ICININFO

Open Data as Universal Service. New perspectives in the Information Profession

L. Fernando Ramos Simón*, Rosario Arquero Avilés, Iuliana Botezan, Félix del Valle Gastaminza, Silvia Cobo Serrano

Department of Library and Information Science, Complutense University, Madrid 20810, Spain.

Abstract

The Internet provides a global information flow, which improves living conditions in poor countries as well as in rich countries. Owing to its abundance and quality, public information (meteorological, geographic, transport information... and also the content managed in libraries, archives and museums) is an incentive for change, becoming invaluable and accessible to all citizens. However, it is clear that Open Data plays a significant role and provides a business service in the digital economy. Nevertheless, it is unknown how this amount of public data may be introduced as universal service to make it available to all citizens in matters of education, health, culture... In fact, a function or role which has traditionally been assumed by libraries. In addition, information professionals will have to acquire new skills that enable them to assume a new role in the information management: data management (Open Data) and content management (Open Content). Thus, this study analyzes new roles, which will be assumed by new information professionals such as metadata, interoperability, access licenses, information search and retrieval tools and applications for data queries.

Keywords: Open Data; information profession; Open Content; public information; librarians

1. Introduction

In the last decade of the 20th Century, the e-Government concept arose from applying information technologies to governance functions in democratic countries that, although still quite vaguely defined, has improved efficiency in managing government and providing citizens better services (Relyea, 2002).

One of the most visible results of the change is that application of new technologies has made a large volume of information available to citizens and companies so as to improve the citizens' daily life (transport, health, weather...) and that of companies that may use that data flow in their own technical, commercial and financial
processes. Thus, throughout the first decade of the 21st Century, e-Government has provided a flow of information and data available for re-use known as Open Data Government, although the concepts "open data" and "open government data" suffer from a certain amount of ambiguity and one may exist without the other (Harlan Yu and David G. Robinson, 2012).

The idea of a more transparent government and free access to information resources from the most varied fields (the environment, weather, scientific information, cultural heritage...) took off parallel to the development of the Internet and has become explicit in a great range of activities: The Budapest Declaration in 2001, the European Directive on the re-use of public sector information in 2003 (amended by Directive 2013/37/EU on 26 June 2013), the formulation of the open data principles in 2007 (see below), the Open Government Directive by Obama in 2009, the open access initiative in the United Kingdom in 2009 and the communication on Open Data by the European Commission (2011) that is aimed at creating a major European data portal and putting European cultural heritage on-line. Although some open data principles have been formulated, these eight principles have a general recognition among the advocates of the movement (Open Government Data Principles, 2007):

1. Full: all data is made available.
2. Primary: data must be collected at the source.
3. Timely: available as quickly as possible.
4. Accessible: for all uses and users.
6. Non-discriminatory: without registration requisites or controls.
8. License-free: limits only to protect personal data and security.

In recent years and with backing by the W3C consortium, the open data movement has consolidated itself on the basic lines of promoting the use of raw data made available through open licenses as well as re-usable and linkable formats, which has given rise to the well known five star scheme. That clear explanation still finds difficulty over the distinction between data and information (Davies, 2010 and Yu and Robinson 2012), expressions that are gradually being replaced by those of knowledge, being understood as open if it complies with the conditions “free to use, reuse, and redistribute it – subject only, at most, to the requirement to attribute and share alike” (Open Definition).

2. Open Data vs. Open Content – status of the matter

Development of new technologies and ongoing evolution of society have brought about the appearance of new ways to produce and share information, which involves more transparency and wider access to information, especially to governmental information. In order to keep up with this development, several tools have been implemented to view and process data, mainly for open access (open data) in spite of constantly being “under construction” (open content).

Work groups have been established for these two subjects, such as Open Access Foundation, Open Content Alliance or Data Documentation Alliance. In spite of not being very numerous, papers on the subject have not taken long to appear -articles, books and regulatory literature (standards,…) as well as organizations that promote them. Noteworthy, articles are those published under the auspices of the Open Access Foundation since 2007.

“The mere existence of data is not enough” stated Pascal Heus and Arofan Gregory (2010) in an article published under the auspices of the Open Access Foundation. In order to cover the users’ needs, authors propose that various measures be taken, such as assuring data quality. As data alone is meaningless, data must be accessible to researchers in order for it to be useful.

Important steps have been taken to improve data quality, access, documentation and data exchange. However, there are still things to improve: adopting and using standards, adapting tools and infrastructures, training people, preparing reference materials, implementing change management.
There are fewer papers concerning Open Content initiative, although many of them re-examine the problem of users’ licenses as well as the terms of use. The latter is dealt with in the paper on Open Education by Katarina Pisutova (2012), who returns to the scheme proposed by D. Weley in 2011, called 4Rs Framework:

- Reuse. The right to reuse the content in its unaltered / verbatim form (e.g., make a backup copy of the content).
- Revise. The right to adapt, adjust, modify, or alter the content itself (e.g., translate the content into another language).
- Remix. The right to combine the original or revised content with other content to create something new (e.g., incorporate the content into a mashup).
- Redistribute. The right to share copies of the original content, your revisions, or your remixes with others (e.g., give a copy of the content to a friend).

The contributions to this process of opening up data are innumerable. On the one hand, Hernández Pérez and García Moreno (2013) especially emphasize the enormous potential of scientific open data and the role that libraries will play in development of institutional repositories. On the other hand, the paper by Saorín (2012) concentrates on the impact Linked Open Data may have on libraries or libraries in linked open data. According to Saorín, libraries must be more concerned with the quality of the data they generate and less with development of applications based on these, for which innovation provided by information mediation companies would provide a more feasible development model. In fact, that model is fairly similar to that defended in the European and Spanish laws on the re-use of public information, in which the information mediation companies (re-users) also play a key role. Moreover, according to the amendments introduced in the European Directive on the re-use of public sector information, it seems that Saorín provides a fairly good description of the library situation.

Ferrer-Sapena and Sánchez-Pérez (2013) discuss the advantages and disadvantages linked to Big Data and Open Data concepts. From a scientific research point of view, the authors state the advantages of Open Data, although they raise doubts regarding the effective loss of data protection under certain circumstances or the use (and abuse) of open data in the information markets.

It is important to point out that the use of open data or open content initiatives means a major challenge both in technical terms and in organizational as well as policy terms.

3. Open data: Public information portals

This international current in favor of open public data on the Internet has brought about creation of government portals as Open Data initiatives, introduced as of 2009. In a study of ten government portals in the European Union (Ramos et al, 2012), conclusions are considered to represent the present situation of such portals. In general terms, portals are quite recent (approximately 2 years) and they are in a phase of organization and start-up, so data consistency and coverage are very unequal. Other additional conclusions are as follows:

1. Open data portal is a low cost showcase for public transparency and is relatively easy to build. However, there is scarcely structured data and the information on offer is not very consistent in quality and quantity.
2. There is a big difference between the United Kingdom’s data portal and the rest. This one responds most consistently to an efficient use of public data.
3. Each portal is designed in a different way; there is no sole model.
4. While search forms have similar features, the way in which results are presented differs according to the type of data shown. It is concluded that the main cause of this dysfunction lies in the fact that data is obtained from databases organized with different criteria.
5. The resources inventory is made according to the administrative tradition of each country and its competence structure, both in territorial as well as functional terms. Therefore, the difficulty of comparing datasets or information packages at present. Each one has been prepared with its idiosyncrasy and according to the information needs of each public body.
6. The access conditions and the terms of use established in the licenses are generally aimed at the open license model, which is based on Creative Commons license. Although the detailed terms of use of the
portal sometimes clash with the license principles, a certain reluctance to accept commercial use of the information is noted.

7. Computer applications to use public data, known as APIs, have a highly unequal development, ranging from a quite testimonial presence to a very solid presence.

8. It is difficult to identify common characteristics other than the basic features of the European Directive 2003/98/EC on the re-use of public sector information: transparency in the information access, elimination of discriminatory treatment or exclusive agreements.

As we see, these open data portals are national initiatives that provide access to public data to facilitate their re-use. In general terms, these are scarcely structured initiatives that are supported by developers while no significant presence of information professionals is noted. However, another public information sector has been formed, whose main characteristic is that it manages content or documents, instead of data. Although this sector operates in libraries, archives and museums, it has had a much less innovative role than the open data sector. In this regard, some initiatives, which may be highly important for the future development of public information management, have been established in recent years, such as Europeana or the Digital Public Library of America (DPLA).

4. Information professionals: proposal of content and competences to respond to Open Data requirements

The new Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013, that amends Directive 2003/98/EC on the re-use of public sector information, establishes that the scope of Directive 2003/98/EC should be extended to libraries, including university ones, archives and museums. That Directive 2013/37/EU shows that those institutions have a considerable quantity of valuable public sector information resources that have been multiplied thanks to the digitization projects that have been undertaken by many of those institutions. Undoubtedly, all these cultural heritage collections form a potential base to design and develop digital content products and services that may generate opportunities to re-use their content and, consequently, develop the so-called European content industry.

In addition to the impact the new Directive 2013/37/UE has in these cultural institutions, it must be added the role of the libraries in the future of open access (Harris, 2012), which means acquiring specific competences in the virtual environment and a mentality change in Library and Information Science (LIS); in other words, an additional academic training that responds to the requisites of the Open Data philosophy.

Thus, as teachers and researchers in the field of Library and Information Science, we must reflect on the competences (it is defined as “the set of knowledge, skills and personal attitudes required of the employees to perform their work efficiently” (Sanabria Barrios, 2013, p. 77)) of the future managers of public information. In that sense, the methodology used to detect these new competences and academic training is divided into the following phases:

1. Analysis of 10 European open data portals (Ramos Simón et al., 2012) and analysis of Europeana and Digital Public Library of America (DPLA).


3. Identification of competences and academic training required for the future library professionals.

4. Reflection, synthesis and presentation of our proposal.

In relation to our objective, it must be pointed out that there are many international Library and Information Science schools that are designing educational curricula so as to prepare the future information professionals who must become involved in:

a. Managing new interactions between public sector information, technologies and people.
b. Decision making to set guidelines and policies that benefit the governance of public sector information and its relation with technology (Weill and Ross, 2004).

In addition to managing public sector information and defining strategies or guidelines, it is necessary to include both academic content of Library and Information Science (information technology and management functions) as well as specific content linked to the public administration in the academic training of our future professionals (Lee and Tibbo, 2012).

In line with that approach, Lee and Tibbo (2012) carried out a research in order to define the content and competences of a double Master in Information/Library Science and Public Administration. Their analysis was based on the following framework methodology:

- Identification of functions and activities performed by information professionals based on a review of specialized literature.
- Identification of curricular needs to manage public sector information.
- Analysis of several curricula in Library and Information Science and Public Administration (Master level).
- Semi-structured interviews with public information management experts.

The main result of that research and the analysis performed in this paper leads us to establish a synthetic proposal based on the definition of three main categories so as to identify learning competences and objectives required for LIS professionals, who will be able to work as public sector information managers (Fig. 1).

---

**Fig. 1. Proposal of learning skills and objectives for Library and Information Science professionals.**

- **First category: knowledge of the organizational context:**

  Knowledge of governmental structures and functions, legal framework, mechanisms to render accounts, agreements management, public sector information life cycle, balance between privacy and accessibility, process planning and evaluation, project management and requirement definition –this is the scope of the product, time,
costs, risks and life cycle- (COI Council, 2012), record keeping and knowledge of patterns of use of the public sector information.

- Second category: Technical and operating skills - Technological integration

Knowledge and design of the Content Management Systems (for example: OJS, Drupal or Wordpress), data management quality control, workflow design, process control, storage processes, access methods, conservation techniques, information architecture, metadata, user-based interface design, data modeling, information organization, markup languages, ontologies and Linked Data.

- Third category: Social skills and working mentality:

Communication skills, teamwork, commitment to opening government up to citizens, efficiency and economy in processes, social skills, digital marketing, Web positioning, community management, agents for change, ability to obtain sponsorship or alliances with stakeholders and ongoing training.

5. Conclusions

Internet has changed the way people communicate, including the communicative way of governments and citizens. There are many ways to articulate that information flow which affects both data access as well as its management. Although it is generically known as open data, different expressions are used in different contexts: open data, open content, linked open data, big data...

Pressure by citizens and social dynamics have encouraged a current in favor of free information flow, particularly between governments and citizens. Public data portals, which are created by governments and different inter-institutional initiatives, are the best practice to represent such data and information flows.

As that information flow has been introduced in society and all professions, the information manager’s traditional role has become fairly different. Nevertheless, they fulfill an essential public function in the private and public sector.

In that sense, experiences obtained in different public information scopes (access and re-use, portals, databases...) let managers detect essential competences to manage data as well as information, including libraries and archives.

The research results have emphasized three categories concerning competences and general training objectives must be included in the university curricula: knowledge of the Administration context, technological integration as well as technical skills and, lastly, social skills to act as information mediator.

Should that approach be adopted, Library and Information Science schools should orientate their curricula toward citizens' present information needs.

Bibliography


Open Definition. http://opendefinition.org/