and usability.

Neoliberalisation and Marketisation of Public Services

Jo Bates (2012) argues that ‘open initiatives such as OGD [open government data] emerge into a historical process, not a neutral terrain’. As with all political initiatives, the politics of open data are not simply commonsensical or neutral, but rather are underpinned by political and economic ideology. The open data movement is diverse and made up of a range of constituencies with different agendas and aims, and is not driven by any one party. However, Bates makes the case that the open data movement in the UK had little political traction until big business started to actively campaign for open data, and open government initiatives started to fit into programmes of forced austerity and the marketisation of public services. For her, political parties and business have appropriated the open data movement on ‘behalf of dominant capitalist interests under the guise of a “Transparency Agenda”’ (Bates 2012).

In other words, the real agenda of business is to get access to expensively produced data for no cost, and thus to a heavily subsidised infrastructural support from which they can leverage profit, while at the same time removing the public sector from the marketplace and weakening its position as the producer of such data. Indeed, because the income from data services has disappeared (in cases where it was being funded by fees not the taxpayer), and thus the funds to support in-house production and management, public sector bodies are more likely to be forced to outsource such services to the private sector on a competitive basis or cede data production to the private sector which they then have to procure (Gurstein 2013). Here, data services and data derived from freely available public data have to be purchased by the data creator,

at the same time as the data literacy of the organisation is hollowed out. Moreover, because open data often concern a body's own activities, especially when supplemented by key performance indicators, they facilitate public sector reform and reorganisation that promote a neoliberal, New Public Management ethos and private sector interests (McClean 2011; Longo 2011). Such processes, Bates (2013) argues, are part of a deliberate political strategy to open up the ‘provision of almost all public services to competition from private and third sector providers’ with open data about public services enabling ‘service users to make informed choices *within a market* for public services based on data-driven applications produced by a range of commercial and non-commercial developers’ (2013, original emphasis). In such cases, the transparency agenda promoted by politicians and businesses is merely a rhetorical device. If either party was genuinely interested in transparency and open government then it would be equally supportive of the right to information movement and the work of whistle-blowers (Janssen 2012) and also of loosening the shackles of intellectual property rights more broadly (Shah 2013). Instead, governments and businesses are generally resistant to both.

Politics of the Benign and Empowering the Empowered

A related argument is that much of the open data movement is driven from a technical and economic perspective, and focuses largely on gaining access to the data, not on the politics of the data themselves, what that data reveal, or how they are used and for whose interests (Shah 2013). In other words, the movement largely seeks to present an image of being politically benign and commonsensical, promoting a belief that opening up data is inherently a good thing in and of itself. For others, making data accessible is just one element with respect to the notion of openness. Just as important is what the data consist of and how they can create a more just and equitable society. If open data merely serve the interests of capital by opening public data, but keeping proprietary data locked behind pay walls and protected by intellectual property regimes, and further empower those who are already empowered and disenfranchise others, then they have failed to make society more democratic and open (Gurstein 2011; Shah 2013).

Implicit in most discussions on open data is the notion that the data are neutral and objective in nature and that everyone has the potential to access and use such data (Gurstein 2011; Johnson 2013). However, this are not the case. With respect to open data themselves, as Johnson (2013) contends, a high degree of social privilege and social values is embedded in public sector data with respect to what data are generated, relating to whom and what (especially within domains that function as disciplinary systems, such as social welfare and law enforcement), whose interests are represented within the dataset and whose interests are excluded. Thus, value structures are inherent in datasets and these subsequently shape analysis and interpretation and may work to propagate injustices and reinforce dominant interests.

Citizens have differential access to the hardware and software required to download and process open datasets, as well as varying levels of skills required to analyse, contextualise and interpret the data (Gurstein 2011). And even if some groups have the ability to make compelling sense of the data, they do not necessarily have the contacts needed to gain a public voice and influence a debate, or the political skill to take on a

well-resourced and savvy opponent. Thus, the democratic potential of open data has been overly optimistic, with most users those with high degrees of technical knowledge and an established political profile (McClean 2011). Indeed, open data can work to further empower the empowered and to reproduce and deepen power imbalances (Gurstein 2011). An oft-cited example of the latter is the digitisation of land records in Karnataka, India, where an open data project, which was promoted as a ‘pro-poor’ initiative, worked to actively disenfranchise the poor by enabling those with financial resources and skills to access previously restricted data and to re-appropriate their lands (Gurstein 2011; Slee 2012; Donovan 2012). Far from aiding all citizens, in this case open data facilitated a change in land rights and a transfer of wealth from poor to rich. In other words, opening data does not mean an inherent process of democratisation. Indeed, open data can function as a tool of disciplinary power (Johnson 2013).

Here, it is important to be mindful that government data is generated for the purposes of governance. They consist of two broad types: those related to the activities of the state and those concerning citizens, places and business. The former concern how the state operates and, when made transparent, can be used to evaluate performance and accountability. Few would argue against such data being made publicly available, though the politics of such an unveiling needs to be appreciated, especially if the measures used have the counter-effect of skewing service provision to game the evaluation data (conforming to Campbell's Law, see Chapter 7), in turn negatively impacting what services are being delivered and making the lives of those people receiving the services worse. The latter consist of highly sensitive personal and institutional records. They were not created with the intention of being shared. Indeed, citizens expect them to be protected by privacy and data protection laws. Even when anonymised and aggregated, data can be quite sensitive and political. Consider, for example, social welfare and health data aggregated to relatively refined spatial units (e.g., neighbourhood level). Such data have utility for directing targeted interventions aimed at addressing social disadvantage. They also make useful inputs into data analytics that seek to socially sort and profile citizens with respect to credit and insurance risk, and can be used to create area profiles that stigmatise a locale and reduce inward investment (see Chapter 10). In other words, the data can be repurposed in various ways which have differential outcomes, and there are legitimate reasons to be cautious with respect to what government data are released and to resist the rather simplistic mantra used by some open data advocates of ‘they're our data, we've paid for them, and we should have access to them’.

Sustainability, Utility and Usability

To date, attention has been largely focused on the supply-side of accessing data and creating open data initiatives, and there has been insufficient attention paid to the sustainability of initiatives and how data are being utilised and employed. In a study of a number of different open data projects, Helbig et al. (2012) reported that many are too technically focused, amounting to Tittle more than websites linked to miscellaneous data files, with no attention to the usability, quality of the content, or consequences of its use’. The result is a set of open data sites that operate more as data holdings or data dumps, lacking the qualities expected in a well-organised and run data infrastructure, such as clean, high-quality, validated and interoperable data that comply with data standards and have appropriate metadata and full record sets

(associated documentation); preservation, backup and auditing policies; reuse, privacy and ethics policies; administrative arrangements, management organisation and governance mechanisms; and financial stability and a long-term plan of development and sustainability (see Chapter 2). Many sites also lack appropriate tools and contextual materials to support data analysis. Moreover, the datasets released are often low-hanging fruit, consisting of those that are easy to release and contain non-sensitive data that has relatively low utility. In contrast, data that might be more difficult and demanding to make open, due to issues of sensitivity or because they require more management work to comply with data protection laws, often remain closed (Chignard 2013).

Part of the issue is that many open data sites have been rough and ready responses to an emerging phenomenon. They have been built by enthusiasts and organisations who have little experience of data archiving or the contextual use of the data being opened. They have been supported and promoted by hackathons and data dives, which reproduce many of these issues. As McKeon (2013) and Porway (2013) contend, these events, which invite coders and other interested parties to build apps using open data, can be of little benefit. While they do focus attention on the data and are good for networking, those doing the coding often have little deep contextual knowledge with regard to what the data refer to, or belong to a particular demographic that is not reflective of wider society (e.g., young, educated and tech-orientated), and believe that deep structural problems can be resolved by technological solutions. They are 'built by a micro-community of casual volunteers, not by people with a deep stake in seeing the project succeed' (McKeon 2013). Further, hackathon-created solutions often remain at version 1.0, with little post-event follow-up, maintenance or development. Porway's (2013) answer to building solutions that might enact appropriate change is to try and match those who understand the data and problem (working in a sector) with those who can code, build and sustain apps, and McKeon (2013) suggests pre-planning meetings and post-event mentorship aimed at making the outcome sustainable and creating a longer-term community. In other words, there is a need to match existing expertise with developer expertise. The same need is required with open data sites, learning from those who have been in the archiving and data-sharing business for a long time, including archivists, scientists and geomaticians.

As a consequence of these various teething issues, rather than creating a virtuous cycle, where the release of more and more datasets, in more formats, produces growing use and therefore the release of more data, as assumed by the open data movement, many sites, as noted by Helbig et al. (2012), have low and declining traffic because they do not encourage use or facilitate users, and are limited by other factors such as data management practices, agency effort and internal politics. After an initial spark of interest, data use drops quite markedly as the limitations of the data are revealed and users struggle to work out how the data might be profitably analysed and used. McClean (2011), for example, notes that analysis arising from open data has had limited impact on political debates, and concludes with respect to COINS (Combined Online Information System; government financial data in the UK), that after

> a brief flurry of media interest in mid-2010, in the immediate aftermath of the release … reports explicitly mentioning COINS are now extremely rare and those members of the press who were

most interested obtaining access to it report that it has not proved particularly useful as a driver of journalism.

Where data are released periodically (e.g., quarterly or annually), usage tends to be cyclical and often tied to specific projects such as consultancy reports rather than having a more consistent pattern of use which one might expect for data that is real-time in nature or has high economic worth. In such cases, Helbig et al. (2012) observed that a set of negative or balancing feedback loops slowed the supply of data and use, thus further decreasing usage. Thus, after some initial 'quick wins' the danger is that any virtuous cycle shifts from being positive to negative, and the rationale for central government funding of such initiatives is undermined and in due course cut. The key to avoiding the creation of such a negative cycle is to ensure that any initiative focuses as much on the demand-side as the supply-side, providing users with interoperable data and analytic tools and other services that facilitate use and add value to the data, rather than simply linking to files.

Conclusion

At one level, the case for open and linked data is coninionsensical – open data create transparency and accountability; participation, choice and social innovation; efficiency, productivity and enhanced governance; economic innovation and wealth creation. Linked data convert information across the Internet into a semantic web from which data can be machine-read and linked together. Open and linked data thus hold much promise and value as a venture. However, the case for open and linked data is more complex, and their economic underpinnings are not at all straightforward. Open and linked data might seem to have marginal costs, but their production and the technical and institutional apparatus needed to facilitate and maintain them has real cost in terms of labour, equipment, and resources. Indeed, open data are far from free data, but as yet no established model has been devised to sustainably fund open data initiatives beyond government subvention, and the benefits of open data in terms of producing additional consumer surplus value and new commercial ventures, innovative products, and costs being offset by additional tax revenue are largely hypothetical or only beginning to materialise. Moreover, the potential negative consequences of opening data have not been fully explored.

Much more critical attention then needs to be paid to how open data projects are developing as complex sociotechnical systems with diverse stakeholders and agendas. To date, efforts have concentrated on the political and technical work of establishing open data projects, and not enough on studying these discursive and material moves and their consequences. As a result, we lack detailed case studies of open data projects in action, the assemblages surrounding and shaping them, and the messy, contingent and relational ways in which they unfold. It is only through such studies that a more complete picture of open and linked data will emerge, one that reveals both the positives and negatives of such projects, and which will provide answers to more normative questions concerning how they should be implemented and address issues of sustainability, usability, and their inherent politics.

http://dx.doi.org/10.4135/9781473909472.n3