Creating an open ecosystem for fostering innovation with Public Sector Information (PSI)

Open standards and open access to data drive innovation

Roslyn Docktor, Jochen Friedrich, Peter Haggar, Diane Jordan, Arnaud Le Hors
(all IBM Corporation)

Introduction – Open Data: A Technology Challenge

The European Commission has identified public sector information (PSI) as a major area where the public availability of data can encourage innovation by opening access to significant quantities of raw data for use and analysis. By making the data openly available – as so-called "open data" – new business applications will evolve, making use of this data and providing services and information that are derived from this PSI.

The term "open data" is, however, not clearly defined. It can be construed differently in different contexts by different audiences. "Open" can mean transparent, publicly-available, free to see, free to use, etc. "Data" can refer to raw numbers or more generically mean information, documents, presentations, etc. "Open data", therefore, may be about the terms and conditions, inferring certain copyright or rather copyleft Intellectual Property Rights (IPR) terms; or it may be more about the technological aspects and making data available in a machine readable format.

This paper focuses on the standards behind the big picture of data – the information, text, spreadsheet – as all of this data is stored in digital formats. Accessing, analysing and processing the vast multitude of data will, in general, be done with information and communication technologies (ICT). Accordingly, it is prudent to address the related technology issues and challenges, including formats, protocols, APIs, etc.

The Commission has, therefore, rightly recognised the issue of PSI as a key action item in the Digital Agenda for Europe (DAE):

> "[...] governments can stimulate content markets by making public sector information available on transparent, effective, nondiscriminatory terms. This is an important source of potential growth of innovative online services. The re-use of these information resources has been partly harmonised, but additionally public bodies must be obliged to open up data resources for cross-border applications and services." (DAE, section 2.1.1, p. 5)[1]

Moreover, activities towards the development of a "Pan-European data portal" were started[2], the "Open Data Challenge" was announced in April 2011[3], and issue of open data is top on the agenda of several

1 The Digital Agenda for Europe is available from the website of the European Commission at: http://ec.europa.eu/information_society/digital-agenda/index_en.htm

2 See the report from a "Technical workshop on the goals and requirements for a pan-European data portal" of November 2010, available at http://cordis.europa.eu/fp7/ict/content-knowledge/docs/report-ws-pan-eu-dat-porta_en.pdf.

3 See http://opendatachallenge.org/.

Members States and is addressed in a number of conferences.[4]

Standards, and in particular open standards, play a key role. They ensure a level playing field for non-discriminatory access and interoperability. In other words, open standards are critical for developing an open infrastructure and an open ecosystem for providing PSI. Open standards give any entrepreneur the ability to innovate to create value-add services and information for citizens. Consequently there should be no encumbrances or technological barriers for accessing the data.

This paper briefly elaborates on the benefits of, and the specific requirements for, using standards in the context of PSI and discusses desirable characteristic for standards to be considered open. Finally, it offers some recommendations for actions and next steps.

The value of making PSI available

Open access to PSI fuels greater social and economic development. Governments want to be open to their citizens, open to the increasingly interconnected economies of other nations and open to choose the best technology platform for their administration. Governments want to benefit from the collective knowledge as well as promote the development of new business, e.g. new services and creation of value ad as part of industrial and innovation policy. With access to data, collaborative innovation is possible – different constituencies can build their own reports, coordinate across different data sources and derive more information from raw resources. This can, for instance, address serious problems too complex to be solved by a single individual, agency or government. Problem-solvers thrive in community environments knowing their innovations will work across applications, platforms and networks. And, collaboration can lead to more and better product ideas in less time. Also, data will be available for new and innovative intelligent analyses being done with the data and "on top" of what has already been done.

A sample case: City Forward

It has been clearly demonstrated that as access to public information increases, additional value is realised for both governments and citizens. For example IBM's "City Forward" demonstrates some of the potential benefits open data can bring. City Forward[5] is a free, web-based platform that enables users - city officials, researchers, academics and interested citizens world-wide - to view and interact with city data while engaging in an ongoing public dialogue about cities. Users can explore and analyse data collected from numerous public sources featuring metropolitan areas, cities and smaller geographic areas varying in size, development phase and location. City Forward's exploration tools help identify patterns, trends and correlations in the data that may reveal new insights and point to new areas for further investigation.

4 As one example, the renown re:publica conference held in Berlin, Germany, from April 13-15, 2011, had an explicit section on "open data". For the European Commission Carl-Christian Buhr, member of the cabinet of Vice-President Kroes, gave a presentation giving an overview of the Commission activities in the area and confirming the Commission's intention to drive open data. The slide deck is available at slide share: http://www.slideshare.net/ccbuhr/open-data-eu-policies-and-activities.

5 http://cityforward.org.

Regarding the economic benefit of open data, a recent blog post on the City Forward website gives some interesting estimations:

> "It's true - making government data available to the public has genuine, tangible economic results. A number of official studies have shown that the economic benefits of increasing access to public data can be staggering. In fact, one Danish report estimates that "business reuse of public data in Denmark could be worth at least DKK 600 million (more than EUR 80 million) a year." The 2010 United Nations e-Government Summary cites an example where the state of California spent $21,000 to create a "spending transparency" website - and saved over $20 million over a few months as a direct result of the site. The same report pointed to a $50,000 application development contest (Apps for Democracy) using public data supplied by the District of Columbia. The 47 applications created for the contest would have cost the District $ 2.6 million to develop on its own.
>
> "The growing list of studies on this topic shows that the relatively marginal costs associated with investing in open data can have a tremendous positive social and technological impact on local economies. Open data projects have a demonstrably positive benefit on government - it's important that we get the approach right through tools like City Forward, making sure the data isn't just open, but is also easy to explore."[6]

There are other examples as well where business and governments are stepping up to the challenge of providing open access to data. The BBC, for instance, has made significant progress in providing users with the ability to locate items of interest and correlating information across them using technologies based on semantic web standards. Similarly, governments have found it possible to open new opportunities by making data available with standards which enable linkage providing correlation of information that then makes possible enhanced cross agency communications, investigation of regulation violations, tracking expense and budget recovery, and managing natural disaster relief.[7]

Objectives and challenges

The key objective for making PSI available is to foster innovation. Anybody – an individual, group or business – who has new methods or tools, new business approaches, or new ideas, can find new ways to use the PSI to add value for citizens. This makes government more effective. Consequently there should be no encumbrances or technological barriers for accessing the data. To ensure this openness, decision makers should consider the following principles when setting policy or procuring solutions for opening up PSI:

Flexibility: ensure a level playing field and freedom of action so that no one company can pace, control or block technology; increase technology options for all stakeholders to easily access open data and to easily adapt to requirements and procedures.

6 http://www.cityforward.org/wps/wcm/connect/cityforward_en_us/city+forward/blog?entryID=6bbf3a03-271e-48f1-b30b-386a4a9bd769

7 See http://www.data.gov/ and http://www.data.gov.uk.

Interoperability: eliminate barriers that inhibit communications and information sharing, seamlessly and in real-time, as appropriate, within and across government, especially for critical public services like health care, public safety and education.

Cost effectiveness: avoid vendor dependencies and lock-in, increase competition and drive lower prices; nobody should be obliged to purchase one specific product in order to access the data.

Moreover, starting from today, there are basically three aspects to be considered for making PSI available:

1) **Open up data that is available already:** Data that is available in a variety of different formats and stored in a number of different structures and different ways should be opened-up quickly. Yet, it must be made public in a way so that it can easily be made use of, especially regarding automatic processing and machine readability. Therefore, an analysis should be made of the categories of data that are available and under consideration, and of the formats and data structures in which the data source are available today. On the basis of this analysis – and taking into account the major requirements identified above as critical for effectively making use of open data – it should be decided how, in terms of formats, protocols and data structures the already available data is best provided to meet the objectives of being available for innovative use.

2) **Use non-proprietary formats and API's:** New data, and to the extent possible existing data, should be provided in such a way that no one entity (commercial vendor, government agency, etc.) can control the access to or use of the data, or the technology to leverage it.

3) **Consider common formats, protocols and APIs for providing future data:** the identified categories of PSI, possible use cases and subject to the agreed requirements and objectives for providing PSI provide a basis on which recommendations for the formats, protocols and APIs for future data should be provided. Preferably this agreement should be reached in a harmonised way for all of Europe and also taking similar global activities into account that might already be in place.

It is important that these challenges[8] are addressed properly and with the required sense of urgency. City Forward has provided some insights into the issue. Due to the disparity of how data is being made available today, City Forward has to import data through ad hoc transformation processes on a case by case basis. This makes working with the data not only uncomfortable, but also creates cost and unnecessary technological challenges to ensure that the data is transformed correctly and that information does not get corrupted or lost.

Ideally, data should be made available in machine readable format. This needs to be a key objective for future data. And for data that is available already and is opened up it is of high importance that data

8 See also the proposed 5-star rating from Tim Berners Lee for making data available in general. The first three of the aspects listed in this paper point into a similar direction like the first three “stars” listed by Tim Berners Lee (see http://www.w3.org/DesignIssues/LinkedData.html). Tim Berners Lee's Five Star list was also endorsed by the Technical workshop on the goals and requirements fora pan-European data portal” - see the Report, pp. 6-7 (http://cordis.europa.eu/fp7/ict/content-knowledge/docs/report-ws-pan-eu-dat-porta_en.pdf).

formats are being made available together with the data itself.

Open Standards requirements for open data / PSI

Technology standards including interface, protocol, format and language specifications, such as HTTP, HTML, TCP/IP, XML, ODF and SQL, evolve and mature at different rates and to different levels of openness. There are important factors in determining whether a standard provides appropriate openness for adoption. Similar to other definitions, IBM considers a standard to be open if it meets all of the following criteria:

- **Published without restriction:**; the standard is available at no charge or a charge that is reasonable in cost and can be reasonably administered by parties in the implicated industry.
- **Made freely available for adoption by the industry:** ; historically fair, reasonable and nondiscriminatory (FRAND) and royalty free standards have co-existed. Those standards essential for software interoperability in e-government services, and those essential to accommodate the open source community, should ideally be royalty free.
- **Controlled by an open industry organization with a well-defined inclusive process for evolution of the standard:** This condition guards against the possibility of an individual vendor modifying a standard with the intent of disadvantaging competing suppliers.
- **Implemented by offerings available in the market** ; the acid test for an open standard is whether or not it actually permits substitutability and choice among independent, multi-vendor implementations on different technology platforms with acceptable levels of functionality. Diversity of competing applications that support the standard also indicates its openness and ensures choice for procurers and longevity for use

Open standards ensure a level playing field and are critical for guaranteeing interoperability. And they are a key enabler for innovation. The Internet and the World Wide Web are a prime example here. Open standards are at the core of the boost of innovation we have seen with the World Wide Web. They provide a trusted basis of technology and are available for everybody to implement and innovate on top of them, at the level of the implementation and the applications. Working with PSI and open data requires a similar level of software interoperability as is found in the Internet and the World Wide Web. Consequently Governments should make open standards a priority requirement for data formats and data access for PSI.

Recommendations for actions and next steps

Several initiatives for driving the topic of PSI and open data have been started or are just starting. Most notably, next to the revision of the PSI Directive, there is the PSI Member State Expert Group. Other initiatives include the Open Data Challenge and, of course, Share-PSI. These initiatives should be complemented by an action line addressing the technological challenges and in particular the issue of open standards for formats, protocols, APIs etc. And this action line should be open for broad stakeholder input and expert advice.

The following concrete activities could be part of such an action line:

Develop European Framework for providing PSI:
As part of the revised PSI Directive or accompanying the revised PSI Directive, and in joint collaboration with the PSI Member State Expert Group, a European Framework for the provision of PSI should be developed. This Framework should define the objectives and requirements, including especially those on a technical level. This includes the importance of using open standards, of interoperability, of providing the data in machine readable formats etc.

Launch a study to develop recommendations for standards and specifications:
A number of standards and specifications are already available. The Commission should initiate a study in order to produce an inventory of available open standards and open specifications. The study should also develop pragmatic recommendations for leveraging existing data sources that may be already available in formats which might not be the long term standardisation focus.

Follow-on activities providing guidelines and advice:
Following the basic establishment of a framework and of initial agreements on formats, open standards, protocols, APIs etc. to be used specific follow-on activities should be planned in order to drive and promote the provision of PSI. On the technical level these include in particular the development of architecture guidelines and of implementation services. Collaborative development, sharing of tools and technologies and making re-use of them will help public authorities to stay cost-effective and at the same time reach a pan-European harmonised environment and open ecosystem for the provision of PSI.

Initiate requirements definition process:
Based on the inventory of available standards and specifications a standards roadmap should be agreed on including interoperability agreements and agreements on specific standards and specifications. Also functional gaps need to be identified and it needs to be part of the roadmap to close these gaps. In this context, it is important to consider the global level and similar activities that might already be done in other geographies. Since making use of PSI will not be limited to one specific country or geography, but will be done on a global level where cross-border use of open data may, in some areas, actually be of very valuable and specific interest. Any national or regional specific solutions and technical regulations should be avoided wherever possible.

This action line covers the full spectrum from laying the basic foundation for technological requirements for open data to practical guidance and guidelines on the implementation of the framework. Some of these four proposed actions can be done in parallel, e.g. the first two, the development of a European Framework and the study, don't have any mutual dependencies and could be done at the same time. The proposed actions also take a pragmatic approach on identifying the technologies and the open standards to be used and selected to ensure that information about formats, protocols, APIs etc. is available.

Looking at the technological challenges and pushing for open standards to be used for making PSI available is an important element in the open data debate. Open standards will ensure that an open ecosystem around PSI can be set up allowing all interested parties and innovators to make use of PSI

and contribute with innovative applications to generate added value to the benefit of governments and societies as well as to innovation and growth.

IBM applauds the intention of the European Commission and the Member States to promote PSI and open data. We welcome the opportunity to further this discussion and participate in subsequent actions and the implementation.