Following the user's flow in the Digital Pompidou

Emmanuelle Bermès
Centre George Pompidou, Paris, France.
manue.fig@gmail.com

Abstract:
Since 2007, the Centre Pompidou, major modern art museum in Paris, developed a new digital strategy aiming at providing a global platform for online digital content: the Centre Pompidou Virtuel, which could literally translate as "Virtual Pompidou Center" or more accurately "Digital Centre Pompidou" . This platform provides access through a unique entry point to the whole digital production of the organization and associated institutions (Bpi, Ircam): digitized works of art, documents about art and art history, videos and podcasts, archival material, library books records, etc. The goal of the project was to make the online presence of the Centre Pompidou focus on the content rather than being just an institutional showcase mainly targeting physical visitors of the building in Paris. On the contrary, the Pompidou website is now a reference online tool for anyone interested in modern and contemporary arts, or in the humanities in general.

Keywords: Semantic Web, Library Linked Data, User interfaces, User generated content

0 INTRODUCTION

Since 2007, the Centre Pompidou, major modern art museum in Paris, developed a new digital strategy aiming at providing a global platform for online digital content: the Centre Pompidou Virtuel, which could literally translate as "Virtual Pompidou Center" or more accurately "Digital Centre Pompidou" . This platform provides access through a unique entry point to the whole digital production of the organization and associated institutions (Bpi, Ircam): digitized works of art, documents about art and art history, videos and podcasts, archival material, library books records, etc. The goal of the project was to make the online presence of the Centre Pompidou focus on the content rather than being just an institutional showcase mainly targeting physical visitors of the building in Paris. On the contrary, the Pompidou website is now a reference online tool for anyone interested in modern and contemporary arts, or in the humanities in general.

Hence the Digital Pompidou is not about providing a virtual experience that would try to copy the onsite experience for visitors in our exhibitions, through an interface based on camera views like the Google Art Project.
First of all, we wanted to emphasize the diversity of our cultural activity, which doesn't rely only on the museum, but also involves conferences, live shows, cinema screenings and other live events involving artists of all kind.

Moreover, we're less interested in displaying what can already be seen than in revealing what is usually hidden. Among the 76.000 art works that are part of our museum collection, only about 2.000 are actually exhibited in the Paris building, in the course of temporary exhibitions or presentations of the permanent collection. The rest is either on loan or deposit in other places all around the world, or stored in the museum's storeroom.

Finally, the Centre Pompidou remains convinced that there is no way a virtual experience can replace the actual contact with works of arts, the emotional and sensible approach to our cultural heritage. We hope that making more content available online will unveil new possibilities, either by making a new range of people come to the museum, or by allowing the display of forgotten works that wouldn't have been considered for exhibition had they not been digitized.

These statements led the Centre Pompidou to the definition of a broader scope, defining the "virtual" experience as something completely different from what can be seen onsite. This experience is based on the ability of our web users to create their own flow of meaning, by following links and aggregating content according to their own interests.

Traditional editorial approaches of online development in museums lead to emphasize only those works or artists that are considered of main interest, which results in the prevalence of a mainstream culture above more alternative forms of creation. Following the long tail paradigm, artists that are already well-known tend to have a heavier presence on museum websites than others. The Centre Pompidou is attached to its tradition of openness to new and alternative forms of arts and doesn't want to privilege a certain part of its collection, but rather empower unexpected discoveries and serendipity.

In order to do so, the Centre Pompidou has created an online platform aggregating a great diversity of content, starting from the digitized works, which are the backbone of the website, but also including documents and archival material related with those works and their creators. Most of these digital resources were not created purposely for the website, but rather reflect the actual day-to-day activity of the Centre since its creation in 1977.

A large part of this content was actually already available on the former website, but they were scattered and hardly accessible for a user without a thorough knowledge of information retrieval techniques. In order to allow the average user to benefit from these hidden treasures, the Centre Pompidou adopted a semantic approach in the design of the new platform. The combination of semantic Web technologies and an intense research on end user interface issues resulted in the creation of the Digital Centre Pompidou as it is today. But a lot of other possibilities are still waiting to be unveiled.

1 WHY SEMANTIC WEB TECHNOLOGIES?

One of the main challenges of the project lied in the creation of a global and common information space from data extracted from several databases which all have their own structure. We decided to adopt Semantic Web technologies in order to address this issue.

The Digital Centre Pompidou was created based on the aggregation of existing databases, which are used as management tools for the Centre's professionals in the course of their work. The main databases are:

- the museum collection, a database dedicated to the management of the works of art and their curation; this database is based on a software shared with other French museums, called Videomuseum;
- the agenda, a database which describes all the events (exhibitions, conferences, workshops, visits, etc.) past, present and future;
- the library catalogues, based on traditional ILS sytems (3 library collections are aggregated in the Digital Centre Pompidou: Bibliothèque Kandinsky, Bpi and Ircam);
- archives finding aids, both from the Centre Pompidou's institutional archive and from the Kandinsky library which holds several artists' funds;
- audiovisual databases, usually based on local tools;
- other databases holding biographical information, journal articles, learning resources, shop products, etc.

It was a major challenge to be able to aggregate data from all those databases into one common interface for the public to search and browse. The data is very heterogeneous, as some of it follows library standards (MARC, MODS and Dublin Core), some archival standards (EAD for archives), some an internal locally defined structure (museum and audiovisual material) and even part of this data relates to entities that are not documents by nature (events, persons, etc.)

However, merging all this data reveals very interesting, as all those databases share common entities: for instance, if you're looking for Kandinsky, you could find interest in his paintings, the exhibitions that shown his works, his archives held by the Kandinsky library, books and videos about him, photos of him in the archives, etc. All this information already exists in the different databases, but relating it in a consistent way still is a challenge.

In the course of the project, it was not our purpose to change the habits and tools of the professionals, so the principle of having separate databases was no subject for discussions. Of course, a global change towards a digital-oriented purpose of the activity was needed, and led to new practices such as requesting authorisation for online display when the content is copyrighted (which is almost always the case, as the Centre Pompidou preserves mainly art from the 20th and 21st centuries). Also, indexing the content of the resources was now necessary, as the idea of making those works accessible to the public at large also required different entry points from those used by professionals. However, beside these amendments to the way of describing things, very little change was induced by the project in terms of software or data models.

The new digital platform had to use these databases as sources, and aggregate and relate their content. As the data models were so different, the choice of semantic web technologies was almost natural.

Linked Data offers a powerful way of achieving interoperability between databases of heterogeneous structure. In his final report (Baker et al., 2011), the Library Linked Data incubator group emphasizes the relevance of these technologies for libraries, in particular in the perspective of interoperability across domains (libraries, archives, museums):

> *"The Linked Data approach offers significant advantages over current practices for creating and delivering library data while providing a natural extension to the collaborative sharing models historically employed by libraries. Linked Data and especially Linked Open Data is **sharable, extensible**, and easily **re-usable**. It supports*

multilingual functionality for data and user services, such as the labeling of concepts identified by language-agnostic URIs. These characteristics are inherent in the Linked Data standards and are supported by the use of Web-friendly identifiers for data and concepts. Resources can be described in collaboration with other libraries and linked to data contributed by other communities or even by individuals. Like the linking that takes place today between Web documents, Linked Data allows anyone to contribute unique expertise in a form that can be reused and recombined with the expertise of others. The use of identifiers allows diverse descriptions to refer to the same thing. Through rich linkages with complementary data from trusted sources, libraries can increase the value of their own data beyond the sum of their sources taken individually."

"By using Linked Open Data, libraries will create an open, global pool of shared data that can be used and re-used to describe resources, with a limited amount of redundant effort compared with current cataloging processes.
The use of the Web and Web-based identifiers will make up-to-date resource descriptions directly citable by catalogers. The use of shared identifiers will allow them to pull together descriptions for resources outside their domain environment, across all cultural heritage datasets, and even from the Web at large. Catalogers will be able to concentrate their effort on their domain of local expertise, rather than having to re-create existing descriptions that have been already elaborated by others.
History shows that all technologies are transitory, and the history of information technology suggests that specific data formats are especially short-lived. Linked Data describes the meaning of data ("semantics") separately from specific data structures ("syntax" or "formats"), with the result that Linked Data retains its meaning across changes of format. In this sense, Linked Data is more durable and robust than metadata formats that depend on a particular data structure."

The principles of Linked Data are designed to be applied to the Web at large and across organizations, but they can also be fit for internal use within an institution or company: this kind of use is usually referred to as "Linked Enterprise Data" (Wood, 2010). "LED" is about applying Linked Data principles and technology within the information system in order to increase interoperability between its components.

The four main principles of Linked Data are the following:
- use URIs as names for things
- use HTTP URIs so that if someone looks up a URI he retrieves useful information
- when someone looks up a URI, provide useful information using standards (RDF, SPARQL)
- provide links to other datasets.

The goal of these rules is to provide to the end user a seamless information space where he can "follow his nose" from one resource to the other, following their URIs, without the need to have any knowledge of their structure or storage. This form of interoperability should allow different institutions to publish their databases without knowledge of the software used by others, just as the Web allows web pages and websites to communicate through hypertext regardless of the storage of the pages on different servers and using different content management systems.

This is exactly the kind of interoperability that we wanted to build within the Centre Pompidou information system. We wanted a system that wouldn't enforce the data from the separate databases into one common structure, but still would make it possible to create links between the entities they share. As the museum professionals are very demanding in terms of data quality, we couldn't afford to lower the level of detail of the data by using only the smallest common denominator between the databases. The main advantage of the RDF model, with its triple structure and the use of URIs, is to bind together descriptions of entities of a great variety into a seamless data model.

2 THE DATA MODEL

The Digital Pompidou platform is based on a RDF core that binds together all the data from the different databases. They are thus expressed according to a common model using RDF and URIs. In order to handle this, an RDF ontology was created, based on a few main concepts: Work, Document, Person, Place, Collection, Event and Resource (see Fig.1).
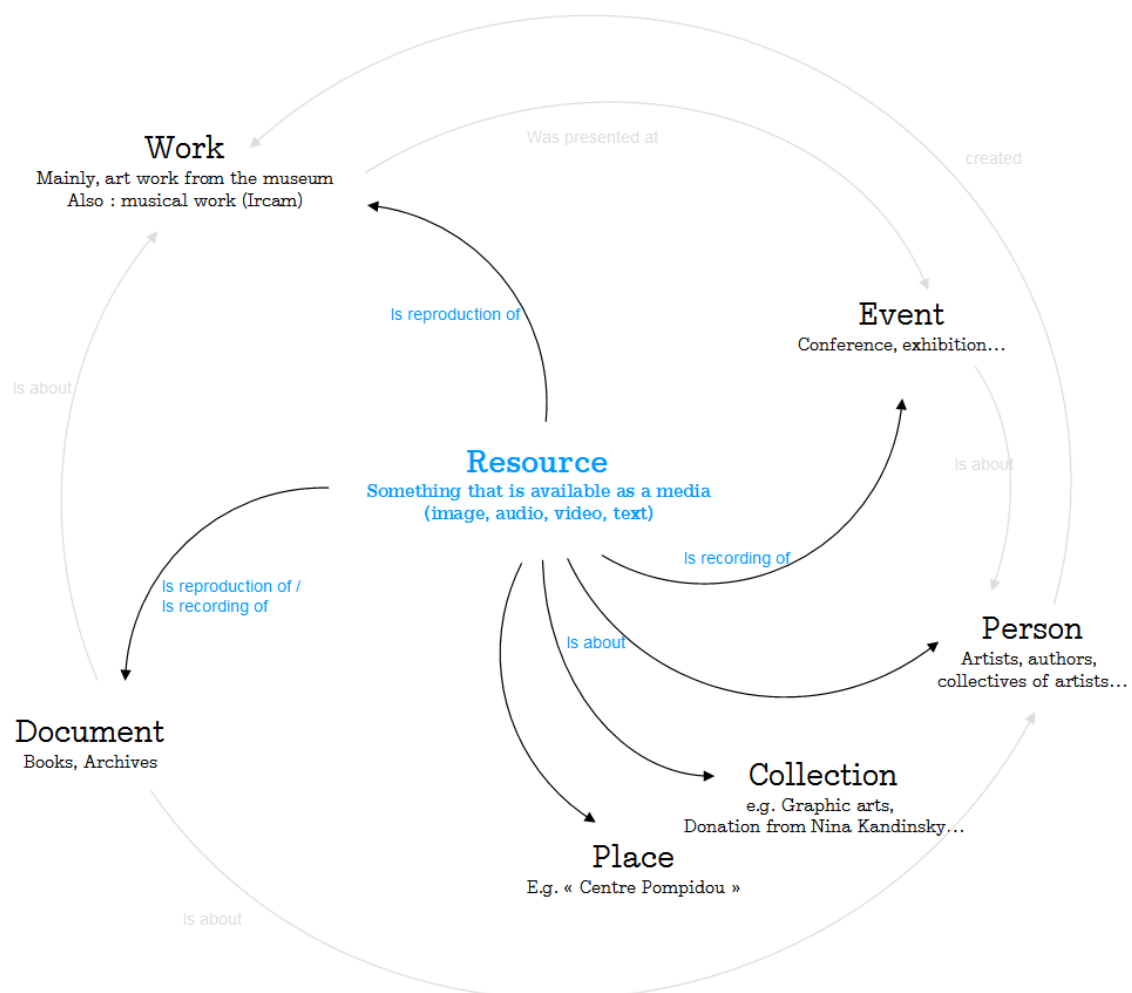

Figure 1. Overview of the Digital Pompidou data model

The main concepts are designed to integrate data from the different sources: data from the Museum mainly relates to Works, Persons and Collections. The Agenda provides information about Events but also the Places where they are located. Data from libraries and archives are aggregated around the concept of Document. Finally, audiovisual material is provided as Resources together with information about the content of videos and audio recordings

(Persons who are speaking during the conferences, Works that are presented, etc.). Art content (works from the museum, recording of musical performances) is linked with event-based information (exhibitions, performances, conferences) and with other relevant resources (posters, photographs, books, archives, etc.) thus allowing to browse the website and discover all these resources in a serendipitous manner.

During this process, we learnt that it was quicker and easier to work with our own schema than try to adapt existing vocabularies, none of which were completely fit for our purpose due to the diversity of our resources. However, this was possible only as long as we intended to use the data within our own system and not to redistribute it to partners or make it available as open linked data. If we were to do so, which is definitely part of our plan for this year, we'd have to transform our local ontology into a standard one. Work has already been done in this respect, in collaboration with a class of students in documentation.

We also learnt that creating links between databases is not a trivial task, even if they are owned by the same institution and supposed to address similar topics. Entities such as Persons can be aligned using very simple keys such as given name plus family name. In most cases, the result is relevant because we are working on a narrow field of interest with little risk of ambiguity (there are, however, a few cases of homonyms). When it comes to Events or Works, it's a whole different story. Events often have very ambiguous names, and even taking into account the event's dates it's difficult to disambiguate for instance the name of an exhibition and the visit for disabled people of the same exhibition, or a series of conferences around the same topic. Works also have ambiguous names, and if you consider the collections from the Photography or Graphic Arts services, "sans titre" is probably the most frequent title in the database...

Moreover, those alignments have to be recreated each time the data is updated in the system, a process that happens every night for most of the data, as the website displays sensible information that requires frequent updates (ongoing events, data about people, right owners, etc.). Hence, even if we were able to edit manually the alignments, in order to disambiguate false positives for instance, the editing would be over-written by the next nightly update when the new source data comes in and erases the existing one.

In order to solve this issue, we worked to cross the identifiers from the different databases directly in the source databases. For instance, the Audiovisual database has been enriched with a new data element which is the unique identifier for an event in the Agenda, imported from the latter. Using a specific interface, the people in charge of describing videos can pick up from a list the proper event the video is related to, in order to make sure that the link between the media and the event will be accurate. This kind of improvement requires evolutions of the source databases and the professionals' practice, even if it is only on the margin of their activity, but is important to ensure that the user experience will be consistent.

Finally, many of the links are still created manually by the Multimedia team who is in charge of curating the data for the website. Thanks to the RDF model, it is very easy to add links between resources. We use an editing interface called "RDF Editor" to create those links, which are basically triples binding existing URIs in our datastore. The RDF Editor then behaves like a new source database which only stores links. Those links are restored daily when the rest of the source data is updated. This process only requires that the URIs are persistent across updates.

3 THE USER INTERFACE

The purpose of the creation of this data model and all the links between the databases is to provide our users with a new experience: being able to browse the semantics of the data. The initial purpose of the project was to make it possible for users to retrieve our content, in particular art works, by using words from the natural language. A query such as "horse" would then retrieve not only those works of art that have the word "horse" in the title, but every representation of a horse thanks to iconographical indexing. This feature is offered in our website by the search engine SolR.

The creation of links also provides a very different way to browse the data. The interface has been designed to allow the presentation of many links on a single page. This required the use of design tricks such as the clickable tabs that unfold vertically to display more and more content. This presentation makes it possible to display all the related links to the resource, hence providing different points of view on the data:
  - If a user is interested in a Work, he can discover the artist who created it, see different digitized versions, have access to audiovisual material such as an interview of the artist, or textual material like an article extracted from a printed catalogue. He can also discover in which previous exhibitions this work was shown, the place where it can be seen now, and browse a series of works ingested in the same collection.
  - If the entry point of the user is the event, for instance if they want to see what exhibitions are ongoing currently in the Centre Pompidou, they have access to information about the event but also audiovisual material, recordings, the catalogue of the exhibition they can buy from the online shop or read at the library, discover related events such as visits for children or conferences, etc.

This ability to browse the graph according to one's own centre of interest is what we call "the flow of meaning": users extract their own meaning from the circulation in the graph of data; they build their own course adapting it to their interest and the time they want to spend on the website. Whereas in most websites the circulation is mainly hierarchical, from the most general to the most specific, in the Digital Pompidou the navigation takes the form of an hypertextual graph where all resources are displayed at the same level. Our website is also different from a traditional database in the sense that usually, you have to express a detailed query in order to reach a resource; if the query doesn't get you to what you're looking for, you would then try to reformulate it. Database records are often dead-ends, with no other choice than creating a new query to find other resources. On the contrary, the Digital Pompidou always offers links to other resources and allows broadening the search to things that were not looked for in the first place.

Before the website was officially launched, we conducted a user study to evaluate this new way of discovering content. We found out that our users did perceive that it was a completely new way of exploring data. They sometimes felt lost in the richness of the content, or had the impression that their browsing was circular; they looked for the site map and requested tools in order to help them visualize a location in this information space.

While expert users (academics, students) said they first had to spend a lot of time on the site in order to understand how it works, users who were just browsing the site out of curiosity liked the fact that they would get lost and discover unexpected resources. However, they often complained that they had found interesting resources but were not able to reproduce the path that had led them there, hence to retrieve the resource.

So it appears that we actually did succeed in creating a user interface that is completely original and specific to the fact that the underlying structure of the website is based on Linked Data. However, we need to develop new tools in order to help our users grasp the advantages of hyperlinked data and non hierarchical models. This is what we intend to develop as a next step.

4 PERSPECTIVES FOR FUTURE DEVELOPMENT

We are currently working on a new version of the website that will bring some improvement to help our users with these aspects. In particular, the personal account will provide a "history" function and will record every resource that the user has displayed so that it will be easier for them to retrieve what they have already seen.

Future evolutions of the platform also include an even greater involvement of users in the construction of the meaning or semantics of the content, as we intend to offer collaborative tools for resources indexing and linking. This new feature will allow users to add keywords in a wiki-like interface, and thus create new paths to help find resources.

Integrating user's contributions into our data model, much alike the integration of several databases, is made easier by the fact that we're relying on a Linked Data model. Any addition to a resource can be managed as a RDF triple, a simple annotation to the content created by the Museum. Provenance information will need to be addressed in order to be able to keep the user generated content separate from the content that is validated by the institution.

Another interesting aspect of having a RDF core for our data is the fact that our current interface is only one of the infinite possibilities of presentation for the many links that we have created from our source data. The Digital Pompidou today provides the official interface displaying this data, but it's only one possible interpretation among many. We can predict that we'll work on a new design within 2 or 3 years, in order to improve the user experience, take into account the feedback we had and adapt to the new material that we're currently digitizing.

However, it would be interesting also to see what other actors could build by interpreting our data and in particular the added value that was created by the aggregation of databases that were distinct in the origin.

Data visualization has been explored since many years now, as an alternate way to provide access to large collections of data, but without succeeding to overcome traditional interfaces such as textual search engines or mosaics made of small images. Now with the development of open data, big data and data journalism, a new interest emerges regarding these techniques, not so much as a querying tool that you would put into users hands but as a storytelling tool that can bring up new perspectives for your data. For instance, we could build a representation out of the Digital Pompidou's data presenting the links between artists based on the exhibitions that have shown their works.

Experiments have been conducted in this regard by IRI (Information technology research institute), a partner of Centre Pompidou on research in IT and digital developments. IRI has created HAD-Lab, a portal dedicated to learning resources in history of art for teachers. HAD-Lab experiments several interfaces for structured data, including a visualization tool

that shows relationships created between different learning resources by tagging them with DBpedia URIs (see fig. 2.).
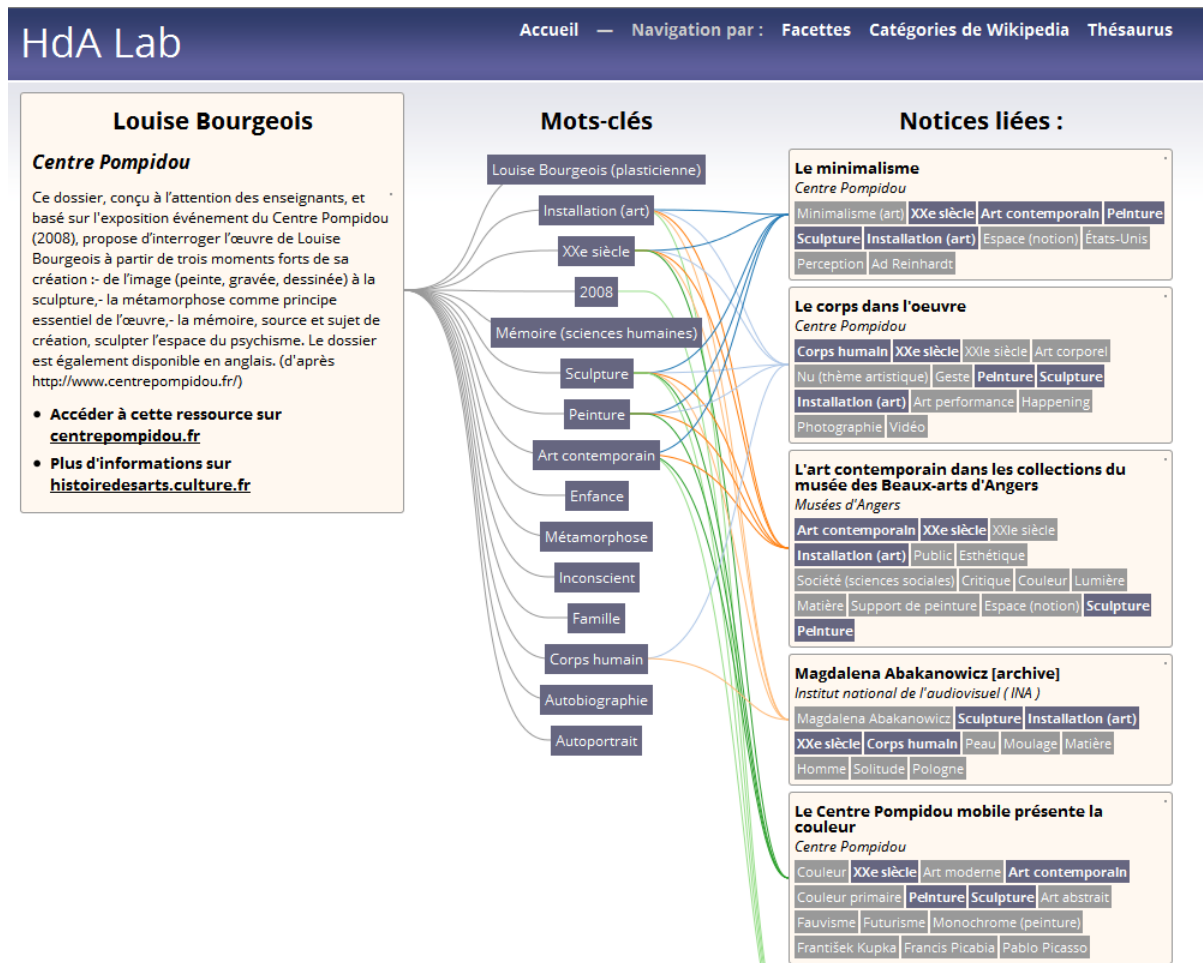

Figure 2. HDA-Lab visualization interface

The important part of this idea is the storytelling: data visualization is only interesting if there is a story to illustrate or if it allows new stories to be discovered. The task of unveiling these stories can't be delegated to the user himself, and information professionals such as librarians don't have experience in this area. For this reason, it is very important that other players (data journalists, data visualization experts or even artists) can have access to the raw data in order to be able to build their own representations, invent their own stories based on the material that we can provide.

During the summer of 2012, several experiments of data visualization were presented during the exhibition Multiversités créatives in the Centre Pompidou (Bernard, 2012). The point was to demonstrate the added value of data visualisation when it comes to understand complex and moving structures such as the Web. In the future, the Digital Pompidou should naturally become a source for this kind of experiments, thus providing a new perspective on the works and the artists presented in the museum.

In this perspective, the Digital Pompidou can be envisioned as an open door to a new interpretation of art and humanities. This tool will be even more powerful when the Pompidou data will be linked to external data such as Wikipedia, Freebase, Viaf or data.bnf.fr

(Wenz, 2010): this is definitely a motivation to open the data and make it available for others to create unexpected new interfaces with it.

5 REFERENCES

Baker, T., Bermès, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., . . . Zeng, M. (2011). Library Linked Data Incubator Group Final Report. Retrieved from: http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/

Bermès, E. (2009). *Linked data and why we, librarians, should care* [Slideshare presentation]. Retrieved from: http://de.slideshare.net/Figoblog/linked-data-and-why-we-librarians-should-care

Bermès, E. (2011). Convergence and interoperability: a Linked Data perspective [PDF]. Paper Presented at the IFLA World Library and Information Congress, Puerto Rico. Retrieved from: http://conference.ifla.org/past/ifla77/149-bermes-en.pdf

Bernard, C.-R. (2012). Multiversités créatives : Entretien avec Antonin Rohmer et Guilhem Fouetillou [Video file]. Retrieved from: http://www.centrepompidou.fr/cpv/ressource.action?param.id=FR_R-fcfa20aad7968c0c76cecc245537a6&param.idSource=FR_E-fbe91867c63213578548d4c568cb35a

Koster, L., (2011). Missing links: the challenge of generating linked data from legacy databases [Blog post]. Retrieved from: http://commonplace.net/2011/03/missing-links/

Keller, M. A., Persons, J., Glaser, H., Calter, M. (2011) Linked Data for Libraries, Museums, and Archives: Survey and Workshop Report. Retrieved from: http://www.clir.org/pubs/abstract/reports/pub152

Wenz, R., (2010). data.bnf.fr: describing library resources through an information hub [Video file]. Retrieved from: http://scivee.rcsb.org/node/27093

Wood, D., (2010). *Linking Enterprise Data*. Boston, Mass.: Springer. DOI: 10.1007/978-1-4419-7665-9