

Two birds with one stone. An economically viable solution for linked open data platforms

Riccardo Bonazzi

University of Applied Sciences Western Switzerland (HES-SO), Switzerland

Riccardo.bonazzi@hevs.ch

Zhan Liu

University of Applied Sciences Western Switzerland (HES-SO), Switzerland

Zhan.liu@hevs.ch

Abstract

Linked open data has been described by scholars as the logic evolution and the main benefit of open data. Nonetheless, the cost of data integration and platform management cannot be simply covered by selling the data, which is freely available by definition. Moreover, existing classifications of business models for linked open data platforms are rather descriptive instead of being prescriptive, and they do not take into account the notion of economic sustainability. Hence, this paper extends the existing literature in order to understand how to define a value proposition and a revenue model to increase the adoption of linked open data. We have developed a simple typology and we have identified a new revenue model for a linked open data platform, which is currently being tested.

Keywords: open data, linked open data, business model, revenue model, value proposition, open innovation

1 Introduction

This paper describes the first phase of an on-going project, and it is addressed to managers and scholars seeking for new ways to assure economic resources to develop and maintain a linked open dataset. The notion of *linked open data* (hereinafter referred to as LOD) comes from two concepts: (1) open data and (2) linked data. *Open data* is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike (Open Knowledge Foundation, 2012). *Linked data* describes a method of publishing structured data, upon standard Web

technologies, in a way that can be read automatically by computers so that it can be interlinked and become more useful (Bizer, Heath, & Berners-Lee, 2009).

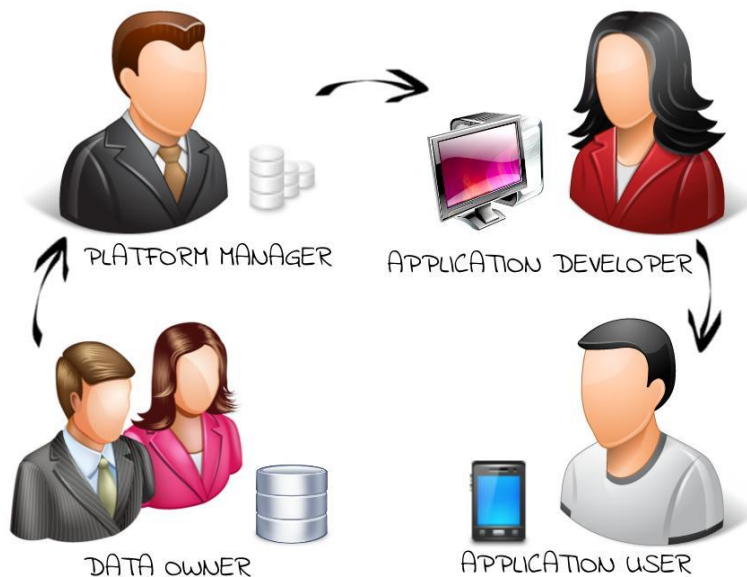


Figure 1: Stakeholders of a linked open data platform

Driven by the success of Linked Data (LD), LD related business models have been discussed in the literature. Hence, we refer to a set of different provider's roles proposed by (Latif, Saeed, Hoefler, Stocker, & Wagner, 2009) to support the conceptualization of successful business cases: raw data provider, linked data provider and linked data application provider. Based on these roles, (Tammisto & Lindman, 2011) claim that the main benefits of open data related activities is LOD transformation, consulting, and the application development by using these data.

Figure 1 represents four stakeholders of a LOD platform: (a) the user of the application, who sometimes is willing to pay for contextual information obtained by aggregated data; (b) the application developer, who looks for a large amount of consistent data to exploit, in order to increase the usefulness of the application; (c) the owner of open data, who hopes to increase the usage of the dataset; (d) the manager of the LOD platform, who has to offer a service that is outperforming traditional data services while finding new ways to cover the cost of data integration and platform management.

Accordingly, we refer to the notion of business model, as defined by (Osterwalder & Pigneur, 2010), and we focus on two key elements: (1) the revenue model, which is the description of how a business monetizes its services and (2) the value proposition, which is a promise of value to be delivered to the customer.

Therefore, our research question is: **how to define a revenue model to increase the adoption of linked open data?**

The rest of the paper proceeds as it follows. Section 2 briefly illustrates the existing literature, which addresses our research question. Section 3 illustrates the methodology used to address the gap in the literature. Section 4 illustrates the theoretical model obtained. Section 5 briefly illustrates the evaluation procedure, which is currently ongoing. Section 6 summarizes the key elements of the paper and illustrates further directions of investigation.

2 Literature review

In order to obtain a descriptive review, we followed the argumentative strategy suggested by (Rowe, 2014). Accordingly, we used the keywords "linked open data" "value proposition" "revenue model" on Google scholar and we selected academic articles that were available online on January 2015, that offered insights about revenue models for linked open data platforms, possibility under the shape of classification or typologies. We initially obtained nine papers: three were dismissed since out of topic, one was dismissed since it was not containing any sort of classification; one was dismissed since it was not an academic paper and one was dismissed since it was not available. Of the remaining three articles, one focuses on linked data, whereas the other two on open data. On the one hand, (Vafopoulos, 2011) proposes eleven distinct business model categories for linked data. On the other hand, (Lindman, Kinnari, & Rossi, 2014) induce from a set of case studies the open data value network structure and propose five business model for the data network profiles, whereas, (Zeleti, Ojo, & Curry, 2014) merge emerging value disciplines for open data businesses into five major categories: (1) freemium; (2) premium; (3) cost saving;(4) support primary business;(5) razor and blade. Table 1 compares the elements used by (Zeleti et al., 2014) with the elements of the other two classifications. In the following sections we intend to extend these models by introducing the notion of economic performance of the LOD platform.

Table 1: Mapping among elements of the three classifications

(Zeleti et al., 2014)	(Vafopoulos, 2011)	(Lindman et al., 2014)
Free as Branded-Advertising Freemium Dual-Licensing Charging for Changes Open source	Advertising Freemium Affiliate program Subscription (timely access) -	User experience provider - - - -
Sponsorship Demand-Oriented Platform Support and Services Supply- Oriented Platform White-Label Development Premium	Sponsorship Marketplace Subscription (Convenient Access) Subscription(Archival access) Customized service -	User experience provider Extract and trasform Data analyzer Extract and trasform User experience provider -
Quality through Participation Cost Avoidance	Public service; community service	Support and Consultation Commercial OD publisher
Support Primary Business	Multi-sided platform; Traffic Generation /SEO	-
Infrastructural Razor and Blades	Subscription (full/on demand/block access)	-

3 Methodology

In this section, we describe how we use design science to obtain a theory under the shape of a typology. According to (Hevner, March, Park, & Ram, 2004), design science addresses wicked problems and seeks out usefulness, rather than truth. According to (Doty & Glick, 1994), typologies are conceptually derived, interrelated sets of ideal types that meet three criteria: (1) they contain explicitly de-fined constructs that can be quantified, (2) relationships among the constructs are articulated, and (3) predictions associated with the typology are testable and subject to disconfirmation. Constructed in this way, a typology can account for multiple causal relationships in a given setting, and it can reduce complexity to manageable levels both conceptually and methodologically.

4 Our typology

In this section we present (1) the constructs of our typology, (2) the relationships among the constructs, and (3) the predictions associated with the typology.



Figure 2: Money flows among stakeholders of the linked open data platform

4.1 Our first order constructs

Figure 2 represents the money flows among the four stakeholders already introduced in figure 1. In this study we focus on recurring transactions. The profit of the LOD platform manager depends on three flows: (a) the money paid by the application developer, who uses the data, which can be nothing or any amount above zero. Therefore, **Price** = [Low; High]; (b) the money that the platform has to pay to the data owner, which can be nothing or any amount above zero. Therefore, **Cost** = [Low; High]; (c) the money paid by a third-party, who takes indirect advantage from the platform, which can be nothing or any amount above zero. Therefore, **Support** = [Low; High].

In this article we do not take into consideration: (d) the money flows between the application user and the application provider and (e) the money flow, which goes between the data owner, if it is a public institution, and the application user, intended as citizen.

4.2 Relationships among our first order constructs

Following what stated in the previous paragraph, we obtain the following relationship.

Equation 1: Profit_{Platform Manager} = Price_{Application developer} – Cost_{Data Owner} + Support_{Third-Party}

Moreover, we assume that the amount of money obtained by third-party is less than the amount of money obtained by application developers.

The resulting typology and its associated predictions

Table 2 illustrates the resulting set of ideal types, which we named by using the five categories of (Zeleti et al., 2014). Nonetheless, since we obtained eight ideal types, we had to split some categories into two sub-components.

The first ideal type offers community services (Vafopoulos, 2011), which are not meant to be profitable. Dbpedia.org is an example of service offering Wikipedia as LOD.

The second ideal type refers to public services (Vafopoulos, 2011), which are supported by public institutions. ItoWorld.com offers public LOD.

The third ideal type is used to increase traffic towards other services. Google public data explorer is meant to increase the overall traffic.

The fourth ideal type is used to promote data owners. Musicbrainz.org offers linked music data to promote artists and it is supported by Google, which uses the dataset to improve its search results.

The fifth ideal type refers to a LOD platform offering paid services beside its free datasets. According to their website, Mapbox.com offers custom online maps for major websites such as Foursquare, Pinterest, Evernote and the Financial Times.

The sixth ideal type does not appear as such in the existing literature, and it will be described in the next section.

The seventh ideal type refers to a LOD platform offering high quality data at a price, which is given back to the data owner (minus a transaction fee). The Azure data market allows Microsoft to obtain some money while increasing usage of the Azure platform.

The eighth ideal type offers paid dataset in exchange for money from users for complementary services, while receiving supporting money from the provider of the complementary services. The recent acquisition of Datamarket by Olik (Park, 2014) can be seen as an example of this type of platform.

Table 2: Our typology

(a) Price data user	(b) Cost data owner	(c) Third-party Support	Profit for platform owner	Link to (Zeleti et al. 2014)	Exemple
Low	Low	Low	Low	Cost saving (cost avoidance)	Dpedia for Wikipedia
Low	Low	High	Medium	Cost saving (quality through participation)	ItoWorld
Low	High	Low	None	Support primary business	Google public data
Low	High	High	Medium	Premium (Sponsorship)	Musicbrainz
High	Low	Low	High	Freemium (Charging for Changes)	Mapbox
High	Low	High	High	??	??
High	High	Low	Medium	Premium (demand-oriented platform)	MS Azure data marketplace
High	High	High	High	Razor and blade	Datamarket/Qlik

5 Evaluation

According to (Snow & Ketchen, 2014) most typologies fail to be assessed by using the five guidelines offered by (Doty & Glick, 1994). Therefore, we explain in details how we have addressed each guideline.

5.1 Typological theorists should make explicit their grand theoretical assertion(s).

We refer to (Karahanna, Straub, & Chervany, 1999), who use the theory of diffusion of innovation to show that the perception of usefulness increases the chances of acquisition and retention of new users. Indeed, (Zeleti et al., 2014) have claimed that the freemium and the premium value propositions, which increase usefulness. Therefore, we claim that:

Proposition 1: Over time, the diffusion of the ideal types *premium* and *freemium* will be greater than the other ideal types.

5.2 Typologies must define completely the set of ideal types.

We have defined the full set of ideal types, and we have discussed the soundness of each result obtained. For one ideal type (the sixth) we did not find a correspondence in the existing literature. Indeed, it describes a service that relies on paid services based on freely available data, while obtaining sponsorship from third-party. Therefore, we named it 2b1s (“two birds with one stone”), and we speculate that it could refer to a LOD platform that offers high quality data at application owners, while selling to third-party the usage statistics of its datasets. Indeed, one could expect that, data owners belonging to public institutions would be willing to know how to fine tune their datasets to increase usage.

5.3 Typologies must provide complete descriptions of each ideal type using the same set of dimensions.

We have presented our ideal types and we have done two actions: (a) we gave an example for those that are currently implemented and (b) we suggested a business case for those that are theoretically sound.

5.4 Typological theories should explicitly state the assumptions about the theoretical importance of each construct used to describe the ideal types.

We have derived three first order constructs by extending the roles of (Latif et al., 2009). For sake of simplicity we have simplified our set of first order constructs in order to obtain the lowest set of ideal types that answers our research question.

5.5 Typological theories must be tested with conceptual and analytical models that are consistent with the theory.

The testing of our typology is currently on-going. We are collecting experts' opinions to validate our theoretical model and we have been collecting second-hand data about LD, OD and LOD platforms to falsify our testable proposition.

In parallel, we have partnered with a public institution and we have developed a LOD platform, which will follow the guidelines of the sixth ideal types (2b1s), to verify if it is feasible in practice.

6 Conclusions

The purpose of this paper was to extend the existing literature to understand how to define a revenue model to increase the adoption of linked open data. We have developed a simple typology and we have identified a new revenue model for linked open data, which is currently being tested.

We recognize that the major limitation of our paper is the lack of first-hand data. Nonetheless, we believe that this study already offers a major contribution in the field of business model for LOD by combining existing classifications for open data and linked data into a new prescriptive model that introduces the notion of business performance in the equation.

Acknowledgements : This project was supported by the HES-SO Valais-Wallis under grant number 40160 (OverLOD Surfer).

7 References

- Bizer, C., Heath, T., & Berners-Lee, T. (2009). *Linked data-the story so far*.
- Doty, D. H., & Glick, W. H. (1994). Typologies as a unique form of theory building: Toward improved understanding and modeling. *Academy of Management Review*, 19(2), 230–251.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, 28(1), 75–105.
- Karahanna, E., Straub, D. W., & Chervany, N. L. (1999). Information Technology Adoption Across Time: A Cross-Sectional Comparison of Pre-Adoption and Post-Adoption Beliefs. *MIS Quarterly*, 23(2), 183. doi:10.2307/249751
- Latif, A., Saeed, A. U., Hoefler, P., Stocker, A., & Wagner, C. (2009). The Linked Data Value Chain: A Lightweight Model for Business Engineers. In *I-SEMANTICS* (pp. 568–575). Citeseer.
- Lindman, J., Kinnari, T., & Rossi, M. (2014). Industrial open data: Case studies of early open data entrepreneurs. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on* (pp. 739–748). IEEE.
- Open Knowledge Foundation. (2012). *Open Data Handbook Documentation. Release 1.0.0*. Retrieved from <http://opendatahandbook.org/>
- Osterwalder, A., & Pigneur, Y. (2010). *Business Model Generation: A Handbook For Visionaries, Game Changers, And Challengers Author: Alexander Osterwalder, Yves.* Wiley.
- Park, H. (2014, November 4). Qlik Acquires DataMarket | Blue Hill Research. Retrieved from <http://bluehillresearch.com/qlik-acquires-datamarket/>
- Rowe, F. (2014). What literature review is not: diversity, boundaries and recommendations. *European Journal of Information Systems*, 23(3), 241–255.
- Snow, C. C., & Ketchen, D. J. (2014). Typology-driven theorizing: A response to Delbridge and Fiss. *Academy of Management Review*, 39(2), 231–233.
- Tammisto, Y., & Lindman, J. (2011). Open data business models. In *The 34th Information Systems Seminar in Scandinavia, Turku, Finland*.
- Vafopoulos, M. (2011). A framework for linked data business models. In *Informatics (PCI), 2011 15th Panhellenic Conference on* (pp. 95–99). IEEE.
- Zeleti, F. A., Ojo, A., & Curry, E. (2014). Emerging business models for the open data industry: characterization and analysis. In *Proceedings of the 15th Annual International Conference on Digital Government Research* (pp. 215–226). ACM.