

# When we come together: when libraries, archives, museums and galleries are linked together

---

*Nicolaie M. Constantinescu,  
Information architect, Kosson Community  
kosson@gmail.com*

## Abstract

Libraries, Archives, Museums and Galleries are all together bound to cross a very turbulent period in the advent of a golden age. The debate over the convergence of these fields landed for the first time on neutral ground when they were coined as memory institutions and invited to collaborate in building common services like World Digital Library, Multilingual Inventory of Cultural Heritage in Europe, Access to Cultural Heritage Networks Across Europe, The European Library and Europeana. In the light of the unfolding economic situation, the struggle to keep a valid social profile is often taken on the grounds of collaboration where enough resources are attracted in order to keep balance and even progress. The future presents a leverage point that should not be ignored or treated lightly: building the informational context through open and standard compliant technologies. LAMG specialists claim to have switched to a functional paradigm that puts the user needs as the core business. Doing so is to prepare for the future and the future is one where the resources are aggregated and linked in what we call today Semantic Web. Here we take into consideration the building blocks of this future starting from what we have, and advancing towards an inclusive picture where we are together. Open standards, open technologies, and open licenses are the key components of this space for collaboration among us. We should be swift in embracing them as our users have done this already.

**Keywords:** libraries, museums, archives, galleries, collaboration, open standards, open technologies, open licenses, semantic web, rdf, library linked data, linked open data, open data

## In search for context

Knowledge cannot be achieved without having a rich context for the information. Knowledge is a state of refined information that is highly interconnected, and tuned to cover as many aspects of a particular class of problems. The refinement of the information is achieved through establishing as many as possible connections that taken together form a particular meaning. What are these connections in the context of

yesterday information consumption? How do they look now? From the old times of antiquity scholars put a lot of efforts in finding meaning and order in things. Today, the specialists of the information science field are still searching for the best models to build a context needed to extract meaning out of the information deluge. The so-called memory institutions are at the forefront of these efforts. The high valued product is usually information about their collection of documents and cultural artifacts. The trend is showing that libraries are pulling their products together<sup>1</sup> in big data aggregation services or they are contributing as partners in different project frameworks like The European Library<sup>2</sup>. Many of these efforts are shaped in the form of a portal. Unfortunately, the sum of the portal's features does not always cover the complexity of the information or the entire scale of the potential users<sup>3</sup> uses for that particular collection. In fact, a clear case for the scenario when all the memory institutions will come together sharing the same interest is the moment when they are making the user part of their business case and not a final beneficiary. When the users feel involved and are given the means to contribute and shape certain aspects of the informational contexts they are exposed to, something magical happens and this is the moment when "the user" becomes a community member. This new member has more value as it comes with his/her unique set of habits attached to the manner in which they retrieve and judge meaningful information. The contributions of the members are manifested knowledge parts that add to the rich context mentioned earlier. From this moment on, the users will be referred as members – members of a community we all serve.

## Information related habits

Once we are on the Internet, we do it *naturally*, as our members do! We do very well in using searching instruments and the unique search field is a generous invitation to input a query string. The most advanced features like autocomplete and suggestion prove to be useful in the effort of getting relevant results. In any case this habit of opening the browser and search for information is in direct competition with the reference librarian serving targeted answers. It is proper to say that these new reflexes are muscled by the most precious factor and resource in human existence: time. Search engines sell the world dream of relevant timely information. The issue of relevance comes immediately into the debate of establishing a proper informational context. But, it is observable that our members tend to sacrifice relevance and to train their noses in picking up relevant sources rather than calling a library specific service<sup>4</sup>.

Although the rich offer libraries are proposing the user to exploit, he/she takes another path choosing online search engines where multiple types of queries may be shaped and forged on incremental trial error bases.

---

<sup>1</sup> [WorldCat](#) – “the world’s largest network of library content and services”

<sup>2</sup> [www.theeuropeanlibrary.org](http://www.theeuropeanlibrary.org) - “The European Library is a membership organisation for national and research libraries. Its members include 48 national libraries and an increasing number of research libraries.”

<sup>3</sup> “It doesn’t always give me the results that I want.” Excerpt from Milena Dobrova, Pierluigi Feliciati, et al. (March 2010). *User and Functional Testing* (Final report, Europeana v1.0, Grant Agreement Number: 558001). Retrieved from [http://pro.europeana.eu/c/document\\_library/get\\_file?uuid=1c25ae28-9457-4b0f-be62-654a7cf6c5b7&groupId=10602](http://pro.europeana.eu/c/document_library/get_file?uuid=1c25ae28-9457-4b0f-be62-654a7cf6c5b7&groupId=10602)

<sup>4</sup> Keren Markey, The Online Library Catalog: Paradise Lost and Paradise Regained? in D-Lib Magazine, Vol. 13, Nr. 1/2, January/February, 2007, source: <http://www.dlib.org/dlib/january07/markey/01markey.html>

This behavior should not be fought over by the information specialists, but embraced and remolded in new information behavioral patterns yet to be developed by the research core of the field. More and more online services are inviting the user to be part of the collection building processes.

A possible way to cross the gap between the memory institutions and the new demands of the members is to force an engaging action. This action would drive their creative and collaborative will towards meeting common goals of the community, a community that in the case of finding the right stimulate is contributing back.

Crowdsourcing<sup>5</sup> in some particular aspects concerning the digital collection build up are not anymore isolated cases. A good exercise is letting patrons become a decisional component. More than that, keeping a reserved and elitist view on our jobs not on the long run will prove to be a lost game.

## We, the information specialists

We are used to work with classification and indexing schemas that are hierarchical but this was only a natural approach since knowledge is considered to be a “tree”. It is no surprise to find this approach as an inherited philosophical good.

Starting with Aristotle’s famous tenfold division<sup>6</sup> and arriving to the first considerable effort aimed to catalogue the entire world based on a decimal system (Paul Otlet and Henri La Fontaine) – Mundaneum, all of us trained the eye and hone the professional skills to this way of seeing the informational space organized.

But what is to be observed is that the Aristotle’s classification is closer to the faceted classification and also beyond that he saw the seeds of another approach with regards to ontologies (the four fold division): semantic relations (Predicamenta). Semantically structuring the known universe is not a new idea, as it may seem at the first look, but a very old search that only recently met information technology means.

Almost everything that comes in touch with organizing information has the idea of a “tree” in its conceptualization and practical application.

Examples: searching algorithms, models for organizing a file system, traditional book classification, etc.

Apparently all that makes relevance to organizing information seems to be the works of a “tree” conceptualization. That is because we need a reference system to organize our manifested thinking. A desired development would be the natural language processing of the information, but this needs new patterns for processing the very structures that keeps human communication meaningful.

Still, until the systems will be able to “understand” natural language and be able to shape discrete information resources into meaningful contexts useful for us, we should take a look into a recent approach developed as a next step for the electronic

---

5 Rose Holley, Crowdsourcing: How and Why Should Libraries Do It?, in D-Lib Magazine, Vol. 16, Nr. 3/4, source: <http://www.dlib.org/dlib/march10/holley/03holley.html>

Supplementary resources  
<http://www.zooniverse.org/>  
[http://www.galaxyzoo.org/how\\_to\\_take\\_part](http://www.galaxyzoo.org/how_to_take_part)

6 (1) Substance, (2) Quality, (3) Quantity, (4) Relation, (5) Where, (6) When, (7) Position, (8) Having, (9) Action, (9) Passion

communication. A new way opens and the chance comes in the form of all the efforts hidden under the general canopy of “semantic web”<sup>7</sup>. This is the point in time where hierarchical thinking (data structuring and presentation) meets semantics.

We, the information specialists are the first to understand the real value of the semantic web. That is because in the end it doesn’t deliver a rich informational context as long as the data start to have meaning for the machines, but also the technologies needed to exchange data across systems, and most important query discrete sources seamlessly.

Why do we need to go on this road?! Because an inclusive environment it is needed to be able to host and exploit after the transitional phase of digitization of the LAMG’s collections. There are two components that need our full attention: the catalogues and the digitized resources. Both of them are the most valuable products, and they need a rich context to function properly.

## A semantic approach

The inventor of the World Wide Web and the director of W3C, Tim Berners-Lee coined the term “semantic web” as “a web of data that can be processed directly and indirectly by machines”.

W3C is actively engaged in adding to the “Web of documents” the “Web of Data”. The final goal is to link those data into a machine actionable data stores. There is a technology stack to support achieving a “semantic web”: RDF (Resource Description Framework, SPARQL (the query language used on RDF formatted data), OWL (Web Ontology Language) and SKOS (for designing knowledge systems). There is also a very important aspect that has to be take into consideration: trust. No semantic web technologies could be developed and used without being sure that the internet transactions that carry data are not safe and secure.

A new grammar is in store for delivering an increasing resource space that has the potential to transform the way we learn about the world.

## Resource Description Framework – a new promise

“Resource Description Framework (RDF) is a framework for representing information in the Web”<sup>8</sup>.

All things described by RDF are called resources<sup>9</sup>.

RDF is a language<sup>10</sup> based on an XML syntax.

RDF makes use of vocabularies<sup>11</sup> based on the use of URI<sup>12</sup>s.

---

7 <http://www.w3.org/standards/semanticweb/>

8 <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>

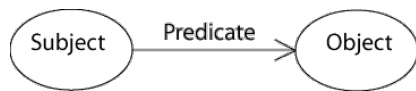
9 <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>

10 <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>

11 idem 8

12 Uniform Resource Identifier – represents a path to where a particular resource is located.

<http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/#dfn-URI-reference>



The structures of the expressions used in RDF are called triples. Triples consist of a **subject**, a **predicate (also called the “property of a triple)”** and an **object**. A particular set of triples form at one time the RDF abstract syntax.

A set of triples is called a **RDF graph**.

Resource Description Framework is one of the technologies we call *semantic web*, which in turn is called the Web of Data.

RDF holds the promise of becoming the standard used to describe the components of the information space. It has the means to describe resources and to connect them to form meaning not only for us humans, but for the machines as well.

Much of what the RDF graphs are describing are in fact data.

For the information specialists working in LAMS the data are the catalogues we build for our community members to help them find valuable information and resources in our collections. But this data contained in the catalogues have a new chance to tell a meaningful story as long as we provide them with linkage means.

This is what we call Linked Data. At this moment the RDF gives the common format for this data.

## LAMGs need Linked Data, Linked Open Data

Many cases of good use of the semantic technologies are to be found on the dedicated site of W3C<sup>13</sup>.

The foundation of all these efforts is data.

Starting with May 2010 W3C has established a Library Linked Data Incubator Group having a focus on linked data and interoperability of library data (bibliographic data, authorities and concept schemes). This important step forward puts the linked data as a prime enabler for:

- “Remove cultural bias in metadata format and content.
- Replace aging standards and models that are not conducive to web-scale use.
- Recommend a basic level of functionality and basic data requirements for metadata created by national bibliographic agencies.
- Improve ability to re-use licensed metadata by shifting focus away from the metadata collection and record.
- Improve matching and de-duplication of metadata.
- Improve metadata aggregation from all sources.
- Extend coverage to linked data from other communities.
- Add functionality to services.”<sup>14</sup>

Beside these positive changes there is a point on which the paper finds its balance: uniformed data has the power to offer more visibility for the collection as long as it is

---

<sup>13</sup> <http://www.w3.org/2001/sw/sweo/public/UseCases/>

<sup>14</sup> [http://www.w3.org/2005/Incubator/llid/wiki/Cluster\\_BibData](http://www.w3.org/2005/Incubator/llid/wiki/Cluster_BibData)

created and exported in a Linked Data format. This way “convergence” is in memory institution’s grasp:

“This scenario is expected to facilitate the creation of semantic links between heterogeneous material from libraries, archives, and museums.”<sup>15</sup>

The Group just issued the Final Report<sup>16</sup> which draws some important conclusions:

- **Library data is not integrated with Web resources.** This is mostly an issue linked to the architecture of the library software that use a database. The data contained is not linked to the internet resources. Linked Data technology offer this opportunity;
- **Library standards are designed only for the library community.** Much of the library standards today still are the standards developed for a certain period of ICT infancy. This has to change. **Library technology changes depend on vendor systems development.**
- **Library data is expressed primarily in natural-language text**
- **The library community and Semantic Web community have different terminology for similar metadata concepts.** Librarians are not talking in terms of metadata statements. On the other hand the members of semantic web community do not have in their scope notions like “headings” or “authority control”. Here LAMGs are shining and are able to make a significant contribution to the field.

One concrete step towards a seamless integration of all types of LAMG resources, is the recent pilot of the Europeana project: Europeana Linked Open Data.

Data.europeana.eu pilot is an effort of making the metadata Europeana aggregates as Linked Open Data. The data is represented in the Europeana Data Model (EDM) and the resources are addressable and dereferencable by their URIs.

This should be taken as a clear signal that a common informational space is built for the LAMGs. The new paradigm of the linked data as a building block for the semantic web makes the convergence a reality and not just a dream afar.

## The future is not a trend

The enabling mechanisms and technologies are here already doing their work and for the profit of all those who understand to put a little more effort.

LAMGs have the tools and the information space already at their feet. The only thing missing is a leap of trust into the future. This leap of trust is in fact accepting the change and reacting positively by learning all the new skills needed.

The rich semantic context in its making is the very moment when we all come together.

---

<sup>15</sup> <http://www.w3.org/2005/Incubator/lld/XGR-lld-usecase-20111025/>

<sup>16</sup> <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>