Service innovation: the hidden value of open data

Muriel Foulonneau, Slim Turki - Luxembourg Institute of Science and Technology

1. Introduction

The development of a data driven economy has been a major orientation of economic policies over the past few years based on (i) the wider availability of data promoted in particular by the Open Data movement (Chui et al., 2013) and (ii) the development of dedicated tools to support heterogeneous data and data in large quantities (Big data). Reports anticipate the creation of enormous amounts of economic activity and growth opportunities (e.g. Buchholtz et al., 2014). However the promise of the data-driven economy lies to a large extent in the development of new services (Deloitte, 2011). The return on investment of open data policies for instance should be evaluated from the services created based on open data sets. Open data promoters couple more and more open data initiatives with actions dedicated to the promotion of the datasets for the creation of new services (Foulonneau et al., 2014). Nevertheless the results in terms of services created remain below the expectations of open data promoters. Indeed most services created are not sustainable and / or do not use the variety of datasets. They are to a wide extent relying on a limited number of very popular datasets. In order to make the promise of the data-driven economy a reality, it is therefore necessary to increase reuse and value extracted by services from data. Our hypothesis is that service innovation approaches can help understand the mechanisms that drive the creation of services. We therefore propose to analyse the roles that the data can have in the design of services based on a theoretical framework of service innovation.

2. The service design process

Before it becomes an innovation, one service idea has to gain its spurs: Does it have a market, technological or research potential? Is it worth investing in its development? The lifecycle of an idea is composed of 3 phases of their life: (i) birth, (ii) maturation and (iii) evaluation. These phases correspond to the steps of the iterative process of service design (Vidou, 2013), i.e., idea generation, maturation, and concept evaluation (**Figure 1**).

Figure 1: Service design process

The (i) ideation phase corresponds to the birth of the idea. It can be born spontaneously or from the systematic exploration of various fields of innovation. It can be triggered by a stimulus at the occasion of a call for ideas or an ideation contest. The (ii) maturation phase consists in exploring the idea related issues, validate its options or not, make some investigation which other technologies and services are out there

already. When the idea has reached a level of maturity, the potential of the idea can be assessed to decide to invest in its development – this is the (iii) evaluation phase.

During the idea lifecycle, some aspects of the idea should be addressed in order to avoid missing elements: synopsis, context, target, resources, service system, innovativeness and sustainability. The (i) synopsis is the minimum level of description of the service, the summary of the concept. The (ii) context dimension describes the context in which the service is delivered (time, space technological components, regulatory context: norms, standards...). The (iii) target describes the customers of the service and the reason why they would buy it. The (iv) resource component describes the type of resources required to deliver the service (human, technological resources, process & organizational resources, norms and standards, partners, financial resources…). The (v) service system component describes the way in which resources are combined to deliver the service to the target in the context (key activities, key partners as stakeholders). The (vi) innovativeness and sustainability highlights the innovative aspects of the service system through its ingredients and the expected economical, societal & environmental impacts. For each dimension, elements can be at different levels of maturity to support the progressive maturation of the service.

In innovation, there are multiple opportunities for the creation of innovative services. A demographic change, such as the increase of the elderly population can lead to opportunities for the creation of services for retired people. A technology such as indoor GPS can lead to the creation of services based on the new possibilities offered by the availability of that technology. Just like a technology and a demographic change, the availability of data can create opportunities for the creation of new services.

3. The role of the dataset in the services

As summarized in **Figure 2**, data can therefore play different roles in the service design process: 1) The service is based on data, 2) the service uses data as a resource, and 3) The service is validated or enriched with data but the data is not directly used or is not directly visible in the service.

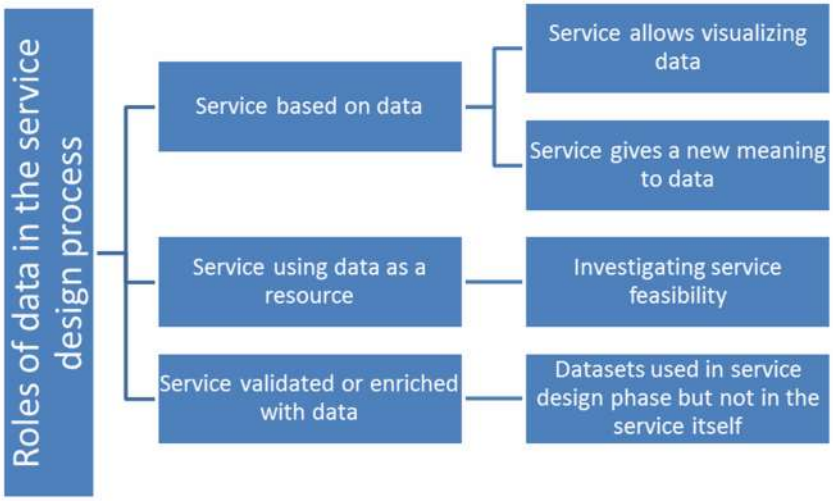

Figure 2: Roles of data in the service design process

3.1. Service based on data

When the data is used as impulse to the service ideation process, it represents the core of the service concept. The objective of the ideation process is to determine with one or more datasets which services

could be designed based on them. The data producer should wonder who currently uses the data; who else could be interested by these data; and if there is any combination of these data with other data which could be of interest to a stakeholder.

The service can for instance allow visualizing the data. The Website *publicspending.net* for instance allows citizens to view budgetary data. The *nosdeputes.net* service shows the activity of French members of parliaments. The objective of these services is to provide a new didactic access point to the data and thus enable transparency. The *handimap.org* service in the French cities Rennes and Montpellier show paths through the cities for disabled citizens. By combining public and private data sources on the creation of companies in Europe, the European SPOCS project has shed a new light on the comparative ease of creating a new business in European countries (Foulonneau et al., 2013a). The Narrative science company generates texts from data so as to make it more user friendly (http://goo.gl/Cmwzf2). It is for instance possible to generate sport news articles from the raw results of local competitions which would not be covered by newspapers.

Finally, it is possible to give a new meaning to the data. *Google ngrams* (https://books.google.com/ngrams) for instance benefits from the *Google* book digitization program. In this context, Google has digitized millions of books from various countries at various times. It then applies an Optical Character Recognition process so as to enable full text search functionalities in the books. By combining the bibliographic data (including the date of publication) and the individual words used in each book, *Google ngrams* allows visualizing the evolution of the use of particular words over time (Figure 3).

Figure 3: Google ngram viewer representation of the evolution of the words "republic" and "democracy" between 1800 and 2000

At the maturation phase, it is then necessary to analyse the characteristics of the datasets and their impact on the feasibility of the service. This includes the update frequency, the data quality, such as its reliability, completeness etc. (see Stvilia, 2007), the data source, its maintenance processes, intellectual property rights attached to the data and its conditions of use, its cost, its accessibility, including its technical accessibility (e.g., API, data dump …) as well as its formats (e.g., RDF/XML, JSON, spread sheet), its interoperability with other datasets, typically to mix it with third party datasets, and its documentation including the documentation of its underlying semantic model in order to adequately interpret and use it.

Indeed the update frequency and the maintenance processes attached to the data may be critical and impact the quality of the service. This can jeopardize its viability. In addition, specific processes may have to

be in place in order to ensure the collection of updated data for instance. The cost has a direct impact on the business model of the service, while the lack of interoperability with other datasets of interest may lead to additional costs to make aggregate them in a single service.

In this type of services, the data represents the main resource or one of the main resources. Its characteristics should therefore be analysed in detail at the occasion of the maturation phase of the service design.

3.2. Services with data as resources

In other types of services, the concept is defined at the ideation phase without any specific relation with the datasets. However when investigating the feasibility of the service at the maturation phase, the datasets have to be taken into consideration as necessary resources. Location and traffic data for instance is not the core of a delivery service concept. However they are resources that will help design the service.

Data enrichment can use for instance Wikipedia for the translation of concepts or a dataset of postcodes to automatically fill the city in an address form. In a service designed to assess the reading difficulty of educational texts (Foulonneau et al., 2013b) we used a lexical base (WordNet) to identify the relation between words in contiguous paragraphs so as to define breaks in the texts. The final system does not display the external data. It may even work without the external dataset (with simple sentence similarity metrics based on exact word matching for instance). However the external dataset increases the quality of the final system.

As for the services based on data, the datasets should be analysed in terms of intellectual property rights, reuse conditions, maintenance processes, update frequency, technical accessibility, formats, data quality, costs, interoperability with other datasets, and documentation in particular.

A major challenge for open datasets for instance is that in many cases the reuse conditions of the datasets are not clear (Martin et al., 2013a). When a service has to reuse multiple datasets, the multiplicity of reuse conditions may increase the barrier for the reusers. The *Europeana* digital library has therefore promoted the CC0 licence to ensure that all datasets would be reusable without limitation.

3.3. Service validation

Finally datasets may be used in the design phase of the service but not in the service itself. Indeed, in the concept validation phase it is possible to use external datasets only for testing a service concept or validating data which are already hold by the service designer.

Datasets can be used for validating the data already used in the service. A dataset of postcodes can be used to validate postcodes provided by users in a form. Recommendation systems are often tested against standard datasets, such as MovieLens (http://grouplens.org/datasets/movielens/) which has gathered ratings of users over many movies. Authors of new algorithms can then test their algorithm against the dataset to verify that it can accurately predict the ratings provided by users and compare its performance to the performance of the numerous other algorithms which have been tested against the MovieLens dataset and for which performance has been reported in scientific publications.

Datasets can also be used to validate a business model, through gathering economic indicators, typically from statistical institutes that publish their datasets. Simulation environments which require many datasets

to recreate the context of execution of a service are often used to predict the difficulties or identify the optimal distribution of resources, for instance in traffic related services.

In these cases the external datasets do not appear in the final service. However they play a critical role to increase the quality of the service and ensure its viability.

4. Conclusion

This paper explored the roles that the data can have in the design of services based on a theoretical framework of service innovation. A market has emerged with companies dedicated to supporting access and the design of new services based on data (e.g., Enigma.io, Narrative Science). Even Google offers a visualization service for open data sets (http://www.google.com/publicdata/). More and more data access portal build APIs and mechanisms that lower the barrier for service designers. Dedicated software has been designed such as the Callimachus platform to reuse Linked open Data (http://callimachusproject.org). In addition, many platforms also organize competitions, such as Hackathons (http://goo.gl/cINX7L), display successful services or even co-finance the development of services based on the data (e.g., in Singapore (Calvin, 2013)). Competitions especially often focus on the first type of service for which data are the core resource. However it is important to understand the roles that the data can have in a service and most importantly that they can also help at the maturation and the validation phases of the service design. This offers new opportunities for the reuse of data and suggests a different approach to measuring the impact of opening datasets beyond the mere number of services created. Indeed in the future it would be necessary to ensure that the benefit of opening data can be adequately measured and that datasets reuse is not only a matter of promotion but also a matter of asset value, i.e., the characteristics of the datasets are critical to ensure their effective reuse by service designers.

References

Buchholtz S., Bukowski M., Śniegocki A., (2014). “Big & Open Data in Europe: A growth engine or a missed opportunity?” demosEUROPA & WISE Institute, ISBN: 978-83-925542-1-9, Jan 2014.

Chui M., Farrell D., Van Kuiken S. (2013). “Generating Economic Value through Open Data”. Beyond Transparency: Open Data and the Future of Civic Innovation. October 2013. ISBN-10: 0615889085.

Deloitte Analytics Institute (2011). “Unlocking growth: How open data creates new opportunities for the UK”.

Foulonneau M. et al. (2013). “SPOCS: A semantic interoperability layer to support the implementation of the service directive”. Proc.of ECEG’2013, 13th European Conference on e-Government. June 2013, Como, Italy.

Foulonneau M., Martin S., Turki S. (2014). "How Open Data Are Turned into Services?”, Proc. Of IESS 2014: 31-39, Exploring Services Science. Geneva, Switzerland, February 2014. Springer 2014 Lecture Notes in Business Information Processing ISBN 978-3-319-04809-3.

Foulonneau M., Ras E., Abou Zeid E., Atéchian T. (2013). "Reusing Textual Resources in Educational Assessment: Adding Text Readability Metrics to Learning Metadata”. Dublin Core conference 2013, Lisbon.

Klessmann et al., (2012). “Open Government Data Germany”, Short Version of the Study on Open Government in Germany Commissioned by the Federal Ministry of the Interior. July 2012.

Martin et al., (2013), ”Risk Analysis to Overcome Barriers to Open Data”. Electronic Journal of e-Government Volume 11 Issue 1 2013, (pp348 -359), available at www.ejeg.com

Martin et al., (2013). “Open Data: Barriers, Risks, and Opportunities”. Proc. Of ECEG’2013, 13th European Conference on e-Government, Como, Italy, June 2013. ISBN: 978-1-909507-22-7.

Martin, S., Foulonneau, M., Turki, S. (2013). "1-5 stars: Metadata on the Openness Level of Open Data Sets in Europe”. In Metadata and Semantics Research 2013. Communications in Computer and Information Science, Springer Verlag, Heidelberg.

Stvilia et al., (2007). “A framework for information quality assessment”. Journal of the American Society for Information Science and Technology 58.12 (2007): 1720-1733.

Vidou, G. (2013). “The Service Value Pathway: the 3-6-3 tool”, The XXIV ISPIM Conference, Helsinki, 2013.