

PERSONAL VERSION

This is a so-called personal version (author's manuscript as accepted for publishing after the review process but prior to final layout and copyediting) of the article: Lindman, J & Tammisto, Y 2012, 'Definition of Open Data Services in Software Business' in *Proceedings of the third international conference on software business* .

http://link.springer.com/chapter/10.1007%2F978-3-642-30746-1_28

This version is stored in the Institutional Repository of the Hanken School of Economics, DHANKEN. Readers are asked to use the official publication in references.

Definition of Open Data Services in Software Business

Yulia Tammisto

Aalto University School of Economics
P.O. Box 21220
00076 Aalto
+358 44 035 4667
yulia.tammisto@aalto.fi

Juho Lindman

Hanken School of Economics
P.O. Box 479
00101 Helsinki, Finland
+358 40 353 8091
juho.lindman@hanken.fi

Abstract. New software businesses are extending the software industry via private enterprises that build new and innovative services on top of Open Data (OD) sets released by government and public bodies. What are the tenable value propositions and income-generating mechanisms for these private enterprises and what are the new opportunities for service development? This paper combines conceptual and empirical investigations of OD definition to clarify the benefits open data sets hold for service development.

Keywords: Open data, open data definition, open data services

1 Introduction

The prevailing wisdom is that new services and novel businesses can be created by opening up the data archives collected by government [3, 4, 9, 13]. Contrary to popular belief, releasing the data may provide value to society rather than limiting the archives to within its organizations. Organizations may create new ways to capture the benefits of this new value in the ecosystem by building new services. This is why also entrepreneurs push for OD policy initiatives related to government data sets. Opening of data for public use also provides a variety of new market opportunities for start-ups and other small companies.

However, it remains unclear what the term OD in this sense actually entails. We position our study in the field of research (see for example, [16]) that recognizes similar confusion as related to any new ICT innovations (some previous examples [16] of other novel technologies including case tools, intranet, and open source [12]). Furthermore, the confusion does not have to be considered only negative. Ambiguity of the exact meaning of OD may help a movement, a new process, or a tool to grow a

usage base. This diffusion of innovation happens when consultants and business press are charting the business interest and potential, the demand in their customer organizations, and also, at the same time, defining the legal and commercial implications of the wide range diffusion of the said phenomenon [16]. In short, both constructing meaning and building legitimate use for the new ICT innovation characterized by the term ‘Open Data’ in the industry. However, such ambiguity may also hinder the adoption by confusing some of the actors who might benefit from it, especially if the legal and commercial implications of opening the data remain unclear.

To summarize, we investigate a gap in the research literature related to defining the term OD in a way that would be both informed by relevant research and take into account how the term is currently applied in business. We address this issue by posing two research questions: one about the meaning of the term and the other about the arguments given to its benefits. The first question is: *What is open data?* The second question is: *What are the legitimizations for data openness?*

Our data set includes both small and large companies operating in the Finnish software sector to simplify the legal environment, but we posit that similar characteristics of the OD definition are likely in other legal contexts. We also note that there are political issues related to the OD definition, but we limit them outside the scope of this paper and focus on service development in commercial context.

2 Open data service as software business

According to the Open Knowledge Foundation [13], data can be called open if it can be, “freely used, re-used and redistributed by anyone without legal, technical or social restrictions.” How the data is published determines its potential for re-use (for example, how well-structured the data is [1]). To improve data’s usability, develop a service on the top of it one needs to convert it into open and re-usable format. This conversion requires resources and involves a number of technical challenges [2].

Latif et. al [10] provide one classification role in linked data publication that can be applied to both corporate entities and non-corporate actors, i.e.: persons, enterprises, associations, and research institutes. The adapted version of this classification and brief descriptions of the roles is as follows:

1. Raw Data Provider (or Data Provider) possesses and provides any kind of data;
2. Linked Data Provider (or Data Service Provider) possesses the expertise to convert the raw data into linked data machine-readable format;
3. Data Application Provider (or Application Developer) possesses the expertise to develop applications, visualizations, and mash-ups – all kinds of human-readable outputs, on top of data and linked open data.
4. End Users are persons who consume the data in human-readable format, not raw or unstructured data.

How to build a solid business model [6]? Latif et al. [10] offers a starting point for charting the roles of open data value chain and the service providers’ business models [15]. Building service-based business models on top of public good includes

challenges [18, 19]. Open source [5] and innovation studies [11] identify different processes related to inbound acquisition and outbound data sharing/data publication.

The processes identified above describe the movement of resources (such as data) from an internal organizational environment to external and vice versa. Organizations can apply these approaches to generate and capture value by offering their internal assets externally, or by making use of external assets internally. Moreover, different third party organizations can build their businesses by helping to implement these processes.

3 Empirical analysis

In order to answer our research questions about the meaning and benefits offered by Open Data, we chose several OD service providers and their customers for explorative interviews. The interviews were semi-structured and centered around the phenomenon of OD and elements of business models related to OD. Interviewees shared their perception of OD from their own business perspectives. To conduct the interviews and perform the analysis we applied an interpretive approach, as described by Klein and Myers [8].

Most of the respondents are from Finland and work for companies that build services on top of OD (Service Providers), or are the employees of the client companies of OD service providers (Data Providers). We conducted a total of 16 interviews including 5 sessions with small data service providers, 5 sessions with large data service providers, 5 sessions with a large data provider, and 1 session with a small data provider. Among interviewees there are CEOs, project managers, consultants and developers who deal with OD services in their companies.

4 Findings and discussion

Coming back to the research questions: *What is open data?* and *What are the legitimizations for data openness?* our research indicates that the views on openness as well as views on the data varied between respondents from the different organizations. Interviewees also expressed a variety of opinions on how OD could be applied in their organizations. This points out that there is indeed a certain tension between policy documents surrounding OD and the empirical realities in the companies that actually build the services.

Talking about data openness some interviewees referred to externally open data – freely available for everyone. Others referred to internally open data – data that is only accessible by employees of an organization that provided this data. Another characteristic of OD that was frequently mentioned by interviewees was the technical format of open datasets. A number of interviewees distinguished several data formats or stages of data development. These formats correspond to the classification of roles in linked data development processes.

As expected, interviewees expressed distinctive opinions on what OD is technologically, legally etc. The two alternate opinions to this question, as expressed

by the interviewees are: (1) data is open when it is published publicly (over the internet), and (2) data is open when it is distributed freely within one organization or a network of organizations (intranet or extranet). Results show a clear transition: the archives that were previously unused or closed within an organization or certain part of the organization are now being opened either to the public or to some particular networks, communities. In both situations, whether the data is available for everyone or just for several units of the same organization, the interviewees referred to it as OD. Moreover, many interviewees pointed out that the term OD is more about the process of application than about the opening of the data set itself.

OD can bring benefits when the data is opened and applied internally (within an organization or a network of organizations): *“It might be surprisingly big that an organization that opens its data might develop an internal tool, because the data from different departments, units will be available for use, then sharing of the data and collaboration between departments would become much easier. When the data is available it provides inter-operability between the departmental systems and it has a potential to change dramatically the way organizations work today.”*

This also applies when data is opened externally (publicly): *“I think the added value comes from having more clever people look at it [data]. They [data providers] don’t know how to deal with their data. In their case the added value comes from outsiders being more apt at doing that.”*

During the analysis phase, we compared the ways of using data that is open internally to data that is open externally. As a result of this comparison we found substantial similarities between the ways of using internally and externally open data and consequently found similarities between corresponding aims of using OD. For example, the following aims of using OD to increase visibility of performance and assets, increase transparency, change organizational structures, and express organizational identity were similar as they involved improving the communication within an organization (in the case of internally OD) or society (in the case of externally OD). Table 1 contains an outcome of the comparison of why and how internally and externally open data is used.

Table 1. Compilation of interviewees’ opinions on why to use internally and externally OD.

Aims of using Internally Open Data	Aims of using Externally Open Data	Similarity
Increase visibility of performance and of assets	Increase transparency	Improve communication
Change organizational structures	Express organizational identity	
Change public sector	Benefit from combination of many datasets	Improve decision-making
Commercial use	Enable external contribution to service development and provision	Develop and provision new services
	Boost the economy	Create economic value

Accordingly, (on an organizational or internal level), an organization opens data internally to improve internal communication and decision-making, to support internal service development internally and expects monetary return on these activities. On a societal or external level, a national government publishes its datasets to improve communication between different organizations within society and to provide entrepreneurial minds opportunities to develop new services and grow the economy. Consequently, OD turns out to be a process that can appear on different levels of organization or society.

Interviewees distinguished different OD formats. OD service providers referred to three stages of data development, namely: raw data, linked data, and applications built on top of the data. All the interviewed data providers recognized two data formats: 1) raw data – “the data as it is” or as it was collected and 2) various data applications (services built on the top of data, data visualizations, and mash-ups). Some data providers also identified linked OD format, or, “*semantically enriched open data*.” Each data provider decides which format to choose. The framework that reflects how interviewed data service providers and their clients perceive different data formats is presented in Table 2.

Table 2. Interviewees’ views on OD technical formats.

Format	Readability	Examples
Raw data	Neither machine-, nor human-readable data	CSV files that are contained in data providers’ databases.
Linked open data	Machine-readable data	A rather technical concept that consists of 5 rankings [1] - requirements that should be fulfilled in order to call the data – linked open data.
Applications build on the top of data	Human-readable data	Applications, visualizations, mash-ups that are easier to comprehend than large rows and columns of data.

Based on empirical analysis OD service providers differentiate OD according to the level of its technical development. In terms of OD usability or applicability, “Raw data” is the hardest to apply. It takes a professional to first scrape it and convert into a machine- or human-readable representation of the original data set. “Linked open data” is easier to handle from a professional point of view, but still is not understandable for an inexperienced user. The last data format, “applications” is basically a user interface that allows regular users to clearly comprehend the content of the dataset that it was built upon.

5 Conclusion

There are several perceptions on what the phenomenon of OD implies. The classifications of these perceptions that we found are following: the degree of data openness (internal vs. external) and OD technical format (raw data, machine-readable data, and human-readable data).

Some people call the data that is available internally within one organization open. Others argue that only the data that is available for everyone externally can be called open. Both name a number of benefits that their version of OD that provide. In case of internally open data those benefits apply to an organization where the data was opened, such as improving internal communication and organizational performance. When talking about data open for everyone or externally OD people refer to national scale benefits like increasing transparency and boosting economic development (see 2). Consequently, the benefits of OD are expected to occur within the system where the data was opened whether it is an organization or a nation.

The entities and different actors can open their data externally or internally to pursue various goals and improve organizational processes. Some organizations may adopt a completely open model by making their data open externally - for example, non-profit organizations that collect statistics on particular area like government agencies and cultural institutions. Others can consider opening their data internally, for example, to increase interoperability between different departments or organizational network members. Also, both approaches may be applied simultaneously, when some parts of data are released only for internal use and others are available to everyone.

OD technical format is related to data usability. The more technically developed the data, the easier it is for the not-versed user to exploit it. Consequently, raw data is the hardest format to use and applications are the easiest way to get the information out of an original dataset. From this perspective we logically assume that the OD's capability to achieve expected benefits of its usage (listed in Table 1) correlates with OD format. By choosing and maintaining an appropriate OD format a data provider encourages OD usage thus increases its chances to improve communication, decision-making, develop and provide more services, and create economic value either at organizational or national level.

References

1. Berners-Lee, T.: Linked Data Design Issues (July 2006), <http://www.w3.org/DesignIssues/LinkedData.html>
2. Bizer, C., Heath, T., Berners-Lee, T. Linked Data – the Story So Far, to appear in: Heath, T., Hepp, M., and Bizer, C. (eds.). Special Issue on Linked Data, International Journal on Semantic Web and Information Systems, (IJSWIS). <http://linkeddata.org/docs/ijswis-special-issue>.
3. Data.Gov: Data <http://www.data.gov/about>
4. European Commission: Open data An engine for innovation, growth and transparent governance. Brussels, COM (2011) 882, Brussels, Belgium, (2011).

5. Fink, M. *Business and Economics of Linux and Open Source*. Prentice Hall, New Jersey (2002).
6. Halb, W., Stocker, A., Mayer, H., Mülner, H., Ademi, I.: Towards a commercial adoption of linked open data for online content providers. In *Proceedings of I-SEMANTICS the 6th International Conference on Semantic Systems, Graz, Austria, (2010)*.
7. Huijboom, N., Van den Broek, T.: Open Data: an International Comparison of Strategies. *European Journal of ePractice*, 12, (March/April 2011)
http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf
8. Klein, H. K. and Myers, M. D.: A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems, *MIS Quarterly*, 23(1), 67-94 (1999).
9. Kuk, G., Davies, T.: The Roles of Agency and Artifacts in Assembling Open Data Complementarities. In *Proceedings of Thirty Second International Conference on Information Systems, Shanghai, China, (2011)*.
10. Latif, A., Saeed, A.U., Hoefler, P., Stocker, A., Wagner, C.: The Linked Data Value Chain: A Light Weight Model for Business Engineers. In *Proceedings of I-SEMANTICS '09 International Conference on Semantic Systems, 568-575, Graz, Austria (2009)*.
11. Lichtenthaler, U. 2011 Open Innovation: Past Research, Current Debates, and Future Directions. *The Academy of Management Perspectives*, 25(1), 75-93 (2011).
12. Lindman, J., Rossi, M., Marttiin, P. Open Source Technology Changes Intra-Organizational Systems Development – A Tale of Two Companies. *Proceedings of European Conference of Information Systems, Pretoria, South Africa, 7-9.6, (2010)*.
13. Open Knowledge Foundation: <http://okfn.org/>; <http://www.opendefinition.org/okd/>
14. Somus (Social media for citizens and public sector collaboration) project – final report (January 2011), <http://www.vtt.fi/inf/pdf/publications/2011/P755.pdf>.
15. Reference is withdrawn for reviewing purposes
16. Swanson, B., Ramiller, N.: The Organizing Vision in Information Systems Innovation. *Organization Science*, 8, 5, 458-474, (1997).
17. Van de Ven, A. H., Johnson, P. E.: Knowledge for theory and practice. *Academy of Management Review*, 31, 802– 821, (2006).
18. Vargo, S. L., Lush, R. F.: Evolving a services dominant logic. *Journal of Marketing*, 68, 1-17 (2004).
19. Von Hippel, E., Von Krogh, G.: Open Source Software and the 'Private-Collective' Innovation Model: Issues for Organization Science. *Organization Science*, 14, 2, March-April (2003).