

## LAM From the Trenches: The LODLAM Community and the Building of LAM Culture, Practices, and Education

**M. Cristina Pattuelli**

School of Information and Library Science, Pratt Institute, New York, USA.

E-mail address: mpattuel@pratt.edu



Copyright © 2014 by M. Cristina Pattuelli. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

---

### **Abstract:**

*Linked Open Data in Libraries, Archives, and Museums (LODLAM) is an international community of information professionals, researchers and educators working towards the application of Linked Open Data (LOD) principles and technology to digital cultural heritage. Emerging in 2011, LODLAM has gone from a small grassroots effort to a movement at the forefront of fostering the culture of integration and sharing across different cultural heritage domains through LOD research and implementation. This paper discusses the role and activities of LODLAM from the perspective of an active participant in the community.*

**Keywords:** Linked Open Data; Digital Cultural Heritage; Semantic Web; Libraries, Archives and Museums; Professional Organizations

---

### **1. Introduction**

Linked Open Data in Libraries, Archives, and Museums (LODLAM)<sup>1</sup> is an international community of information professionals, researchers and educators engaged in the research and implementation of Linked Open Data (LOD) technology to digital cultural heritage. LODLAM can best be described as an open movement that emerged in the last few year rather spontaneously, driven by the goal to transition cultural content to the LOD ecosystem and realize the vision of the Web of Data, the semantic extension of the web as we know it. Self-organized and a-centric, LODLAM is rapidly growing through a string of activities, offline and online, global and local, that bridge different cultural contexts and strengthen the ties of various research and professional communities. As the LODLAM movement is too novel and evolving to be captured in its complexity, in this paper we intend to simply

---

<sup>1</sup> <http://lodlam.net>

highlight its cultural and technology context as well as some of the stages that have been pivotal to its development and consolidation.

## **2. LODLAM Background and Vision**

A practical outgrowth of the Semantic Web, the LOD initiative provides the context for fostering greater collaboration among libraries, archives, and museums (LAM), as well as other cultural institutions such as galleries, which inspired GLAM, the acronym often used interchangeably with LAM. The Linked Data initiative is defined as a set of practices for publishing structured data on the web in ways that make this data machine-readable, interlinkable and open to novel and enhanced uses. Technically, linked data methods rely on a few common open standards including the HyperText Transfer Protocol (HTTP) and the Unique Resource Identifier (URI). The Resource Description Framework (RDF) serves as the unifying model to structure the data (Bizer, Heath and Berners-Lee, 2009).

As a subset of Linked Data, LOD focuses specifically on freely sharable and re-usable data. This open model works as a catalyst for data and metadata from distributed and heterogeneous sources to be shared, connected, and re-used. The linked data “technology stack” addresses well-known challenges such as data interoperability and integration and makes it possible to steer this data into the web to support open and seamless access and use. In other words, the web itself becomes the management platform and the global and unified space for data discovery and navigation. We have only begun to envision and understand the potential for cultural heritage organizations to offer integrated access to their content.

This new and boundless scenario calls for a strategic reconsideration of our traditional documentation, curatorial and discovery methods. Libraries, museums and archives have historically developed diverse and, to some extent divergent, descriptive practices that are tailored to modes of access and use of resources rooted in the analog world. LOD technology offers a common framework for data structure and representation where newly created as well as legacy metadata can converge and connect. To permeate the web with “radically open cultural heritage data”, Voss (2012a), it is essential to depart from the “walled garden” model of siloed data and metadata. Back in 2009 Berners-Lee, founder of the LOD initiative, had urged to “stop hugging your data” and to start connecting data to each other in order to enable unprecedented opportunities for data analysis and interpretation.

The culture of openness and sharing across domains drives LODLAM leaders, innovators and enthusiasts to create the conditions for overcoming invisible institutional walls and to contribute to the expansion of traditional access to cultural content. From the humanities to the sciences, it is the mashup culture that underlies the idea of interlinking granular content from diverse sources and mixing and matching open vocabularies from different communities to create new meaningful pathways to inquiry and discovery. This is a culture of cooperation that doesn’t necessarily require coordination as everybody works within the same data representation paradigm.

It is also a culture of experimentation as a new array of methods and sets of tools are needed to build this new scenario. LOD is considered the practical implementation of the grand vision of the Semantic Web and it is driven by a very pragmatic approach. This means a good deal of crafting new tools and learning by doing. Start small and progress through piloting and prototyping is a LODLAM mantra, especially for small-scale independent projects. Such

an iterative process makes LOD research and implementation less intimidating and more pervasive, while the LOD framework provides the thread that ties all the efforts together.

There is an exciting feeling of being at the cusp of a profound cultural shift in the way digital cultural heritage is managed and “consumed”, as a staggering volume of cultural linked data is being made available by memory institutions and organizations worldwide. LODLAM is contributing to the emergence of a rich web of semantically linked data while at the same time addressing the array of challenges—from licensing and data trust and quality to scalability of query processing, to name a few—that still prevent a large-scale adoption of LOD technology.

### 3. LODLAM – The Story So Far

Jon Voss (2012b), one of the driving forces behind the movement and the co-founder and chair of the LODLAM Summit, recalls that the idea of gathering together a group of information professionals working in the domain of cultural heritage came up in 2009. The goal was to interconnect heterogeneous datasets of American Civil War archival content. It was then that the group realized that there was “an appetite for educating the broader community of libraries, archives, and museums about the concept of Linked Open Data” (p. 41). The next step was the creation of the logo LODLAM and the organization of the first summit, which took place in San Francisco in June 2011.<sup>2</sup> Though originally intended for fifty attendees, the call for participation raised significant interest and the summit ended up with double the number of delegates. A diverse mix of librarians, archivists, researchers, developers, and technologists came together for a two-day event from around the world (seventeen countries were represented).<sup>3</sup>

This first LODLAM Summit is reminiscent of the pioneering spirit of the Dublin Core Metadata Initiative (DCMI)<sup>4</sup> of 1995, when an international and heterogeneous community self-described as “freaks, geeks, and people with sensible shoes” flocked to Dublin, Ohio. Their aim was to find a solution that would overcome the limitations of the search engines of the time and improve online discovery (Weibel, 2005). Over the years, the Dublin Core community has grown and evolved, almost organically, through annual meetings and other venues, and their focus and activities have expanded well beyond the universal metadata schema they developed.

Hosted by the Internet Archive and funded by grants from the Alfred P. Sloan Foundation and the National Endowment for the Humanities, the first LODLAM Summit was organized as an “unconference,” and was thus bound by the conventions of academic and professional meetings. Without a pre-set agenda or paper presentations to attend, participants were free to self-organize and discuss the burning issues of the day. A facilitator and an easel were all that was needed to aggregate people around topics of their interest. Delegates were encouraged to move around and join the discussions that inspired them. This level of spontaneity and openness made it possible to engage people with varying degrees of expertise and technical experience on different approaches to the challenges facing LOD in the cultural heritage arena. Vocabularies were a particularly hot topic because of the critical role they play in the construction of RDF triples – the building blocks of LOD. Concerns related to vocabularies’

---

<sup>2</sup> <http://lod-lam.net/summit/>

<sup>3</sup> The list of delegates can be found at <http://lod-lam.net/summit/participants/>

<sup>4</sup> <http://dublincore.org>

expressivity, issues of dealing with different levels of granularity and the need for mapping and alignments were all part of defining a common space for sharing semantics among different cultural domains and representation practices. Data quality and the need to deal with the messiness that comes with heterogeneous datasets were paired with discussions on provenance and trust. The user side of LOD development also began to emerge in discussions of new modes for data consumption including the lack of usable and intuitive interfaces for LOD-driven services targeted at new user communities. The realization that we were converging around questions that we were all facing or thinking about, regardless of the different practices and traditions, was a strong element of cohesion and an indicator that a community was emerging.

A second Summit followed two years later in June 2013, this time hosted by the National Library and Archive of Quebec in Montreal. Once again, more than one hundred participants from sixteen countries came together.<sup>5</sup> The agenda was created free-style during the conference with only a few sessions proposed in advance. This time it felt to some extent like a reality check where people shared the progress they had made on their work as well as lessons learned. A number of developments had happened in the intervening two years. As Voss (2012a) points out, the amount of Linked Open Data had increased exponentially. Libraries had continued to release their bibliographic and authority data as LOD at a high rate (according to Pohl (2011), the Linked Open Data cloud had grown by 1000% with library data in 2010). Europeana had integrated Linked Open Data capability into its framework and announced plans to publish the metadata for over two million resources as LOD. BIBFRAME announced in the fall of 2011 that it would leverage Semantic Web and LOD principles and methods to represent and connect bibliographic data.

A good deal of conversation at the summit was devoted to the methods and tools needed to implement the LOD model. Key initiatives such as the upcoming release of a handful of Getty's vocabularies as LOD were announced and discussed. This time the summit included a new component – the “LODLAM Challenge” competition.<sup>6</sup>

Intended to highlight use cases, the LODLAM Challenge solicited submissions from projects at various stages of completion and encompassing visualizations, tools, and prototypes. This showcase offered a range of current approaches to LOD development and demonstrated that the implementation of LOD for domains like cultural heritage was still highly experimental. Regardless of the scale of the projects and their level of maturity, prototyping remains common practice and efforts continue to be put into building use cases that can demonstrate the potential of LOD applications.

Teams were asked to submit a video with a description of their project. Two rounds of online voting took place from September 2012 through February 2013. Eighteen projects participated and a total of five finalists were selected to receive travel stipends, delegate seats at the Summit, and eligibility to compete for a prize. Submissions included tools and efforts to convert legacy data of various types to LOD, projects to visualize collections, map and text annotators, and initiatives related to access to specific archives.

The winning project was a web tool called Pundit; originally created by the Semblib project, the tool is currently developed by Net7 and partially funded by the Digital Manuscript to

---

<sup>5</sup> The list of delegates can be found at <http://summit2013.lodlam.net/about/participants-flat/>

<sup>6</sup> <http://summit2013.lodlam.net/challenge/>

Europeana project.<sup>7</sup> A semantic annotating tool, Pundit's focus is on giving users the ability to create a linkable data layer on top of web content including text or images. Annotations are stored on a server and can then be leveraged to create visualizations and graphs.

Linked Jazz, a finalist in the competition, is another independent research project focused on creating broadly applicable tools for LOD development and use.<sup>8</sup> Directed by the author of this paper, the project's aim is to use LOD to represent and visualize the relationships among the densely interconnected network of jazz musicians, as described in digital oral histories. Several applications have been developed to accomplish this goal, including a name mapping and curation tool that allows for automated and human disambiguation of personal names through mapping with name authorities, an analyzer for name recognition and extraction from textual documents, a crowdsourcing tool that allows users to contribute by determining relationships between musicians, and an interactive visualization tool for social network analysis (Pattuelli et al., 2013).

Another finalist, the MisMuseos project, is an example of an application of LOD to museum content.<sup>9</sup> Focused on Spanish public institutions, MisMuseos gathers metadata from multiple museums and libraries as well as the CER.ES collection in Europeana comprised of Spanish art and artists. Using the GNOSS as its publishing platform, it combines museum and Europeana metadata with data from Dbpedia, Geonames and Didactalia, a social network and user-contributed open dataset of educational resources. It currently presents over 17,000 pieces of art and 2,650 artists. Works of art can be browsed by facets including collection headings (based on traditional museum department divisions), location (museum), artist/author, time period, movement, school, type, technique, and medium. Artist entries can be browsed using the collection, movement, and school facets as well as a birth-place facet.

In the domain of archives, ReLOAD (Repository for Linked Open Archival Data), a project of regesta.exe, takes data from the Istituto dei Beni Culturali Regione Emilia Romagna and the Archivio Centrale dello Stato and makes it available as Linked Open Data.<sup>10</sup> Leveraging archival descriptions in the form of XML EADs and authority records in EAC-CPF, the project's aim is to display the connections between the material using authority records and georeferencing. Not conceived of as a portal, the current iteration of ReLOAD has developed a shared space for archival description metadata, which can at this phase can be browsed by various facets.

The WWI Linked Data project, which received an honorable mention in the competition, also seeks to mash up archival data. In its initial stages, the project is developing an ontology for describing events and linking places and agents to documents. The specific dataset under consideration is an online collection at the University of Colorado Boulder, and the project grew out of practical concerns about resource discovery in the University environment.

Rounding out the Challenge finalists, Free Your Metadata is an outreach project focused on building best practices and providing training for the creation of Linked Open Data.<sup>11</sup> A collaboration between Multimedia Lab (ELIS Ghent University / iMinds) and MaSTIC (Université Libre de Bruxelles), the project attempts to make LOD accessible to the LAM

---

<sup>7</sup> <https://thepund.it/>

<sup>8</sup> <http://linkedjazz.org/>

<sup>9</sup> <http://mismuseos.net/comunidad/metamuseo>

<sup>10</sup> <http://labs.regesta.com/progettoReload/en/>

<sup>11</sup> <http://freeyourmetadata.org/>

community. As an example they provide a step-by-step participatory exercise using metadata downloaded from the Powerhouse Museum, so users can follow the process of metadata cleaning in action.

#### **4. LODLAM Education and Outreach**

This last project touches on the broader issue of the need for education about LOD principles and applications, also reflected in one of the sessions at the Summit devoted to the topic of teaching LODLAM. One of the educational initiatives discussed during the session was the IMLS-funded project lead by the University of Washington's iSchool titled Learning Linked Data. With different types of audiences in mind — from professionals seeking to understand Linked Data on the job to library and information science (LIS) students and educators — the project identified key topics to introduce the theoretical and technical aspects of LOD development and envisioned a learning environment including videos, micro-tutorials, and virtual courses.

In addition to this more traditional pedagogical approach, the Learning Linked Data project also emphasized the importance of peer-to-peer community engagement—which has been part of LODLAM since the beginning. LODLAM has been actively engaged in community building and outreach with regional meetings, workshops, web presence and social media. The upcoming LODLAM Training Day, which will be part of the 2014 Semantic Technology & Business Conference,<sup>12</sup> is one example of the type of informal education taking place, while the Twitter hashtag #LODLAM, heavily used for sharing news and projects related to LODLAM, demonstrates the effective use of social media. The LODLAM Challenge itself was an integral part of the efforts to educate the community about LOD by showing compelling real-world applications. Its relevance extends beyond the event as the volume of the documentation generated by those projects offered a valuable resource for those interested in learning about LOD in the LAM context.

#### **Conclusion**

LODLAM is a growing movement engaged in building the foundation for a new knowledge environment where rich information held in memory institutions and cultural organizations around the world can be connected and shared. A self-organized and distributed community, LODLAM has evolved through an array of initiatives and events, from biannual international summits to regional meetings, that bring together professionals, researchers, and enthusiasts from different cultural contexts and organizations. The ultimate goal and drive of this movement is to implement the radical vision of LOD, where digital cultural heritage becomes an integral part of an extended and expanded web in which data can be discovered and used in both new and anticipated ways. While LODLAM is too novel, multifaceted, and complex to be condensed in a single account, this paper has sought to highlight some of the key aspects and phases of the development of this community.

---

<sup>12</sup> <http://semtechbizsj2014.semanticweb.com/LODLAM>

## Acknowledgments

I would like to thank Alexandra Provo for her assistance in the preparation of this paper.

## References

- Berners-Lee, T., 2009. *The Next Web*, TED2009. [video online] Available at <[http://www.ted.com/talks/tim\\_berners\\_lee\\_on\\_the\\_next\\_web.html](http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html)> [Accessed 25 June 2014].
- Bizer, C., Heath, T., and Berners-Lee, T., 2009. Linked Data—The Story So Far. *International Journal on Semantic Web and Information Systems* 5(3), pp. 1–22 [preprint] Available at <<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>> [Accessed 25 June 2014].
- Neubert, J. and Pohl, A., 2011. *SWIB11 28. - 30. November 2011 in Hamburg; Focus: Scholarly Communication in the Web of Data* [conference announcement] Available at <[http://swib.org/swib11/index\\_e.html](http://swib.org/swib11/index_e.html)> [Accessed 24 June 2014].
- Pattuelli, M.C., Miller, M., Lange, L., Fitzell, S. and Li-Madeo, C., 2013. Crafting Linked Open Data for Cultural Heritage: Mapping and Curation Tools for the Linked Jazz Project. *Code4Lib Journal*, [online] Available at <<http://journal.code4lib.org/articles/8670>> [Accessed 25 June 2014].
- Voss, J., 2012a. Radically Open Cultural Heritage Data on the Web. In: *Museums and the Web, Museums and the Web 2012*. [online] San Diego, California, 11-14 April 2012. Available at <[http://www.museumsandtheweb.com/mw2012/papers/radically\\_open\\_cultural\\_heritage\\_data\\_on\\_the\\_w](http://www.museumsandtheweb.com/mw2012/papers/radically_open_cultural_heritage_data_on_the_w)> [Accessed 28 June 2014].
- Voss, J., 2012b. LODLAM State of Affairs. *Information Standards Quarterly*, [online] 24 (1/2) Available at <<http://www.niso.org/publications/isq/2012/v24no2-3/voss/>> [Accessed 23 June 2014].
- Weibel, S.L., 2005. Border Crossings: Reflections on a Decade of Metadata Consensus Building. *D-Lib Magazine*, [online] 11 (7/8) Available at <<http://www.dlib.org/dlib/july05/weibel/07weibel.html>> [Accessed 26 June 2014].