From Toddler to Teen: Growth of an Open Data Ecosystem

A Longitudinal Analysis of Open Data Developments in the UK

Maximilian Heimstädt

School of Business & Economics, Free University, Berlin, m.heimstaedt@fu-berlin.de

Fredric Saunderson

National Library of Scotland, Edinburgh, f.saunderson@nls.uk

Tom Heath

Open Data Institute, London, tom.heath@theodi.org

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

Keywords: Open Data, Ecosystem, Open Government Data, Case Study, United Kingdom

A**ttempts to describe** economic phenomena using the terminology of evolutionary biology have been around for more than a century (Veblen, 1898). Several decades after Veblen, Schumpeter developed his theory of evolutionary economics (1934) as one in which competition and disruptive destruction lead to innovation. Half a century later, this theory was revived through Nelson and Winter's (1982) volume "An Evolutionary Theory of Economic Change". In this paper we make use of a related analogy from evolutionary biology to describe the socio-economic developments of our time: the ecosystem. The concept of Open Data began as a social and political idea, but has since emerged as a resource that everyone is free to use, reuse and redistribute (Open Knowledge Foundation, 2005). Open Data is available to the individual for free, can be reproduced with no additional cost and is accessible to everyone interested. Open Data can therefore be seen as a public good, or a digital commons comparable to Wikipedia or OpenStreetMap. However, Open Data has now also entered into political agendas, the media and digital growth projections of leading consultancy firms.

This research began with a consideration of how different organizations position themselves around Open Data as a topic of interest and as a resource that is non-exclusive in technical terms and not subject to any property rights. This research project therefore looked at the shared meaning of Open Data, relevant groups of actors and structural properties of their interactions which enable or constrain the growth and health of the ecosystem. To investigate the driving forces within a national Open Data ecosystem in the United Kingdom (UK), the authors utilised a business ecosystem framework to analyse developments between the late 1990s and mid-2013. The Open Data Barometer of October 2013 ranks the UK's Open Data initiative as world leading (Davies, 2013). The Open Data Index, aggregated by the Open Knowledge Foundation, also ranks the UK's Open Data ecosystem as the world's most developed, giving the country an overall score of 940 out of 1000 (Open Knowledge Foundation, 2013). Based on this evidence the UK can therefore be

regarded as a best-in-class case. In this paper, the mechanics of Open Data ecosystems are generalised, with an aim to exploring how this structure can inform Open Data development in other geographies. In the first two sections of the paper the different Open Data definitions are reviewed and aspects of business ecosystem theory are highlighted. In the third section a narrative timeline of the Open Data ecosystems in the UK is assembled before the authors create a more general conceptualization in the final section. Methodologically, the work is informed by a combination of discourse analysis and in-depth interviews, undertaken during the summer of 2013.

1.   Finding a (Mutual) Standard: Open Data Definitions

Open Data has emerged as a global and distributed movement involving various governmental and non-governmental actors. Productive communication within this system necessitates the development and use of technical and terminological standards. To illustrate the development process of this shared understanding from a UK perspective, the existing body of Open Data definitions and principles that have played a role in the UK's Open Data ecosystem has been reviewed, compared and contextualised. The list of definitions at hand is by no means collectively exhaustive, but rather represents the most influential attempts at capturing the state of the art according to our empirical data and experiences in the field.

The Open Knowledge Foundation (OKF) sought as one of its first projects to define digital openness through its Open Knowledge Definition (Open Knowledge Foundation, 2005). When the OKF was launched in 2004, the idea of openness had already gained some ground in academia, sections of the media and, notably, in the software community. The OKF, in recognition of this growth, developed its definition in an effort to prevent the concept from being diluted by a plurality of understandings (R. Pollock, personal communication, July 19, 2013; T. Steinberg, personal communication, July 17, 2013). In order to ease dissemination and understanding, the OKF outlined its criteria in a single phrase: "A work is open if it is accessible, reproducible and re-usable without legal, social or technological restriction" (Internet Archive, 2006). Over the years this summary has been adjusted into its present wording, which was released as Version 1.1 in November 2009: "A piece of data or content is open if anyone is free to use, re-use, and redistribute it – subject only, at most, to the requirement to attribute and/or share-alike" (Open Knowledge Foundation, n.d.).

In September 2007, thirty Open Government advocates gathered in Sebastopol, California, to discuss how opening government data could benefit democratic systems. The results of this meeting were eight principles (see Table 1), which define the structural properties government data must possess to be considered "open". The US non-profit organisation Sunlight Foundation sponsored this Sebastopol gathering and in 2010 released an updated version of the results, which contained two additional principles – permanence and usage cost – for Open Government Data (see Table 1).

Table 1: Open Data Definitions and Frameworks Influencing the UK Ecosystem

#	Definition or Framework	Release Data	Summary
1	Open Definition	October 2005	Use, Reuse, and Redistribution
2	Sebastopol Principles	December 2007	8 Principles for opening up explicitly governmental data
3	Sunlight Principles	August 2010	#2 plus "Permanence" and "Marginal Usage Cost"
4	5 Star Linked Open Data	May 2010	#1, #2, and Semantic Web Technologies

5	OPQUAST Checklist	April 2011	72 Principles, 17 Themes, 3 Levels of Importance
6	Open Data White Paper	June 2012	PSI, made available as Open Data according to #1, #2, and #3
7	Open Data Certificates	June 2013	Merges #1, #4, #5 into four levels of Open Data publishing quality

It is important to clarify that Open Government Data (OGD) is not an equivalent to, but a subcategory or subset of, Open Data, which may equally originate in other domains. As Yu and Robinson (2012) explain, the term Open Data remains neutral in regards to the content of the data sets and only describes its technical and legal shape. The question of subsets becomes even fuzzier when government data is distinguished from, for example, public service data (e.g. transport or water supply data) or GLAM data (whereby GLAM is the common acronym for galleries, libraries, archives & museums). Whilst government agencies by definition have a public affiliation, public service providers and cultural institutions may be privately owned, or at least organized as public-private partnerships. Kloiber (2012), however, mentions that in the majority of relevant articles, reports and strategy papers, the term "Open Data" is used synonymously for OGD. Therefore, for clarity and consistency the authors of this paper simply use "Open Data" as they refer to OGD as well as data from other public institutions (such as public transport data).

Strict interpretation frameworks, such as the Open Definition and the Sunlight/Sebastopol Principles, emphasise a dichotomous classification of data: data is either open or closed. Under these conceptualizations, there is limited scope for assessing further aspects of the data, such as the data's potential usability and adaptability beyond the scope of its initial intended use. Dichotomous models offer simple acknowledgement that data is available with proprietary restrictions ("closed") or without such restrictions ("open"), but fail otherwise to capture facets of the data's production or publication. These further aspects can have very real impacts on data effectiveness and added value.

With his five star rating for Linked Open Data, Berners-Lee (2010) highlights the importance of not just legal openness, but also technical aspects of openness. Assessment of technical openness can, for example, capture the use of open standards and non-proprietary file formats for publication. Berners-Lee's model makes the argument that there is a clear added value in focusing on technical openness, such as file formats, beyond the dichotomy of legal openness. More broadly, Berners-Lee and others (Berners-Lee, 2009; Bizer et al., 2009; Heath & Bizer, 2011) began to promote the concept of Linked Open Data (LOD) to transform "data on the web" into "the web of data" by encouraging the linking of disparate (e.g. separately hosted, dislocated, etc.) datasets to each other.

A more recent initiative, Open Data Certificates[1], was launched in 2013 by the London-based Open Data Institute (ODI). These Certificates enable data publishers and other actors reliably to assess the extent to which Open Data is published according to recognised best practices. Open Data Certificates, conceived within the ODI's development work, bring together into a single system the dichotomy of legal openness and the varied facets of technical openness. Significantly, the Certificates address the need to consider both the core legal openness of data as well as the day-to-day practical aspects of developing credible and usable open datasets that meet the technical needs of data consumers. The Certificates, according to Jeni Tennison who oversees development work within the ODI, act both as an assessment - they tell you how well you've opened your data - and as an advisory checklist - they tell you what you can do to improve the openness of your data (J. Tennison, personal communication, July 11, 2013). The Certificates also take practical and social factors into account, beyond technical aspects, in order to provide this holistic assessment

---

[1] https://certificates.theodi.org

framework. The scheme is a confluence of the OKF's Open Definition, the 5 star scheme, and the OPQUAST Open Data initiative checklist [see table 1] (J. Tennison, personal communication, July 11, 2013), the latter of which is a grouping of thirteen themes (e.g. metadata, format, and licence) for Open Data assessment.

Open Data Certificates represent an interesting step in the development of a shared understanding of Open Data. On the one hand this "badge" serves as a helpful tool that data holders may use to assess their own work. The Certificates are a tool to educate on the idea of Open Data and also serve to expand the community by spreading best practice and advice. Since the Certificates were introduced, however, there have been voices from the civil society community calling for an alternative scheme developed by a broader array of stakeholders and available at no cost. At the time of writing, such an alternative has not been made available and the question remains whether the need for such an alternative is widely recognised across Open Data communities. The theory of network effects may suggest that the market for Open Data certificates is characterized by a "winner-takes-all" dynamic, due to compatibility reasons. However, it is likely too early to assess whether the ODI's Certificates will hold a winner-take-all position in the Open Data appraisal framework.

Lastly, whilst there has been a review of definitions developed by individuals, including both formalised and un-formalised civic actors, it is important also to note how government defines Open Data. As a key stakeholder, it is essential to examine how the UK government itself understands the concept of Open Data (and Open Government Data). HM Government's (2012a, p. 8) Open Data White Paper states that Open Government Data is "Public Sector Information that has been made available to the public as Open Data". The document defines Public Sector Information (PSI) as "data and information produced, collected or held by public authorities, as part of their public task" (HM Government, 2012b, p. 8), and sees Open Data, in general terms, as data that is accessible (ideally via the internet) at marginal cost and without discrimination, available in digital and machine-readable format, and provided free of restrictions on use or redistribution (HM Government, 2012b).

2.   The Ecosystem Analogy

To analyse the provision and use of Open Data by a variety of actors, a suitable framework for investigation is necessary. As outlined above, there exists an investigative tradition which makes use of biological analogies when examining economic and other social phenomena. It is in this tradition that this piece of work is placed.

2.1.   Theoretical Grounding

This study employs the biological understanding of the "ecosystem" to examine development of Open Data in the UK. This methodological approach has previously been applied in a variety of forms (Mars, Bronstein, & Lusch, 2012) and various commentators have developed distinct strands of examination using the ecosystem framework. Hannon (1997), for example, explored the commonalities existent between ecology and economics, noting how both disciplines are concerned with the study of dynamic systems that incorporate methods of production, exchange, capital stocks, and storage. Lewin (1999) likewise observed how biological ecosystems and economic systems are complex, adaptive systems and thus follow the same deep laws. Use of the ecosystem analogy in relation to business practices, in particular, has been notably strong. By developing a survey discussion of the industrial ecosystem, the economy ecosystem, the social ecosystem, and other such analogous pairings, Peltoniemi and Vuori (2004, p. 13) position a business ecosystem as being "a dynamic structure which consists of an interconnected population of organisations".

The existing literature contextualises digital ecosystems, more specifically, as cyclical (Pollock, 2011), sustainable (Boley & Chang, 2007), demand--driven (Boley & Chang, 2007) environments oriented around the agents of various species who are mutually interdependent (Harrison, Pardo, &

Cook, 2012) in the delivery of effective and efficient value. Just as methods of production, capital stocks, etc. are interrelated in a business ecosystem, within the concept of a digital ecosystem it is the sets of data, as well as the systems and actors supporting that data, which can be understood as analogous to a cyclical, biological environment. Ultimately, the difference between a digital ecosystem and a business ecosystem is one of content: digital information (e.g. government data) in the case of the former and entities of commerce (e.g. capital and means of production) in the case of the latter. In respect of principles, the various ecosystem analogies (including those noted here) are largely comparable, in that they are about understanding and appreciating interrelationships and interdependencies between agents and entities. It is the subject matter that crucially varies between them. The foregoing conceptualizations of business ecosystems, and digital ecosystems in particular, is used for the theoretical structure of the following research, which drills further into the analogy and applies it to the concept of Open Data development.

2.2.   Methodological Approach

The authors began their research focused on the range of social and economic ecosystem analogies that have been developed from biology. The data collection consisted of a broad review of UK policy documents related to Open Data as well as a series of semi-structured interviews with members of the ecosystem (referenced as 'personal communication' throughout the text). For the document analysis, a critical discourse analysis (CDA) methodology, as outlined by Chandler and Munday (2011), was employed. Using a range of documentation written by or directly connected with the UK government, and concerning or relating to Open Data (from legislation to blog posts). CDA allowed the authors to explore the "social contexts" (Chandler and Munday, 2011) that give rise to the UK government's views on and understanding of Open Data.

In July 2013 the authors conducted 15 interviews to complement the CDA research. These lasted between 15 and 45 minutes each with an average duration of 25 minutes. Fourteen interviews were conducted and recorded face-to-face, while one was conducted and recorder using VoIP. Ethnographic data was also collected from Open Data-related events in London. The London-based Open Data Institute was used as a focal point for this research and the authors sought balance in the number of interviewees and events attended. In total, 7 employees or affiliates of the ODI were interviewed and 4 events hosted by the ODI attended. The authors also spoke to 8 stakeholders without any direct affiliation to the ODI and attended 3 external events. However, it is important to note that at the time of research, the ODI was less than a year old and many interviewees from the Institute were consulted in relation to their previous work.

The data was analysed in two ways. First, to formulate a better understanding of the UK ecosystem, a timeline charting key events, projects and decisions in the lead-up to and emergence of the UK ecosystem was developed. Second, the authors applied a strategy for qualitative content analysis as proposed by Gioia (2013). In so doing they sorted the data and developed first order codes, which helped to condense the data and highlight emerging themes. Subsequently, these first order codes were organized into second order themes, which were deduced from CDA-derived data from the literature.

The results of the research are illustrated in the following two sections of the paper. In part 3, the development of the UK's Open Data environment over recent years is presented. Part 4 removes the analogy from abstraction and investigates its applicability to Open Data ecosystems, as well as what practical implications the theory can provide.

3.   Open Data in the UK: Results of a (Brief) Longitudinal Analysis

When considering how to structure a narrative describing the UK Open Data ecosystem, two approaches were apparent. First, the environment could be dissected thematically (looking in turn at government reports, licensing frameworks, etc.). Second, the ecosystem could be analysed according to its temporal development. The latter approach provides greater insight into the gradual evolution of the ecosystem and the historical context surrounding major milestones in its

development, and is therefore adopted in this work. Previous Open Data timelines, most notably that compiled by Davies (2010), inform the research. However, the work presented here adds a number of original contributions beyond the state of the art, through the use of expert interviews, the extension of the timeline up to 2013, the inclusion of legal developments dating back to the late 1990s, and the focus on 2009 as a pivotal point in the evolution of the ecosystem.

3.1.   The Toddler Years: UK Open Data from 1998 to 2009

A strong community of activists and civil servants in the UK has driven initiatives to unlock the potential of Public Sector Information (PSI) since the late 1990s. This phase of early development proceeded from central policy papers and changes in legislation. An important precondition for development of Open Data practices in the UK was fulfilled in 1998 when the Cabinet Office published its green paper "Crown Copyright in the Information Age". This paper initiated a liberalisation process crucial to the development of open Public Sector Information (PSI). As proposed in the paper, a new "Click-Use" licensing scheme was introduced in 2000 by the Office of Public Sector Information, which allowed the commercial and non-commercial use of Crown copyright material under the precondition of attribution. At this time, use of information already gathered by public institutions was emerging not only on the UK government agenda, but on other levels as well. In November 2003 the European Union adopted the "Directive on the Reuse of Public Sector Information" with the aim of creating a common legislative framework for public bodies across Europe to release public data. In 2005 two pillars of the developing Open Data movement were firmly established with the UK's Freedom of Information Act coming into force in January and the EU directive entering into effect in November.

In addition to these governmental efforts to reimagine the use of PSI, between 2004 and 2006 civic activism also increased. In May 2004 the Open Knowledge Foundation[2] was founded in Cambridge and began to develop into the world's largest civil society organization promoting the use of Open Data. The beginnings of the organization were inspired by at least one of the founders' past experiences of Free and Open Source Software projects (R. Pollock, personal communication, July 19, 2013). These projects had managed to challenge established institutions by speaking with the single voice of a legal entity. "I was quite inspired by the Apache software foundation and it seemed that it was a natural thing to incorporate [...]." (R. Pollock, personal communication, July 19, 2013). After smaller initial events, the Open Knowledge Foundation organised a World Summit on Free Information Infrastructure in October 2005, which subsequently grew into the annual Open Knowledge Conference.[3]

The OKF was assisted in organizing these early field-configuring events in part by the fact that " [...] there was already contact to government" and that the early activists simply "[...] invited people from local government – some of the civil servants that are still around today" (R. Pollock, personal communication, July 19, 2013). When asked about whether these early connections set the ground for accelerated development of the ecosystem, Pollock stated that "there was crystallizing a group in government that we talked to a lot. [Prime Minister Gordon] Brown seemed to become more interested. [...] There were some people inside government and they managed to get Tim Berners-Lee and Nigel Shadbolt appointed [as Information Advisors to the government in 2009]. That really helped." (R. Pollock, personal communication, July 19, 2013).

In March 2006 the Guardian launched its "Free Our Data" campaign. Lead by the journalists Michael Cross and Charles Arthur, the campaign argued that government trading funds, like the Ordnance Survey and the Met Office, should provide citizens with easy access to their data, on the premise that taxpayers fund data collection. In March 2008 Newbery (University of Cambridge) and Pollock (OKF) published their study "Models of Public Sector Information via Trading Funds", which

---

[2] Open Knowledge Foundation was renamed Open Knowledge in 2014. However, this paper will use the initial name of the organization as it was during the period of empirical research.

[3] The Open Knowledge Foundation alternates annually between hosting the OK Conference and the OK Fest.

criticised the way trading funds commoditised publicly funded data. The study argued that "the problem is not the Trading Funds themselves, but the government policy" (R. Pollock, personal communication, July 19, 2013). One, largely symbolic, central element of the study was a formula, which claimed to quantify the economic benefit of Open Data and thereby provide a highly rationalized argument for future debates. The formula, irrespective of its effectiveness, made its way subsequently into other position papers and studies proposing the use of Open Data (e.g. Preische, 2014).

Simultaneous to the Guardian's campaign, the Cabinet Office began allocating resources to the emerging idea of open PSI. As a result, the civil activists Tom Steinberg and Ed Mayo, together with the Cabinet Office, published in June 2007 the "Power of Information Review", which took a "practical look at the use and development of citizen and state-generated information in the UK" (Mayo & Steinberg, 2007, p. 3). In reaction to the report and in order to further investigate application of Steinberg and Mayo's recommendations, the UK government established the Power of Information Task Force in March 2008. The Task Force was comprised of representatives from business, civil society and the government, and three months after establishment, together with the Cabinet Office, it announced the "Show Us A Better Way" competition (The National Archives, 2010). This competition made large and previously closed PSI datasets – for example, healthcare information from NHS Choices, Official Notices from the London Gazette and a list of all of the schools in England and Wales – available to developers. One of the competition winners was "Where Does My Money Go?", a service launched in December 2009 for visualising the government budget and has now been folded into the OKF-run OpenSpending.

3.2.   Teenage Years: Implementing UK Open Data since 2009

Impressive as initial efforts were, the year 2009 represents a major breakthrough, with significant developments in the United States as well as in the UK. Both countries launched data portals in that year - the so-called and much cited "data.govs". With these, a notably strong environment of reciprocal enforcement emerged as the two countries both witnessed rapid growth in the opening of PSI. As Pollock stated in our interview, 2009 saw significant shifts and "even the phrasing changed... [w]e started talking about Open Government Data" instead of reusable PSI as in the years before 2009. In January 2009, Barack Obama issued his memorandum on the Freedom of Information Act, committing his government to information openness. In the UK, when the Power of Information Task Force published its final report in February 2009, the Cabinet Office immediately began to implement its recommendations. One of these was creation of a single point of access for government data. Subsequently, beta work on the UK's Open Data portal - data.gov.uk - began in September 2009. Notably, the February 2009 release by the OKF of the first version of its Open Database Licence (ODbL) laid important groundwork for the international application of Open Data, particularly in Europe[4] (J. Tennison, personal communication, July 11, 2013) in and amongst the year's government-led changes.

The UK national data portal was set up by the Cabinet Office, and as government and civil society confirm, there were extensive exchanges between both parties regarding the layout of the portal. As Richard Stirling, an ODI-based former employee of the Cabinet Office remarked in our interviews: "In a good policy making process you are talking to all the people who are interested, working out what their relative areas of expertise are and bringing them into the solution. […] I was talking to Rufus [Pollock, founder of Open Knowledge] when we were writing the independent report [which recommended a single point of contact for government data], I was talking to Rufus when we were thinking about what we would do, I was talking to Rufus when we were doing the initial design." (R. Stirling, personal communication, July 18, 2013). A tangible outcome of this

---

[4] Although copyright is increasingly harmonised around the world, the legal situation for the protection of databases is frequently unclear in many jurisdictions. Databases in the US, for example, are not necessarily protectable by law, whereas in the EU they are. The ODbL, combined with an appropriate content licence, allows the reuse of data sets under the paradigm of Open Data.

close relationship between the Cabinet Office and the Open Knowledge Foundation can be found in the software behind data.gov.uk, the open source data management system CKAN, which is primarily maintained by the OKF. As stated on the software project's website, "the Open Knowledge Foundation's CKAN team were involved in the site from its inception, and helped develop and maintain it for the first two years. In early 2012 the UK government took its CKAN work in-house, but they continue to work closely with the CKAN team and make regular code contributions back to CKAN [...]" (CKAN, n.d.)

In May 2009 the US government launched its own data portal - data.gov - initially containing 47 data sets. The launch of this first fully operational national Open Data portal was a pivotal point for the global community of Open Data advocates. This development represented tangible proof of high-level government support for Open Data. To further solidify this support, in June 2009 the UK government appointed Berners-Lee and Nigel Shadbolt to advise government on how to open up government data in a similar manner to the US approach. Shortly after his appointment, Berners-Lee officially launched data.gov.uk to the general public in January 2010, less than a year after the US portal launched.

Ahead of the 2010 general election, the Conservative Party leader David Cameron released the "Conservative Technology Manifesto" (March 2010), which called for legislative change in favour of a "Right for Government Data" (Conservative Party, 2010, p. 3). Later that month, the incumbent Prime Minister Gordon Brown published the Labour Party's "National Digital Strategy", which called for the creation of a Web Science Institute to be directed by Berners-Lee and Shadbolt. Upon winning the election, Cameron cancelled plans for a Web Science Institute in May 2010 and focused instead on the establishment of a new Transparency Board.

Further movement commenced when Cameron sent a letter to his Cabinet Ministers in June 2010 calling for practical implementation of a transparency agenda. Tennison, who worked on development of the UK's legislation.gov.uk, likened this action to a policy implementation wake-up call: the letter was like "being hit by a big stick", and demonstrated Cameron's personal commitment to the agenda, as well as his ministers' initial lack of enthusiasm (J. Tennison, personal communication, July 11, 2013). In September 2010 the government took a major leap towards solidifying its Open Data ecosystem by releasing specifications for a new Open Government Licence (OGL) to replace the Click-Use Licence.[5] This change is notable as a "move from the transactional Click-Use to the non-transactional Open Government Licence" (employee of The National Archives, personal communication, July 7, 2013). Available through the National Archives and now in version 2.0, the OGL, which is produced by her Majesty's Stationery Office (HMSO), enables legal access to a significant corpus of Crown copyright content in a manner which is, as well as being non-transactional (and therefore not premised on the need for any form of request or authorization), in line with popular non-government open collective licensing schemes, namely the Creative Commons system. This position as a non-transactional, normative licence gives the OGL significant clout, both in terms of legal openness (the licence grants largely open use of content) and in terms of technical openness (the licence is comparable to other licences, and therefore can aid the linking of data in the manner advocated by the norms of LOD).

In September 2011, the American and Brazilian governments launched the Open Government Partnership (OGP), an international initiative for promoting transparency, civil participation and digital administration. As one of the eight founding members, the UK released its first OGP National Action Plan that same month. The following April, the UK assumed the co-chairmanship of the OGP and a month later the government responded to lobbying by interested parties and announced plans for an Open Data Institute to be set up in London. As with Labour's plans for a Web Science Institute, Berners-Lee and Shadbolt were appointed as president and chairman

---

[5] On the data.gov.uk blog, Nigel Shadbolt describes the new licence: "Based on the world-leading Creative Commons family of licences, the new licence works in parallel with them and mirrors their Attribution Licence and the Open Data Commons Attribution Licence, whilst covering a broad range of information, including Crown Copyright, databases and source codes, and applying to the whole of the UK." (http://data.gov.uk/blog/new-open-government-licence)

respectively of the ODI, which officially opened in November 2012. Speaking about its beginnings, Gavin Starks, current CEO of the Open Data Institute, commented and thereby highlighted that it is "exciting to see that Nigel [Shadbolt] managed to get a project through a change of government. That is a tough one." (G. Starks, personal communication, July 26, 2013). The Open Data Institute has since its inception secured significant match funding from the private sector and has served as an incubator for various Open Data startups. In addition to the previously discussed Open Data Certificates, the ODI in 2013 began to expand into other cities and countries on a franchise system. Organizations from across the UK and around the world were invited to apply to be an "ODI node" either on the country, city or communications level. These nodes - centres that in various localities can now act as hubs for Open Data evangelism - are all asked to adopt the ODI Charter, which describes the organizational processes and structures aligned with the principles of digital openness that the ODI embraces.

Outwith the Open Data Institute, in May 2012 a new Open Data User Group was established by the UK government. Cabinet Minister Maude appointed Heather Savory to be the group's first chair, and tasked her to run a group that would capture users' perspectives on the process of Open Data policy-making. The following month, the UK government published its foundational and highly significant "Unleashing the Potential – The Open Data White Paper", as well as individual departmental Open Data strategies and an updated version of data.gov.uk. "Open Growth", a study released by the consulting firm Deloitte in December 2012, made another attempt to quantify the economic value of Open Data for the UK economy. The report formed an integral part of the widely received "Shakespeare Review of Public Sector Information" published in May 2013. This comprehensive report was accepted by the UK government as a foundation for future policy decisions, as reflected in the "Government Response to Shakespeare Review" of June 2013.

4. Properties of Open Data Ecosystems

As the psychologist Kurt Lewin famously put it, "nothing is as practical as a good theory" (1945, p. 129). While Open Data has become practice around the world, little has been done to develop a theory on how and why this global project should develop. Current Open Data surveys usually follow a "best in class" approach, benchmarking different efforts against the best examples of the same class. However, a theoretically grounded vision of well-developed Open Data ecosystems will complement and bolster these efforts. Therefore, drawing on the literature of business ecosystem analogies, this section is focused on a handful of key properties in order to generalize on Open Data ecosystems.

Open Data ecosystems consist of organizations and individuals, which generate, share and process datasets mainly within their natural boundaries. Applied to the case at hand, this means that the UK Open Data ecosystem mainly works with UK-specific datasets, although there may be cases where organizations use datasets specific to other countries or to higher-level entities (e.g. World Bank data). Furthermore, Open Data ecosystems have to be understood as being nested, with municipal and regional ecosystems at the micro level, national ecosystems on the meso level and one global ecosystem at the macro level. All of these ecosystems mainly deal with data from their own level, but they may intersect with the other levels at times.

Furthermore, business ecosystems have been conceptualized as being (1) cyclical, (2) sustainable, (3) demand-driven environments oriented around agents that are (4) mutually interdependent in the delivery of value (Boley & Chang, 2007; Harrison, Pardo, & Cook, 2012; Pollock, 2011). In this section these structural properties are checked against empirical observations in order to study the properties of Open Data ecosystems.

By definition, the central resource of an Open Data ecosystem is Open Data. Any data that is opened up has the potential to be processed cyclically, in that it will feed back to the system/agent it originated from. However, empirical evidence shows that the realisation of this potential varies between data categories. In June 2010 the UK government released the heavily requested COINS database as Open Data. COINS contains extensive public spending data that has enabled

institutions like the OKF and The Guardian to develop in-depth spending analyses and visualisations. These have in turn been consumed by the data suppliers and have informed, or even influenced, their later decisions. Public transport data, on the other hand, serves as an example of lower cyclical potential being realised. Transport applications, like *Mapumental*, likely influence users' actions (e.g. which trains they travel on), but such services appear from our research less likely to affect the organisations that supply the data (e.g. rail operators), at least directly. Naturally, over time or with use this data may well feed back and complete the cycle. As Olivia Burman, the Cabinet Office's Head of Relationship Management and a long-term stakeholder in the UK's Open Data ecosystem noted, it is in the transport data area that the average citizen is perhaps most likely to encounter the products of Open Data (O. Burman, personal communication, July 25, 2013). They will likely do this, Burman specifies, without recognising it. When a commuter uses an app on a mobile phone to check the arrival time of the next bus, it is unlikely that data openness, let alone data provenance, as Burman remarks, will be given real consideration. In spite of this, the commuter's use of the app, and by extension the data (made usable through its openness, legal and technical), may result in (perhaps gradual) real-world changes in travel patterns, developed in response to the data. The adjustment of a bus service to accommodate changing passenger patterns, while certainly not alone a result of app use, could in this situation nevertheless represent a clear and tangible example of Open Data completing a cycle by impacting at all stages and generating new data, which in turn will be opened and can generate subsequent impact. In the travel app example, the rate of the cycle is perhaps slower than the cycle generated by COINS, but the possibility still exists.

In the literature on business ecosystems, sustainability is understood as the ability to survive without government intervention (c.f. Peltoniemi and Vuori, 2004). Although it has to be acknowledged that this is a rather narrow definition it seems applicable to evaluation of the development of Open Data ecosystems. In order to do so, it is first important to differentiate between two types of government interventions. Subsequently, the necessity of these interventions can be evaluated. The first type of intervention protects ecosystems from deteriorating through external pressures (e.g. agricultural subsidies such as the European Common Agricultural Policy). The second type of intervention supports the creation of business ecosystems with positive societal effects and high entrance barriers. Examples of the latter include subsidies for regenerative energy or a government's embracing of Open Data. In the UK, government expects positive economic and societal impacts from Open Data, ergo it nurtures the ecosystem. These interventions are differentiated by the idiosyncrasy that the government itself is the bottleneck of the ecosystem, as it is the majority data holder. Government must intervene *in itself* and ensure internally sustainable data provision. However, long-term sustainability can only be achieved when the relevant data suppliers experience a tangible benefit, and the longer and less clear the process to obtaining that benefit is, the less likely that sustainability may be realised, as suppliers can shy away.

It is a task for the UK government-funded, but independently operating, ODI to generate and disseminate case studies that clarify the benefits of Open Data and in turn promote sustainability within the ecosystem. The ODI functions in part as a startup-incubator for Open Data initiatives, and in this capacity works to promote the legal and technical openness of data across sectors. As illustrated by this need for clarity and confidence, government intervention not only has to tackle the supply side of data openness and data provision, but the demand side as well (S. Coleman, personal communication, July 15, 2013; G. Starks, personal communication, July 26, 2013). Governments historically intervene to support and give weight to sectors, industries, processes, etc. when those are, inter alia, new and upcoming. It is exactly this sort of support that governments - stable actors - can provide on the demand side of Open Data, and so lend the ecosystem a healthy route to sustainability.

In (many) models of healthy, non-monopolistic business ecosystems demand regulates supply. However, Open Data ecosystems operate slightly differently. The main resource – Open Data – is often produced within natural monopolies, due to high fixed costs, low variable costs and a rather small number of potential customers for the data (take for example, the existence of a single

statistics agency in a national administration, such as the UK's Office for National Statistics). These natural monopolies within public services likely have certain economic benefits, but they also preclude or exclude the competitive environment that is so often the cornerstone of innovation. As has been shown above, data-holding bodies only publish their data sustainably *if* they experience demand, which they can see will benefit them in turn (for example, an app built from a transport operator's data could serve to help the operator meet demand accurately and therefore win more travellers, or in other words earn more revenue). In the UK it has been observed that in some data areas – such as transport, financial and health data – an initial release triggered significant demand, noticeable through early use cases such as *Prescribing Analytics*, *Where Does My Money Go?*, and *CityMapper*. However, other key datasets, such as the Postcode Address File (PAF), which are in high public and academic demand (Shadbolt, 2013), have not been opened (Savory, 2013) (demand has not generated supply).

Lastly, the authors investigated whether agents in Open Data ecosystems are mutually interdependent in their delivery of value. The minimal value chain within Open Data ecosystems consists of three elements: data suppliers, data intermediaries, and data consumers. Whilst intermediaries and consumers usually conduct a traditional exchange of goods, suppliers (embodied largely by public agencies) are required to provide Open Data to the public as part of their operational mandate. If, for whatever reason, a data-collecting agency stops providing data openly, its core business would likely be unaffected (at least directly). However, external developers, who build businesses and services based on that data, would be unable to continue working. In this manner, Open Data ecosystems do not often show robust mutual interdependence, but rather they normally demonstrate significant one-sided dependency. This is likely to have adverse effects on the ecosystem when it is not regulated by the government.

The final property examined is the reciprocal relationship between individuals within the government and civil society organizations that work to develop Open Data practices. In the case of the UK Open Data ecosystem, the noted high rate of development since the late 1990s was in part made possible through personal ties between innovation officers within government and leading activists outside of government. These close connections enabled a high rate of information exchange between the parties and enabled cooperation beyond certain bureaucratic procedures. However, the single case study used in this research should not be used to generalize on this parameter for a healthy Open Data ecosystem.

5. Conclusion

The United Kingdom has incubated and advanced a robust and world leading Open Data ecosystem over the past 15 years. In that time there have been two primary trajectories: the push of activists and the initiative of government itself. With a greatly accelerated pace since 2009, the UK has seen the latter of these two seize the agenda ever more and establish meaningful Open Data policies as part of a determined agenda for growth. This paper examined the UK's experience of establishing a functioning Open Data environment and focused its analysis on the applicable notion of an analogous ecosystem: a geographically nested system of organizations producing, sharing and processing Open Data, in a cyclical, sustainable and demand driven system where actors are mutually interdependent.

Our work has found that in many respects the last 15 years have shaped the early movement around Open Data in the UK – the toddler years – into what is now a relatively healthy ecosystem in its teens. The ecosystem remains for the time being an adolescent, rather than a fully-fledged adult, because it is yet to function with total independence, although it is clearly on a strong course to do so soon. There are clear signs in this ecosystem of a cycle, of sustainability, of demand encouraging supply and of dependence developing between suppliers, intermediaries and users. However, it has also been found that significant gaps and shortcomings remain in this teenage system. Most prominently, demand is not yet fully encouraging supply and actors have yet to experience fully mutual interdependence.

On the basis of these findings, the authors propose that Open Data initiatives, in the UK and elsewhere, be assessed by ecosystem criteria. It is well established that the tenets of an ecosystem generate strong, sustainable and meaningful survival. The UK, a leader already by many Open Data assessments, is well on its way to establishing a fully functioning ecosystem. But there is, as outlined, more to be done. It is the view here that once the four ecosystem criteria are fully and comprehensively met, the environment can be considered developed and sustainable: a fully-grown adult ecosystem. The UK is not yet at the end of the road, nor are other governments' Open Data endeavours. However, understanding the significance of an operational ecosystem (as outlined here) and what this entails in the Open Data context can clearly be of benefit to the initiative.

References

Berners-Lee, T. (2009). Putting Government Data Online. Retrieved August 11, 2013, from http://www.w3.org/DesignIssues/GovData.html

Berners-Lee, T. (2010). Linked Data – Design Issues. Retrieved November 18, 2013, from http://www.w3.org/DesignIssues/LinkedData.html

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS), 5(3)*, 1-22.

Boley, H., & Chang, E. (2007). Digital Ecosystems: Principles and Semantics. Digital EcoSystems and Technologies Conference, 2007. DEST'07. Inaugural IEEE-IES (pp. 398-403). IEEE.

CKAN (n.d.). CKAN Website – Case STudy on data.gov.uk. Retrieved July 18, 2014, from http://ckan.org/ case-studies/data-gov-uk/

Conservative Party (2010). Conservative Technology Manifesto. Retrieved December 4, 2013, from http://www.theguardian.com/politics/2010/mar/11/conservative-technology-manifesto

Chandler, D. & Munday R. (2011). Critical Discourse Analysis (CDA). *A Dictionary of Media and Communication*. Oxford: Oxford University Press. Retrieved July 8, 2013, from http://www.oxfordreference.com/view/10.1093/acref/9780199568758.001.0001/a cref-9780199568758-e-0567#

Davies, T. (2010). A Timeline of Open Government Data. Retrieved November 18, 2013, from http://www.opendataimpacts.net/2010/06/a-timeline-of-open-government-data/

Davies, T. (2013). Open Data Barometer – 2013 Global Report. Retrieved November 18, 2013, from http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf

Hannon, B. (1997). The use of analogy in biology and economics: From biology to economics, and back. *Structural Change and Economic Dynamics, 8(4)*, 471–488.

Harrison, T. M., & Pardo, T. A., & Cook, M. (2012). Creating Open Government Ecosystems: A Research and Development Agenda. *Future Internet, 4(4)*, 900-928.

Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology, 1(1)*, 1-136.

HM Government (2012a). Open Data White Paper Unleashing the Potential. Retrieved June 10, 2013, from https://www.gov.uk/government/uploads/system/uploads/attachment_ data/file/78946/CM8353_acc.pdf

HM Government (2012b). Public Sector Transparency Board: Public Data Principles. Retrieved June 14, 2013, from http://data.gov.uk/library/public-data-principles

Internet Archive (2006). The Open Knowledge Definition. Screenshot from January 11, 2006. Retrieved December 3, 2013, from https://web.archive.org/web/20060111191539/http://www.openknowledgefoundation.org/okd/

Kloiber, J. (2012). Open Government Data – Zwischen politischer Transparenz und Wirtschaftsförderung. Retrieved August 11, 2013, from http://igitur-archive.library.uu.nl/student-theses/2012-1004-200536/Masterthesis_3806316_JuliaKloiber.pdf

Lewin, R. (1999). *Complexity: Life at the edge of chaos*. Chicago: University of Chicago Press.

Mars, M., & Bronstein, L., & Lusch, F. (2012). The value of a metaphor: Organizations and ecosystems. *Organizational Dynamics*, *41*, 271—280.

Mayo, E., & Steinberg, T. (2007). *The power of information*. London: Cabinet Office.

Nelson, R. R., & Sidney, G. (2005). *Winter (1982) An evolutionary theory of economic change*. Cambridge: Belknap.

Open Knowledge Foundation (n.d.). History of the Open Definition. Retrieved December 3, 2013, from
http://opendefinition.org/history/

Open Knowledge Foundation (2005) Open Knowledge Definition. Retrieved November 18, 2013, from
http://blog.okfn.org/2005/10/19/open-knowledge-definition-released/

Open Knowledge Foundation (2013). Open Data Index – Published October 28th, 2013. Retrieved November 18, 2013,
from https://index.okfn.org/country

Ordnance Survey (2010). Ordnance Survey launches OS OpenData in groundbreaking national initiative. Retrieved
December 2, 2013, from http://www.ordnancesurvey.co.uk/about/news/2010/OpenData.html

Peltoniemi, M., & Vuori, E. (2004). Business ecosystem as the new approach to complex adaptive business environments.
Proceedings of eBusiness Research Forum, 267–281.

Pollock, R. (2011). Building the (Open) Data Ecosystem. *Open Knowledge Foundation Blog*. Retrieved July 10, 2013, from
http://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/

Preische, J. (2014). Digitales Gold. Nutzen und Wertschöpfung durch Open Data für Berlin. Retrieved July 18, 2014, from
http://www.tsb-berlin.de/media/uploads/publikationen/2014_Digitales_ Gold_Open_Data.pdf.

Savory, H. (2013). Open Data User Group Response to RM Postcode Address File licensing consultation – September
2013. Retrieved November 18, 2013, from http://data.gov.uk/blog/open-data-user-group-response-to-rm-postcode-
address-file-licensing-consultation-%E2%80%93-september-2013.

Schumpeter, J. A. (1934). *The theory of economic development: An inquiry into profits, capital, credit, interest, and the
business cycle*. London: Transaction Publishers.

Shadbolt, N. (2013). A Cornerstone for Open Data: The Postcode Address File. Retrieved November 18, 2013, from
http://theodi.org/blog/cornerstone-open-data-postcode-address-file.

The National Archives (2010). Snapshot of the website showusabetterway.com taken on 02/04/2010. Retrieved December
2, 2013, from http://webarchive.nationalarchives.gov.uk/20100402134053/showusabetterway.com/

Veblen, T. (1898). Why is economics not an evolutionary science?. *The Quarterly Journal of Economics, 12(4)*, 373-397.

Weinberger, D. (2011). *Too big to know: rethinking knowledge now that the facts aren't the facts, experts are everywhere,
and the smartest person in the room is the room*. New York: Basic Books.

Yu, H., & Robinson, D. (2012). *The New Ambiguity of Open Government*. Princeton CITP/Yale ISP Working Paper.

Yin, R. K. (2009). *Case study research: Design and methods*. London: Sage Publications.

Chadwick, A. (2006*). Internet Politics: States, Citizens, and New Communication Technologies*. Oxford: Oxford University
Press.

About the Authors

Maximilian Heimstädt

Maximilian Heimstädt is a PhD student within the department for organisation theory at Free University Berlin, where he
investigates the emergence and management of municipal Open Data ecosystems. He previously studied an M.Sc. in
Management and Information Technology at the University of St Andrews, Scotland, and worked as a visiting researcher at
the Open Data Institute in London.

Fredric Saunderson

Fredric Saunderson completed his Digital Asset Management M.A. at King's College London in 2013. Following his degree, Fredric
worked as a researcher for the Open Data Institute in London, before moving to the National Library of Scotland in January 2014. In
his M.A. dissertation, Fredric examined public engagement with Open Government Data in the United Kingdom.

Dr. Tom Heath

Tom Heath is Head of Research at the Open Data Institute in London, UK, where he leads a diverse, multidisciplinary
programme of research exploring the impact, applications and implications of Open Data. Tom has a background in
commercial and academic research, and is an internationally recognised expert in the field of Linked Open Data. He holds a
B.Sc. in Psychology from the University of Liverpool and a Ph.D. in Computer Science from The Open University.