

ENABLING AN OPEN DATA ECOSYSTEM: PRELIMINARY FINDINGS FROM THE MARKET

Diego Ponte

Department of Economics and Management

University of Trento

diego.ponte@unitn.it

Research in progress

Abstract

Over the last decades, the interest in open data has increased dramatically in line with the improvements of information and communication technologies. Attention of institutions and private organizations focused on the value of the public sector information. Scholars and practitioners state that open data might have a huge impact in terms of efficiency within the public sector, social and economic wellbeing and might improve transparency and participation of people. While this is true, open data shows some issues that needs to be addressed. Among others, the most important are the technical quality of the published data and the sustainability over time of the whole ecosystem. Both issues are intertwined as actors providing services to the users need to both 1) assure that data is correct and 2) achieve economic sustainability. In such a market, an important role is growingly played by the enabler i.e. those actors that provide the technical infrastructure which allows for the sharing, linking and re-use of data. As the open data market is at its infancy, a survey of the market shows that such actors are very different in terms of nature and business models.

Keywords: Open data, open government, PSI, Business model, ecosystem

1 Introduction

Over the past decade, 'openness' of the public sector information (thereafter PSI) has become one among the most debated topics worldwide. PSI might be defined as the set of data that public sector institutions produce, collect and disseminate. Data belongs to different domains: society, economy and business, geographical territories, weather, tourism, education, etc. The interest has grown in line with the improvements of the information and communication technologies. In the last couple of years, public institutions and large private organizations worldwide have started to open their data on the web for the benefit of the society at large.

Scholars and practitioners underpin that such kind of data might have a huge impact in terms of efficiency within the public sector, might improve social and economic wellbeing and foster more transparency and participation of people (Gurin, 2014; Pollock, 2006; Huijboom and Van der Broek, 2011; Janssen, 2012; Yiu, 2012; Kitchin, 2013).

While in theory this is true, practicing open data might be problematic. Open data shows some issues that needs to be addressed. While sharing, linking and reusing open data are key activities for an open data ecosystem to work, technical quality of the published data and the privacy of people are important issues at stake. Raw data itself has no value, this latter is generated by technical factors as well as a combination of organizational/economics elements that allow the creation of a sustainable value chain.

An important role in this sense is growingly played by the enablers i.e. those actors that provide the technical infrastructure which make it easy to share, link and re-use open data (Immonen, et al., 2014; Zuiderwijk et al., 2013).

While the role of enablers is key for the wellbeing of the market, few empirical research can be found on this topic. There is no systematic analysis of how the specific role of such actors is evolving (Zuiderk et al., 2013). A survey of the big players in this sense shows that these have different nature, business models and goals although the private organizations are playing a predominant role. As the market is at its infancy, this work tries to depict a preliminary picture of the role of such players.

2 The value of open data

Since the last two decades, many countries started to make open data widely available. United States and UK were the first countries to push for this new approach. The European Union endorsed the process of opening the Public Sector Information (PSI, i.e., data and information held by public bodies) since 1998 with the publication of the “Green Book on Public Sector Information” (European Commission, 1998). This initial input translated into actual provisions with the Directive no. 2003/98/CE (Public Sector Information Directive), then modified by the Directive no. 2013/37/UE. Best practices reports have been also published in recent times (Lee, et al., 2014). Currently, the economic value that can be generated by open data is still an open issue as there is no widespread agreement¹. An European Union report, estimates the opening public data might generate direct benefits for as much as 40 billion euros per year (Vickery, 2011) while a study of the United Kingdom states that the direct benefits of open data for the UK only might be worth 1,8 billion pounds (Shakespeare, 2013). In 2013 a report by McKinsey estimated the global direct value of open data as higher as 3 trillion dollars a year (Manyika et al., 2013). The report states that such value might be generated in many ways: for instance, open data might improve the productivity of employees by providing new abilities and knowledge; on the customers’ side, open data might also raise transparency and knowledge. Finally, a recent report of the World Bank (World Bank, 2014), states that, while the measurement of the benefits is still imprecise, the economic potential of open data might be huge; most part of the benefits could be indirect meaning that users rather than data providers could benefit the most from them. The interest in the open data market is also indirectly highlighted by the many intermediaries providing services (like app for smartphones) thanks to the open data (as an instance think about the open data 500, which is the USA list of the biggest 500 companies using open data).

While many authors highlight the value of open data, the open paradigm is also associated with some critical aspects. First, recent research shows that open data lack of utility and are poorly usable. Helbig et al. (2012) highlight that data that are publicly disclosed are often too technical (hence, hard to be interpreted without adequate skills) or published as “little more than websites linked to miscellaneous data files, with no attention to the usability, quality of the content, or consequences of its use” (p. 9). This issue is also linked to privacy issues. Indeed, as European, national, and local provisions aim at protecting citizens’ privacy, often data are disclosed only in aggregated, anonymous and abstract forms thus disclosing very little informative power (Ohm, 2009). As a matter of fact, raw data are per se poorly informative: they require a consistent management strategy (for cleaning, organization, recording and upgrading) underpinned by systematic business models (Janssen and Zuiderwijk, 2014). An effort in this sense is growingly played by enablers, i.e. those platforms that usually combine heterogeneous data (public as well as private data) from different sources (e.g. different public institutions from all over the world). Such infrastructures have been defined as a marketplace for open data (Chui, et al. 2013). Currently, the tumultuous growing of actors providing heterogeneous services in the sector – being these either public bodies or private companies – shows that the sector is at an embryonic stage.

¹ It is noteworthy that Open Data is provided not only by the public sector but by corporations and other communities as well. Said that, the role of the public sector in providing open data is clearly prevalent.

The next section briefly depicts the structure of the sector and describes the critical role of enablers.

3 Infrastructure and sustainability of the sector

A recent study, identified five types of actors in the sector (Deloitte Analytics, 2012). These are:

- **Suppliers** These are organization that collect and publish their data by means of an open interface. While for these actors there could be no direct economic benefit, this might be indirect such as an increased wellbeing of the society at large, or reputation and increased engagement of customers.
- **Intermediaries** Open data market is made of many intermediaries, e.g. those actors that use open data to provide services to the final users. Three main types of intermediaries might be identified:
- **Aggregators** These are organizations that collect and aggregate open data in order to create knowledge such as for instance complex correlations among data.
- **Developers** These actors design, implement and sell application that directly use open data. Typical example might be app for the public transport.
- **Enrichers** these actors use open data to enhance (enrich indeed) their existing services or products.
- **Enablers** These actors help using open data. As said before, one of the main issues of the sector is the reusability and sharing of data. As such, enablers might be the most important actors in the sector. They have an infrastructural role as they provide services such as management, retrieval, storage, of open data to other actors in the market.

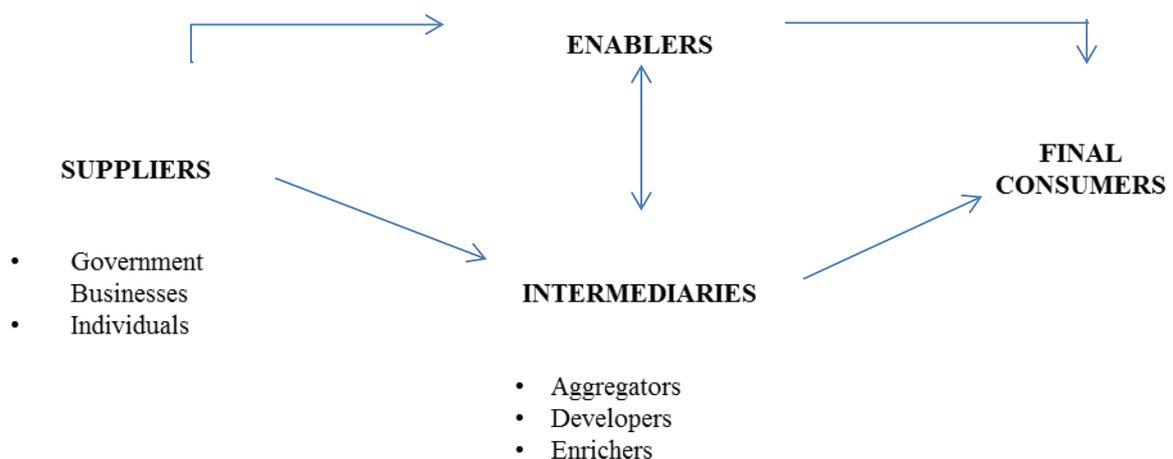


Figure 1. Ecosystem of Open Data. (Own elaboration from Deloitte Analytics, 2012)

The key role of enablers is also highlighted by Ferro and Osella (2013). In their model of the open data ecosystem these authors identify enablers as a separate and critical category of actors. According to these authors, enablers have specific kinds of business models that might allow them to survive and prosper. The business models are as follows:

- **Infrastructural Razor & Blades.** This business model is chosen by organizations that facilitate access to PSI resources. The “razor & blades” model is based on an inexpensive offer that encourages continuing future purchases of follow-up items or services. In the PSI environment, this means that enablers might offer free access to data sets while charging users for certain services (e.g. use of computational resources).

- **Demand-Oriented Platform.** The enabler provides other actors with easier access to open data which usually is collected, stored and harmonized on proprietary servers. Data is then made available to users by means of easy to use interfaces. Users pay for the added value provided by enablers (e.g. consultancy services).
- **Supply-Oriented Platform.** This business model is very similar to the previous one except for the fact that suppliers of data are charged in lieu of intermediaries and users.

In the next section, a preliminary analysis of the market is presented. This analysis lists the actors that at the time of writing are recognized as enablers at a global level. The nature and the business model are also presented.

4 Analysis

The analysis of the market was made in the second half of 2014². The analysis started gathering information about the major organizations working as enablers worldwide. The work was restricted to the biggest players in the market i.e. those players that cover open data archives of many countries; the analysis intentionally left behind national-level actors. This choice relies on the fact that, as recent articles show, the market is fast changing, with new actors coming in, actors acquired by other actors in the market, etc. (Howard, 2013; Anderson, 2014).

The work found 14 big multinational players (table 1). Of these, Datamarket was recently acquired by Qlik and was not taken into further consideration for the analysis.

The analysis shows that almost all actors are private organizations (11 out of 13). Eight of them offer services/products for open data only while five embed open data services in a set of more complex “big data” services. More interestingly, among those who offer big data services, four out of five actors offer a set of more complex services such as advanced analytics. On the other hand, it seems that among the eight actors who offer just open data services, only one offers more complex services.

Coming to the sustainability of the services, it seems that all of the three business model suggested by Ferro and Osella (2013) have been adopted by at least one actor. The first model, the “razor and blades”, seems to be used by four actors i.e. Amazon web services, Enigma, Junar and OpenDataSoft.

Five actors adopted a “demand-side” business model. All of these actors are private companies and most of them offer big data services. As a matter of fact, the focus of this group is the set of services linked with data management and storage regardless of the nature and source of data (except for Microsoft).

The four enablers which adopted a “supply side” approach seem to be more oriented to open data only. This is also confirmed by the fact that the main customers of these actors are public institutions. In this category we find both profit and not-for profit actors. Among others, one of the main actors is certainly the Open Knowledge Foundation (OKFN). Other than consultancy services, OKFN provides CKAN which is an open source portal for the management of open data. While the market is fast moving, at the moment of writing, Socrata, a private firm, seems to be its main competitor. Open Data Institute (ODI) is an not for profit institution but is mainly focused on the UK market.

² The survey is still ongoing as the research is in progress.

Name	Business Model	Type	Type of data	Founded	Advanced analytics
Amazon web services	Razor and blades	Profit	Mixed	2006	no
Cloudera	Demand-Oriented	Profit	Mixed	2009	yes
Collibra	Demand-Oriented	Profit	Mixed	2008	yes
Datamarket	...	Profit		2008	
Engage	Supply oriented	Non Profit	Open data	2012	no
Enigma	Razor and blades	Profit	Open data	2012	yes
Infochimps	Demand-Oriented	Profit	Mixed	2009	yes
Microsoft Open Government Data Init.	Demand-Oriented	Profit	Open data	..	no
MuSigma	Demand-Oriented	Profit	Mixed	2004 (initial creation)	yes
Open Knowledge	Supply-oriented	Non Profit	Open data	2004	no
ODI	Supply-oriented	Non Profit	Open data	2012	no
Junar	Razor and blades	Profit	Open data	2009	no
OpenDataSoft	Razor and blades	Profit	Open data	2011	no
Socrata	Supply-oriented	Profit	Open data	2007	no

Table 1. *Main actors working with open data, their business model and nature.*

5 Conclusions and future work

The preliminary analysis of the enablers in the open data ecosystem shows that the market is populated by a number of actors which have very diverse business models, strategies and characteristics. The role of the enablers as technical infrastructure for open data, while considered key for the wellbeing of the sector, is a spurious role and mainly embedded in a set of other services. The configuration of the sector is quite jeopardized with associations and private companies striving to become big players in a new and promising sector. Indeed, the brief review presented above, shows that - to survive in the sector - many actors play different strategies. Nonetheless, the majority of big players seem to be private companies.

From an organizational point of view, it is possible to say that the growing demand for openness of data seems to push actors that already work in the market of “data analytics”, to embed infrastructural services for open data. On the other hand, it seems that non-profit organizations are more oriented in paying an important role worldwide by promoting and fostering open data as a culture of data management. The sector is also populated by organizations that offer basic platforms for the management of open data while not offering complex services.

As the open data phenomena is quite new, the dynamics of the market will quite certainly evolve in the next future as new public and private institutions will provide open data and users will become more aware of the possibilities. At the moment it seems too early to clearly define the future evolution of the sector.

To conclude, this work needs further development. The analysis is at an initial phase. Further work needs to be done in order to clearly identify the characteristics of the enablers and the elements that will allow them to survive over time.

References

- Anderson, G. (2014). Iceland's Datamarket acquired by Qlik, Arcticstartup, November 04, 2014, Available at: <http://arcticstartup.com/2014/11/04/datamarket-acquired-by-qlik>
- Deloitte Analytics (2012). Open Growth - Stimulating Demand for Open Data in the UK. A Briefing Note from Deloitte Analytics. Deloitte, UK.
- Chui, M., Farrell, D., & Van Kuiken, S. (2013). Generating Economic Value through Open Data. Edited by Brett Goldstein with Lauren Dyson, 163.
- European Commission (1998). Public sector information: a key resource for Europe – green paper on public sector information in the information society, COM(1998)585. Available at: <http://aei.pitt.edu/1168/>.
- Ferro, E. and Osella, M. (2013). Eight Business Model Archetypes for PSI Re-Use. “Open Data on the Web” Workshop, April 23rd-24th, 2013, Google Campus, Shoreditch, London. Available at: http://www.w3.org/2013/04/odw/odw13_submission_27.pdf
- Gurin, J. (2014). “Open Governments, Open Data: A New Lever for Transparency, Citizen Engagement, and Economic Growth”. SAIS Review of International Affairs, 34(1), 71-82.
- Howard, A. (2013). “Open data economy: Eight business models for open data and insight from Deloitte UK”, Radar, January 28, 2013, Available at: <http://radar.oreilly.com/2013/01/open-data-business-models-deloitte-insight.html>
- Helbig, N., Cresswell, A.M., Burke, G.B. and Luna-Reyes, L. (2012). The Dynamics of Opening Government Data: A White Paper. Centre for Technology in Government, State University of New York, Albany. Available at: <http://www.ctg.albany.edu/publications/reports/opendata/opendata.pdf>
- Huijboom, N. and Van der Broek, T. (2011). “Open data: an international comparison of strategies”, European Journal of ePractice, N. 12, March/April. Available at: http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf

- Kitchin R. (2013). Four critiques of open data initiatives. Seminar: Open data and evidence informed decision making. The Programmable City Project. Available at: <http://www.nuim.ie/progcity/2013/11/four-critiques-of-open-data-initiatives/>
- Immonen, A., Palviainen, M., Ovaska, E.. (2014) Towards open data based business: Survey on usage of open data in digital services. *International Journal of Research in Business and Technology*, 4, Jan. 2014.
- Janssen, K. (2012). "Open government data: right to information 2.0 or its rollback version?" ICRI Working Paper 8/2012, Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2152566
- Janssen, M., & Zuiderwijk, A. (2014). "Infomediary Business Models for Connecting Open Data Providers and Users". *Social Science Computer Review*, 32(5)
- Lee, D., Cyganiak, R., Decker, S. (2014). *Open Data Ireland: Best Practice Handbook*, NUI Galway.
- Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S., & Almasi Doshi, E. (2013). *Open data: Unlocking innovation and performance with liquid information*, McKinsey
- Ohm, P. (2009). "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization" *UCLA Law Review*, Vol. 57, p. 1701, 2010, Available at: <http://ssrn.com/abstract=1450006>
- Pollock, R. (2006). The value of the public domain. *IPPR*. Available at: <http://www.ippr.org/publication/55/1526/the-value-of-the-public-domain>
- Shakespeare, S. (2013). *Shakespeare Review: An independent review of public sector information*. London: Department for Businesses, Innovation & Skills, May. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/198752/13-744-shakespeare-review-of-public-sector-information.pdf
- Yiu, C. (2012). *A right to data: Fulfilling the promise of open public data in the UK*. Policy Exchange Research Note. Available at: <http://www.policyexchange.org.uk/publications/category/item/a-right-to-data-fulfilling-the-promise-of-open-public-data-in-the-uk>
- World Bank (2014). *Open data for economic growth*. Available at: <http://www.worldbank.org/content/dam/Worldbank/document/Open-Data-for-Economic-Growth.pdf>
- Vickery, G. (2011) *Review of recent studies on PSI re-use and related market developments*. Paris: European Commission.
- Zuiderwijk, A., Janssen, M. and Parnia, A. (2013) "The complementarity of open data infrastructures: An analysis of functionalities", *Proc. 14th Annual International Conference on Digital Government Research*, 2013.