Risk Analysis to Overcome Barriers to Open Data

Sébastien Martin[1], Muriel Foulonneau[2], Slim Turki[2] and Madjid Ihadjadene[1]
[1]Université Paris 8, Vincennes-Saint-Denis, France
[2]PRC Henri Tudor, Luxembourg, Luxembourg

Abstract: Despite the development of Open Data platforms, the wider deployment of Open Data still faces significant barriers. It requires identifying the obstacles that have prevented e-government bodies either from implementing an Open Data strategy or from ensuring its sustainability.This paper presents the results of a study carried out between June and November 2012, in which we analyzed three cases of Open Data development through their platforms, in a medium size city (Rennes, France), a large city (Berlin, Germany), and at national level (UK). It aims to draw a clear typology of challenges, risks, limitations, barriers, all terms used by the different stakeholders with diverse meanings and based on different motivations. Indeed the issues and constraints faced by re-users of public data differ from the ones encountered by the public data providers. Through the analysis of the experiences in opening data, we attempt to identify how barriers were overcome and how risks were managed. Beyond passionate debates in favor or against Open Data, we propose to consider the development of an Open Data initiative in terms of risks, contingency actions, and expected opportunities. We therefore present in the next sections the risks to Open Data organized in 7 categories: (1) governance, (2) economic issues, (3) licenses and legal frameworks, (4) data characteristics, (5) metadata, (6) access, and (7) skills.

Keywords: Open Data, open government, e-government

1 Introduction

Open Data has gained a lot of interest in the e-government communities over the last years, leading to the implementation of many initiatives and platforms to publish Open Datasets in such areas as mobility (e.g., bus timetables), security (e.g., crime rates), or economy (e.g., statistics on business creations). Open Data is an essential tool for the dissemination of the open government principles. However its wider deployment requires identifying the obstacles that have prevented e-government bodies either from implementing an Open Data strategy or from ensuring its sustainability. We have therefore carried out a study between June and November 2012, in which we analyzed three cases of Open Data development through their platforms, in a medium size city (Rennes, France[1]), a large city (Berlin, Germany[2]), and at national level (UK[3]). In addition, we have studied the context in which the Open Data movement has been developed across Europe, in particular the type of data that have been opened, and the services that were developed by Open Data re-users.

In this study, we aim to draw a clear typology of challenges, risks, limitations, barriers, all terms used by the different stakeholders with diverse meanings and based on different motivations. Indeed the challenges and constraints faced by re-users of public data differ from the ones encountered by the public data providers. Through the analysis of the experiences in opening data in the UK and in the cities of Rennes in France and Berlin in Germany, we attempt to identify how barriers were overcome and how risks were managed.

Beyond passionate debates in favor or against Open Data, we propose to consider the development of an Open Data initiative in terms of risks, contingency actions, and expected opportunities. We therefore present in the next sections the risks related to Open Data organized in 7 categories:

- governance,
- economic issues,
- licenses and legal frameworks,
- data characteristics,

[1] http://www.data.rennes-metropole.fr/
[2] http://daten.berlin.de/
[3] http://data.gov.uk/

Reference this paper as: Sébastien Martin et al "Risk Analysis to Overcome Barriers to Open Data" *Electronic Journal of e-Government* Volume 11 Issue 1 2013, (pp348 -359), available online at www.ejeg.com

- metadata,
- access, and
- skills.

2 Studied experiences

We analysed both Open Datasets and services created based on those datasets. We took into consideration data catalogues from Rennes in France and Berlin in Germany. We used the conclusions drawn by Fraunhofer (Both, 2012) which supported the creation of the Berlin platform. Analyses of services were only carried out for Rennes and the United Kingdom, since the Berlin initiative was more recent.

Berlin (Germany) is a good example of Open Data at the local level, with the additional advantage of showing the relationship between different administrative scales, the level of the city itself and the region (*land*). Berlin platform has served as a model and prototype for the whole Germany and even beyond since it is integrated in in a European research project. It was also prepared with a prospective study by the Fraunhofer Institute to understand of the data opening process since its early stages. The first datasets are progressively added to the portal, launched in September 2011.

Rennes, the administrative centre of the French region Bretagne, also provides an example of local Open Data. The city led the opening of its data in a broader approach to innovation based on digital technologies. Among the first initiatives in France, this is the first community to open a portal in 2010. It has set up an effective support for reuse with a reuse competition and has a rather dense network of re-users.

The UK portal is open since January 2010 and centralizes data nationally. The British approach has also resulted in the publication of reports and scientific articles. The British government has asked each department to publish its Open Data strategy of opening, each one being inserted into the overall strategy outlined in a White Paper[4]. Data from the United Kingdom have also enabled the creation of a large number of services that can be analysed to understand how the data were reused.

These case studies were chosen because they are exemplary Open Data initiatives at different geographical levels and suggest paths to improvement the data opening process and the creation of new services.

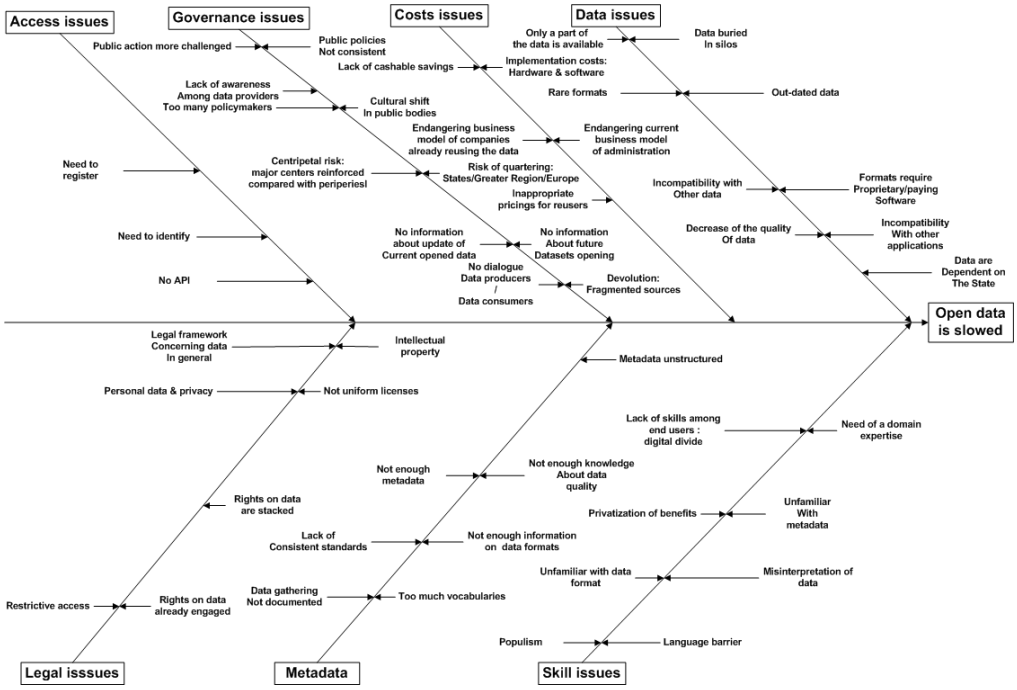

Figure 1: Ishikawa diagram summarising risks and barriers related to data opening

[4] http://data.gov.uk/library/open-data-white-paper

Finally we reviewed existing studies and open data experiences as presented in the literature. Overall we gathered key information on obstacles to data opening, perceived risks, contingency, remediation actions.

3 Risks related to governance issues

Opening data is the result of a political commitment. It raises risks related to the objectives and the sustainability of the initiative.

3.1 Open Data vs. Open government, a misunderstanding

Yu and Robinson (Yu, 2012) highlight a misunderstanding caused by the confusion between open government and Open Data, which are motivated by different objectives. In practice, this leads to consider Open Data as a technical issue, whereas it is clearly not sufficient to engage in a process of open government. An open government process should seek transparency and a profound change in the way in which public bodies operate. Yu and Robinson grounded mostly their analysis on the American case and the first project of the Obama administration in 2009. Within the framework they established, this risk concerns European initiatives to various extents. In Rennes, the concept of open government is not even mentioned. Transparency is discussed. However, the initiative clearly aims to create innovative services for citizens and to generate economic value from data, while strengthening the attractiveness of the city and the region. In Berlin however, open government is one of the stated objectives, of the Open Data initiative[5].

3.2 Reluctance of civil servants

In some cases, Open Data is perceived as a threat by civil servants. The increased control of citizens may lead to protests against public actions, based on an adequate or inadequate interpretation of data which is often de-contextualized. This fear can generate hostile attitudes and finally a reluctance of civil servants to take an active part in the data opening process. In order to overcome this risk, certain actors have engaged in the early mobilization of internal and external stakeholders (e.g., Kéolis in Rennes) and of civil society organizations in order to prevent potential conflicts.

Alonso (Alonso, 2011) also points out the resistance to change in government administrations. He proposes to showcase the social and economic benefits of opening data and to identify champions.

3.3 Inconsistency of public policies

A lack of consistency and perseverance in political behaviors can also put the initiative at risk. Re-users indeed need to be confident that the Open Data policy will be sustainable so that datasets will be updated and continuously available in case they use an API. If Open Data remains the project of a specific team, then it can be questioned as soon as the political configuration changes. However, for the re-users to find a business model and implement a sustainable service, it is necessary that data sources have a certain level of stability and are maintained over time. Therefore, re-users can react heavily to any signal from the authorities. When in France, it was discussed to include *Etalab* (in charge of the French Open Data platform[6]) in a wider agency, many stakeholders expressed concerns about the willingness of the authorities to continue the Open Data policy. However, if Open Data is rooted enough in the administrative culture and operations, if it is supported by a cultural shift in public administrations (Davies, 2010), then it is possible to decrease the risks that the Open Data policy will be overturned.

3.4 The relevant administrative level

Some initiatives have been taken at local level, others at national level. Each local authority makes diverse choices regarding reuse conditions and formats for instance. Each initiative at local level opens datasets which are best suited to its particular context. They raise a risk of fragmentation of the initiative at the expense of the potential reuse of data released beyond the local territory.
A major challenge is therefore to find a balance in the intervention of different political levels, between state intervention (central or federal) that should ensure the consistency of the released datasets and local responsibilities. However, the coordination of efforts at European level (e.g., the European Thematic Network

[5] « Open Data ist ein wichtiger Baustein des Open Government für eine transparente und bürgernahe Verwaltung. » http://daten.berlin.de
[6] http://www.etalab.gouv.fr/

on Legal Aspects of Public Sector Information[7]) and interoperability initiatives (e.g., SEMIC/JoinUp project) can help overcome the fragmentation of projects in Europe.

3.5 The lack of dialogue between data providers and re-users

Another set of risks relates to the relationships between providers and end or intermediate users. It includes the lack of dialogue with the users, the lack of information about the updates of already opened datasets, and the lack of information about the future datasets to be opened.

In Berlin and Rennes, a dialogue has been formalized with groups of re-users (e.g., La Cantine numérique[8]. It can always be strengthened through events dedicated to Open Data (e.g., Berlin Open Data day 2012[9]). However, for both Rennes and Berlin, there is no focus on the last two aspects, i.e., the data update and the new datasets to be released.

Identified risk	**Mitigation**	**Contingency actions**
Reluctance of civil servants	Engage in the early mobilization of internal and external stakeholders, and of civil society organizations. Involve as many people as possible Integrate into day to day process	Ask officials to identify specific issues that explain their reluctance. Identify champions Point out economic and social benefits
Inconsistency of public policies	Favour a cultural shift in the administrations	
The relevant administrative level	Find a balance in the intervention of different political levels	
Lack of dialogue between data providers and re-users	Encourage regular meetings between providers and re-users. Dedicate a specific page for the announcement of updates and future openings, and allow re-users to give their opinions.	

Table 1: Summary of risks related to governance

4 Risks related to economic issues: costs and return on investment

Despite the proliferation of studies on the benefits that can bring Open Data, few of them assess the costs and benefits of each type of data, as it was attempted by the study led by the University of Victoria (Australia) for the various impacts of spatial data and hydrological data (Houghton, 2022). The lack of common standards to assess both the cost and benefits of opening data puts the sustainability of Open Data initiatives at risk.

4.1 The cost of opening data

The cost of opening data is not always calculated by taking into consideration the same type of resources. Implementation costs include hardware, software and human resources. They go beyond the mere technical work to take into consideration the implementation of a process. To overcome the costs of opening data, small communities can mutualize their expenses or rely on national infrastructures. Table 2 shows an estimate of the costs incurred for several platforms. The UK, which has already engaged more resources than most other platforms at national level, has increased its expenses in December 2012 with the provision of a potential credit of £ 8 million for the public bodies that have not yet met their objectives[10].

[7] http://www.lapsi-project.eu/
[8] http://www.lacantine-rennes.net/
[9] http://berlin.opendataday.de/ueber/

[10] http://news.bis.gov.uk/Press-Releases/New-funding-to-accelerate-benefits-of-open-data-684c1.aspx

Platform	Country	Scope	Scale	Cost assessment
data.gov	USA	General	National	Around $10 million/year
data.gov.uk	UK	General	National	2010-2011 : £1,2 million 2011-2012 : £2 million per year[11]
Etalab	France	General	National	€5 million/year[12]
Nantes Métropole	France	General	Local	€ 100 000 (cost of the Portal)
PortalU	Germany	Environment	National	€750 000 / year

Table 2: Cost comparison of Open Data platforms

4.2 Benefits and return on investment

Finally, the way in which it is possible to demonstrate a return on investment is still debated. The uncertainty on the extent and nature of the return on investment represents a clear risk for the sustainability of Open Data initiatives. Very optimistic calculations advanced for example by the organizers of the *Apps for democracy* contest[13] only take into consideration a limited number of parameters : the contest itself deemed the value drawn by the applications worth $2.000.000. The return on investment should take into consideration such heterogeneous benefits as an increased service quality, transparency and trust by citizens, active citizenship through a higher participation in political and public debates, as well as actual cost savings, and the generation of economic activities. In order to match the economic and service related expectations when open datasets, the development of applications and services based on Open Data sets is critical. Application contests help data providers identify new ideas to reuse the data. However, as Janssen (2012) noted « little is known about the conversion of public data into services of public value ». The way in which sustainable services can be created from these initiatives would require further studies. Dedicated actions, such as challenges (e.g., hackathons), applications showcase and the availability of an API are among the key actions which can support the development of applications and services based on Open Datasets (Chan, 2013). However a major issue remains tracing reuse in order to measure benefits. Access through APIs allows tracing reuse more easily than with publication of data dumps.

Finally, the return on investment should take into consideration the benefits to the administrations and public institutions themselves, including benefits related to semantic interoperability (Alonso, 2011) and the change of behaviours of public employees by ingesting a data sharing culture which benefits first of all to eGovernment itself.

4.3 Sustainable business models for the production of data

Data creators are indirectly funded through taxation. They are sometimes also funded by the sale of data or services created from these data. This has led to much debate in the UK about the opening of geographic data from the Ordnance Survey. Releasing data for free on the Internet entails a risk to weaken the data production process and jeopardize the data quality (Uhlir, 2009). Moreover, this might create a distortion in the competition between companies, since certain companies have already established business models based on data they have paid for. The onerous access to this data represents a barrier to entry for new competitors. Opening data in this case entails a risk for the business model of companies that already use public data.

[11] Retrieved from http://www.w3.org/2012/06/pmod/pmod2012_submission_19.pdf
[12] Retrieved from http://www.journaldunet.com/ebusiness/le-net/budget-etalab-1111.shtml
[13] http://www.govtech.com/e-government/Do-Apps-for-Democracy-and-Other.html

Identified risk	Mitigation	Contingency actions
The cost of opening data	Assessing the costs of not opening	Share part of the costs with other Open Data platforms
Benefits and return on investment	Use APIs which allow tracing reuse	Adopt a realistic approach to costs and benefits; Encourage stakeholders who use Open Data to indicate that use. Take into consideration and measure benefits to eGovernment and data producers themselves
Sustainable business models for the production of data		Promote networking between stakeholders; participate in clusters that sustain incubation of companies grounding their business model on Open Data

Table 3: Summary of risks related to the economic issues

5 Risks related to licenses and legal frameworks

5.1 Heterogeneous licenses across datasets

Legal constraints raise mainly the risk of fragmentation of Open Data at all levels, both within countries and internationally, if the licenses and conditions for reuse are mutually incompatible. Indeed, a number of services are based on multiple datasets (e.g., mashups) for which managing heterogeneous conditions of reuse is very challenging. This risk is linked to an incomplete openness of data if re-use or commercial use of the data is limited by the licenses. This was the case in Rennes and the United Kingdom with the first licenses. It has been taken into account by the authorities. Thus, in 2011 the British changed their license. The former license, *Click-Use Licence* designed by the *Office of public sector information* did not allow modifying data. The new *Open Government License* overcomes this limitation by explicitly allowing changes in the data, which is necessary for any complex form of reuse. All the same in Rennes in 2011, this has resulted in the publication of the second version of the « *licence de réutilisation des informations publiques Rennes Métropole en accès libre* ». In both cases, the modifications in the reuse conditions led to a greater openness. The example of the United Kingdom shows the potential value of defining an Open Data initiative at the national level with a partially top-down approach where the national catalogue not only aggregates data, but also defines a real editorial policy. The UK catalog opens over 9000 datasets released with a single license. Coherent conditions of reuse are expected to facilitate the reuse of data.

In Rennes, the whole datasets recorded were covered by the license Rennes Métropole V2 at the time of this study. In Berlin, the datasets follow a slightly less homogeneous licensing model. Despite differences, in both cases, licenses applied to datasets in Berlin are open, except for 4 datasets described as "*keine Freie Lizenz*" (**Table 4**).

License	Occurrences
Creative Commons Namensnennung – Creative commons attribution	54
Creative Commons Namensnennung - nicht – kommerziell – Creative commons attribution-NonCommercial	1
Creative Commons Weitergabe unter gleichen Bedingungen - Share Alike	1
GNU - Lizenz für freie Dokumentation - GNU Free Documentation License	1
Keine Freie Lizenz – No free license	4

Table 4: Licenses in Berlin

5.2 The stacking of rights over individual datasets

Certain datasets may also include data owned by multiple stakeholders with different policies. The stacking of rights on a dataset happens when several organizations claim ownership or control over a dataset and the conditions of its opening. Some may contest the opening and delay it. In Rennes, this has been prevented by the early involvement of the local transporter, Keolis[14], a subsidiary of the French national train company SNCF, in the data opening process.

Certain initiatives, such as the *Europeana*[15] digital library attempt to enforce a single licence (CC0) so that reusers do not have to work with multiple licenses and terms of use for datasets. On the opposite, the Singapore Open Data platform[16] focuses on the publication of clear licensing conditions and terms of use.

While most efforts currently focus on public data, Deloitte in its analysis (Deloitte, 2012) suggests that the main value of Open Data will result from the combination of public data, business data, and personal data. The creation of most services in the coming years would be based on combinations of datasets. This will likely lead to even more complex situations regarding rights and licenses over datasets.

Identified risk	**Mitigation**	**Contingency actions**
Licence is not open enough	Release data complying with the definition of openness	Collect the concerns of re-users and modify licenses if the barriers are too constraining.
Heterogeneous licences across datasets	Awareness raising among stakeholders Enforce a clear policy to publish terms of use	Strengthen the role of the agency that organizes Open Data
Stacking of rights	Governance choices	
Privacy	Data anonymisation	

Table 5: Summary of legal risks

6 Risks related to data

The data also represent risks related in particular to their reliability, their quality, and their format.

6.1 Data accuracy and bias

The dependence of data producers on State funding can raise suspicions on the accuracy of the data. Some data can be sensitive to political pressure (e.g., unemployment figures) and the context in which they were created may raise concerns regarding potential manipulations by the State.

The choice of the datasets to open is also critical to ensure a high return on investment. An initial phase of data analysis is necessary in an Open Data project in order to ensure that the data which will be open both are of high quality and have a high value for reusers (see Martin, 2013). For instance the Bluenove study in France showed that the data most expected by companies are economic and commercial data as well as geographical data (Bluenove, 2011).

6.2 Data quality as a result of a high quality production process

The discontinued funding of certain activities represents significant risks for the quality of the data. The case of the Netherlands cadaster, even though it is not available as Open Data, shows the sensitivity of the datasets to

[14] http://www.keolis.com/
[15] http://www.europeana.eu/
[16] http://data.gov.sg/

financial aspects. Entirely dependent on funds provided by the State, these were cut repeatedly over 1990's, which has led to a sharp deterioration in its quality (Uhlir, 2009).

Open Data advocates discard the risks regarding data quality by showing the opportunities of involving users in the process of data improvement. By identifying errors and warning the data curators, re-users as well as any citizen can contribute to maintain high quality datasets through crowdsourcing mechanisms.

6.3 Data available in heterogeneous formats

In order to efficiently access datasets, users must identify the appropriate software to read the data and work with them, then to choose the best format according to their needs.
Data are made available in a variety of formats (Figure 2The Klessmann report (2012) indicates that approximately 90% of the datasets in Germany are in PDF format, a format which presents the greatest problems for reuse, but a large part (up to 56% depending on the organization) contains structured information that could be made reusable by converting it for instance to the CSV format.
Some formats are proprietary and the combination of Open Data in proprietary formats incompatible with each other already raises conversion difficulties. This also represents an indirect entry barrier for re-users who wish to access the data but could not acquire the required software.

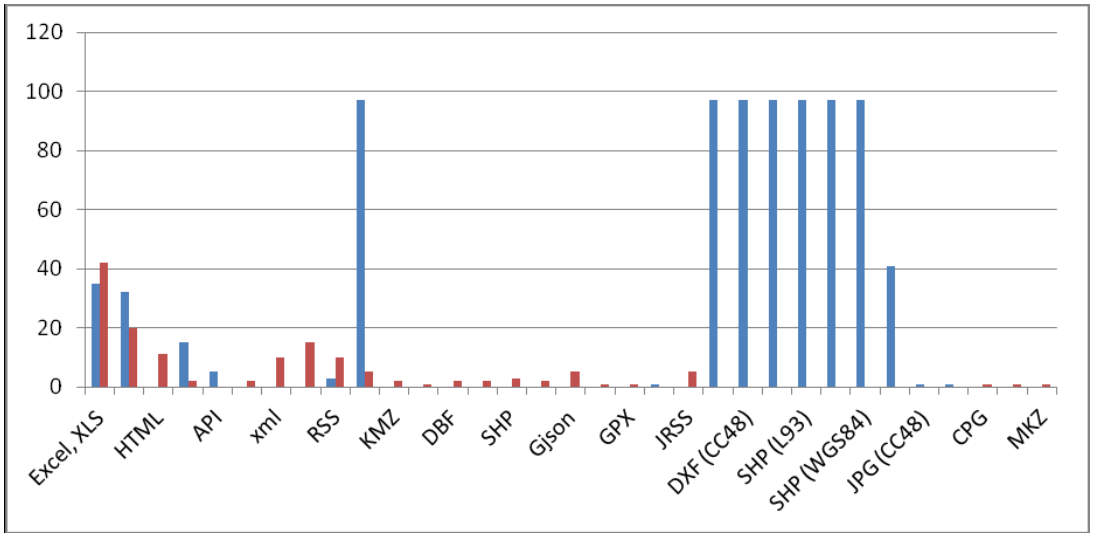

Figure 2 - Formats of datasets in Rennes (blue) and Berlin (red)

Access and reuse can therefore be facilitated if they are produced by software whose code is open (open source software) and published in an open well documented format. In order to ensure that the format in which data are made available is not an obstacle, Rennes and Berlin make many datasets available in multiple formats (Table 67).

	Datasets	**Minimum**	**Maximum**	**Average**
Rennes	137	1	8	5.4
Berlin	61	1	9	2.3

Table 6: Formats by dataset in Rennes and Berlin

The policy of opening a dataset in multiple formats is however not systematic and depends on data creators.Tim Berners-Lee[17] proposes to evaluate Open Data according to criteria that give each dataset a rank based on its openness and its reuse abilities. One limitation of the current Open Data is that most of the data obtained at most three out of five stars, which in the frame of Tim Berners Lee limits the success of the releases and restrains the value of the data.

[17] http://5stardata.info

Identified risk	Mitigation	Contingency actions
Data accuracy and bias	Clarify the context of the data creation process	
Data quality as a result of a high quality production process	Stabilize funding for the creation of data and promote crowdsourcing	
Data available in heterogeneous formats	Publish datasets in various formats but early to create incentive for later standardization	Develop guidelines and encourage standard formats across government bodies Training in semantic technologies

Table 7: Summary of data issues

7 Risks related to metadata

Metadata are assigned to describe datasets. They are very important for the retrieval and reuse of datasets.

7.1 Lack of single standard to describe datasets

For the description of datasets, metadata are most often formatted according to the Dublin Core[18] and DCAT vocabularies[19]. However, there is no single standard to describe Open Datasets. Re-users have to deal with multiple vocabularies. Coordination efforts are then necessary to overcome the difficulties raised by the heterogeneity of metadata models used to describe Open Datasets. In France an initiative has been launched to harmonize metadata from practices identified at the local level[20].

7.2 Incomplete metadata

The lack of metadata, the lack of mechanisms to ensure the quality of metadata, and the lack of information on the objectives and means that have led to their creation or their aggregation also represent risks for the efficient reuse of Open Datasets. For example there is often no information on how data were used in the first instance. Generally, the documentation of data provenance[21] and context which would allow interpreting the data is critical.

In Rennes, the catalogue is being enriched with the recent addition of metadata properties, for instance the precision of geographic data and their original reference system.

In Berlin, Both (2012) demonstrates the key role of metadata for the future of Open Data and even suggests tracing reuses through metadata.

Identified risk	Mitigation	Contingency actions
Lack of single standard to describe datasets	Start with standards as early as possible	Participate in the harmonization of metadata between Open Data catalogues
Incomplete metadata		Gather metadata needs from re-users; implement mechanisms to trace the provenance and use of datasets.

Table 8: Summary of metadata related risks

8 Risks related to access

Open Data should be accessed by both humans (end-users) and machines (through re-users). When setting up an access interface, some platforms request users to register and log in to access the data. This can discourage

[18] http://dublincore.org/documents/dces/
[19] http://www.w3.org/TR/vocab-dcat/
[20] http://opendata.montpelliernumerique.fr/Vers-une-harmonisation-des
[21] http://www.w3.org/2011/prov/wiki/Main_Page

potential re-users by establishing tedious procedures. On the opposite if the platform does not impose any identification, it becomes very difficult to know who is accessing what data and reusing it.

More and more, platforms enable access through APIs (e.g., data.gov in the United States) for re-users who can then automatically access and update the datasets, instead of maintaining their own copy of the data. They relieve service creators of the task of updating data. By ensuring that data used by service creators is up-to-date, the data providers increase the quality of services. They also control better their reputation through the accurate representation of their datasets. Nevertheless, the proportion of data accessible through APIs is still low. For example, in Rennes, there are only five datasets opened through an API. These datasets are also among the most used by applications created from public data. Although it is unclear that this is due to the presence of an API (as they also happen to belong to the domain of mobility, highly popular among data re-users), it suggests that APIs can indeed support the reusability of data.

Identified risk	Mitigation	Contingency actions
Balance between free access and the need to know the use of data	Provide all the data through an API capable of reporting access and use	

Table 9: Summary of access related risks

9 Risks related to user and re-user skills

While Alonso (Alonso, 2011) emphasizes the importance of training data producers in semantic technologies in order to ensure the publication of high quality data, reuse also depends on the skills of potential reusers. Indeed, the risks entailed by the implementation of an Open Data initiative also relate to the potential users and re-users identified for the data.

In particular, analysing the skills of re-users can help understand how to facilitate the reuse and the type of services that can be developed on top of the datasets.

9.1 The language barrier

Whereas for Berlin and Rennes, the question of multilingualism did not arise, other countries such as Luxembourg or Belgium are multilingual. Moreover the creation of services at European level requires that the data published from different countries be understood sufficiently by re-users to be retrieved and used without any risk of misinterpretation. In Luxembourg, the vast majority of the datasets held by public administrations are in French. While there are only few cases where data published in German are not also available in a French version, the creation of transnational services requires implementing mechanisms to guarantee the linguistic interoperability of datasets.

9.2 Skills related to information literacy and domain knowledge

Concerns have been raised regarding the ability of Open Data to equally benefit all social categories. Benjamin, Bhuvaneswari and Rajan (2007) suggest that in some cases opening data can lead to a deterioration of living conditions for a part of the population, while benefiting to a minority who had the necessary skills to make use of the newly released information. In India, an open government project aimed to put online the digitized cadastral data. The better-educated classes, largely those wealthiest, have seized this opportunity to enhance their properties and challenge the rights of other owners. However, this type of risks is inherent to any innovation and only reflects the extent of the digital divide.

The issue of skills is also related to the ability of stakeholders to generate profits from Open Datasets. It is also represented in the concerns about the privatization of public data, with a few people grabbing what should be a common wealth. In this regard, Chignard mentions genealogical data, which represent a very important market (Chignard, 2012).

These risks are to a large extent beyond the scope of Open Data, in particular risks related to the level of education and information literacy. However, they can be addressed through the development of data visualisations, which can ease the understanding and interpretation of the phenomena described in datasets.

In addition, education can help improve the skills of users and re-users through initiatives led by re-user groups which organize training sessions to present available data and the tools and methods to work with them[22].

Identified risk	Mitigation	Contingency actions
The language barrier		Publish data in multiple languages and / or fix the issue through metadata
Skills related to information literacy and domain knowledge		Mention the data provenance and their first use through metadata; provide training to re-users during the events around Open Data
Re-users are unfamiliar with metadata		Assessing metadata formats known by users

Table 10: Summary of skills related risks

10 Conclusion

Many reports have analysed the benefits of Open Data and report on Open Data initiatives. Janssen, Charalabidis and Zuiderwijk (2012) insist on the myths that have accompanied the development of Open Data. From a strategic perspective, Yu and Robinson explore the particular risks related to open government (Yu, 2012), while Lessig has early expressed reservations on the benefits that one can expect from transparency[23].

This calls for a more pragmatic approach grounded in demonstrated benefits and a clear assessment of the risks associated with the implementation of an Open Data strategy. By analyzing the barriers and potential benefits of Open Data, without the ambition of being exhaustive, we propose prevention measures and contingency actions which can be taken. As illustrated by Alonso (Alonso, 2011), barriers to Open Data are not technical. They are rather 1) cultural, 2) economic, 3) legal, and 4) semantic.

However, while Open Data is often considered at the level of general public policies, we note that not all types of data raise the same risks and opportunities. The sale of certain types of datasets is potentially very profitable, whereas others do not have existing markets. Rennes has to a large extent focused on geographic data, while Berlin has opened many economic datasets. The services developed based on the datasets can therefore be of very different nature, making all analyses on costs and benefits very difficult to apply across cases. As noted by Martin (Martin, 2013) who led a survey on barriers to Open Data, the current focus is on Open Data supply. This only represents one aspect to be tackled to match the promise of Open Data. It is necessary to also investigate the creation of services based on Open Data so as to maximize the return on investment of data producers and publishers.

The analysis in terms of return on investment is very different according to the type of data. However, specific actions, such as the definition of complete and standardized metadata can enhance the potential for reuse of datasets and therefore increase the return on investment, whichever the type of data that is considered.

All the same, different types of actors may perceive risks in a different way, due in particular to their local context. Engaging in a risk management framework tailored to the specific context of data providers can help considering Open Data beyond the traditional barriers highlighted by opponents. Most importantly, it demonstrates the need to consider the deployment of an Open Data initiative as a long term process whose sustainability can be improved through the evolution of all stakeholders: users and re-users through the enhancement of skills and the creation of efficient associations; data creators through the prediction and selection of formats necessary to enhance the reuse of data and the release of multiple data formats; finally intermediary platforms such as national aggregators which can help overcome risks related to the fragmentation of datasets, in technical, semantic, as well as legal terms.

[22] Retrieved from http://lemag.lacantine-rennes.net/2012/10/atelier-infolab-a-la-chasse-aux-donnees-rennaises-de-mobilite-1752

[23] Retrieved from http://www.tnr.com/article/books-and-arts/against-transparency?page=0,0#

Future work will be dedicated to the study of the different types of datasets and services developed and the way in which it is possible to optimize the return on investment of Open Data initiatives by selecting relevant datasets and understanding the process by which successful services can be built on top of those datasets.

References

Alonso J.-M., (2011) "Open Government Data (approaches, concerns and barriers, lessons learned)" Share-PSI workshop, Brussels.

Benjamin, S., Bhuvaneswari, R., Rajan, P. (2007) "Bhoomi:'E–governance', or, an anti–politics machine necessary to globalize Bangalore?", CASUM–m Working Paper.

Bluenove (2011). "Open Data : quels enjeux et opportunites pour l'entreprise ? "

Both, W. & Schieferdecker, I. (2012) *Berliner Open Data-Strategie*, Fraunhofer Verlag.

Chignard, S. (2012) *Open Data: comprendre l'ouverture des données publiques*. FYP.

Davies, T. (2010) "Open Data, democracy and public sector reform". Available at: http://www.opendataimpacts.net/report/wp-content/uploads/2010/08/How-is-open-government-data-being-used-in-practice.pdf

Davies, T. & Bawa, Z. (2012) "The Promises and Perils of Open Government Data (OGD)". *The Journal of Community Informatics*. Available at: http://ci-journal.net/index.php/ciej/article/view/929/926

Deloitte. (2 12) *Open Data driving growth, ingenuity and innovation*. Deloitte analytics paper. Available at: http://www.deloitte.com/assets/Dcom-UnitedKingdom/Local%20Assets/Documents/Market%20insights/Deloitte%20Analytics/uk-insights-deloitte-analytics-open-data-june-2012.pdf

Gilbert, D., Balestrini, P. & Littleboy, D., 2004. "Barriers and benefits in the adoption of e-government". *International Journal of Public Sector Management* Available at: http://www.emeraldinsight.com/journals.htm?articleid=868029&show=abstract

Houghton, J., 2011. *Costs and benefits of data provision*. Melborne: Centre for Strategic Economic Studies (Victoria University) Available at: https://www.oerknowledgecloud.com/sites/oerknowledgecloud.com/files/houghton-cost-benefit-study.pdf

Janssen, M., 2012. "Benefits, Adoption Barriers and Myths of Open Data and Open Government". *Information Systems Management*.

Klessmann, J. ; Denker, P. ; Schieferdecker, I. ; Schulz, S., 2012. *Open government data Deutschland. Eine Studie zu Open Government in Deutschland im Auftrag des Bundesministerium des Innern*. Berlin: Bundesministerium des Innern (Germany).

Martin, Ch. (2013). "Understanding Barriers to Open Government Data". Open Knowledge Foundation Blog.

Pasquier, M. & Villeneuve, J., 2007. "Organizational barriers to transparency a typology and analysis of organizational behaviour tending to prevent or restrict access to information". *International Review of Administrative Science*.

Uhlir, P.F., 2009. *The Socioeconomic Effects of Public Sector Information on Digital Networks: Toward a Better Understanding of Different Access and Reuse Policies*. OECD.

Yu, H. & Robinson, D., 2012. "The New Ambiguity of Open Government". Princeton CITP/Yale ISP Working Paper. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2012489&