# Linked Open Data for Artistic and Cultural Resources

**Allana Mayer**

**Abstract**—Linked Open Data (LOD) initiatives are in early stages, even in major cultural hubs, and it is hard to quantify intangible expectations such as improved use of materials, enriched resources for researchers and educators, or increased public awareness for participating institutions. Collections of art and culture objects are a major component of these endeavors. This article defines some key terms, notes some important considerations, details exemplary work in the field, and suggests how an ideal LOD framework may look in the near future.

## INTRODUCTION

Linked Open Data (LOD) is defined by standards and technologies that make information machine-readable, freely accessible, and easily relatable. Since 2009, when Tim Berners-Lee published the rough guide to LOD, and Ross Singer wrote a manifesto arguing for its use in libraries, cultural heritage institutions have struggled with admitting that their resource-sharing mandates will necessitate use of these technically challenging measures.[1] LOD is more frequently found in the case of government data repositories, but arts and culture collections are beginning to adopt the standard.

One example of an LOD mechanism is DBpedia, which scrapes information from *Wikipedia* based on advanced queries. A use case from the DBpedia website describes the process as follows: "One nice thing about *Wikipedia* is that it is kept up-to-date by a large community. Therefore, if you need a table on your Web page with, say, German cities, African musicians, Amiga computer games from the 90s, or whatever, you could generate this table with a SPARQL query against the DBpedia endpoint, and your table will stay up-to-date as *Wikipedia* changes."[2]

Allana Mayer is a recent graduate of the MLIS program at McGill University, Montreal, Quebec, and is currently working as an independent researcher and archivist in Toronto, Ontario; allana.mayer@mail.mcgill.ca.

1. Tim Berners-Lee, "Linked Data-design Issues," 2009, http://www.w3.org/DesignIssues/LinkedData.html; Ross Singer, "Linked Library Data Now!" *Journal of Electronic Resources Librarianship* 21, no. 2 (2009).

2. "Use Cases," DBpedia, 2014, http://wiki.dbpedia.org/UseCases.

*Art Documentation: Journal of the Art Libraries Society of North America*, vol. 34 (spring 2015)
0730-7187/2015/3401-0001 $10.00. Copyright 2015 by the Art Libraries Society of North America. All rights reserved.

An ideal repository would function similarly to *Wikipedia* itself: many users can enter data, based on their research and knowledge, and every piece of data will be registered with the contributing user's name and the date of contribution. Libraries' home catalogs will dynamically refer to the centralized source of information, so that every time something is updated, everyone receives it immediately—but they can also "travel back in time" to research changes and improvements throughout the history of the database.[3] There is room for some philosophical debate as to whether contributing privileges would be granted only to those with certain authorities (e.g., employees of academic institutions) or open to all with some members' work dedicated to oversight and approval. Regardless, the bulk of contributions would come from collections staff who have the best knowledge of cultural artifacts and materials, and so the onus is on galleries, libraries, archives, and museums (GLAMs) to mobilize LOD efforts.

As Becky Yoose and Jody Perkins assert, "libraries, archives, and museums have accumulated an embarrassment of riches in the form of unique digitized resources and structured data as well as unmined, unstructured content, all of which are lying fallow inside a Web of documents and untapped relationships."[4] In order to achieve a level of metadata interoperability that allows researchers cross-collection access to institutional information, LOD is an essential framework.

Opening up one's metadata sets in order to do cross-referencing with other collections has untold benefits. An institution's discoverability will be greatly improved for researchers using digital tools. A metadata set with uncertainties or coarse-grained descriptions can be fine-tuned and expanded, and links to similar materials in other collections will be discovered, developing a global picture of cultural holdings.[5] An institution can improve its outreach and engagement services by starting conversations with new digital users across the world, and even insert itself into discussions with practitioners in related industries—not just art and culture historians and educators but creators, journalists, philosophers, and media producers.[6] Barbara Tillett states that "library data need not be just in the form of a citation in a bibliography or other linear listing, but the descriptive and authority data can be re-used and packaged in creative new ways that can be context-appropriate to a user's needs."[7]

Linking library data to other disciplines "means that [librarians] need to make connections between library data and data that [have their] origins in other communities and resources, whether these come from scientific research, government data, commercial information, or even data that has been crowd-sourced."[8] With this col-

3. See the W3C guide at http://www.w3.org/TR/2014/NOTE-ldp-bp-20140828/ for some best practices on a Linked Data Platform.

4. Becky Yoose and Jody Perkins, "The Linked Open Data Landscape in Libraries and Beyond," *Journal of Library Metadata* 13, no. 2–3 (2013): 209.

5. Barbara Tillett, "RDA and the Semantic Web, Linked Data Environment," *JLIS.it, Italian Journal of Library and Information Science* 4, no. 1 (January 2013): 139; Karen Coyle, "Linked Data: An Evolution," *JLIS.it, Italian Journal of Library and Information Science* 4, no. 1 (January 2013), 53–61.

6. Craig A. Knoblock, Jose Luis Ambite, Aman Goel, Shubham Gupta, Kristina Lerman, Maria Muslea, Pedro Szekely, Mohsen Taheriyan, and Parag Mallick, "Semi-Automatically Mapping Structured Sources into the Semantic Web," in *The Semantic Web: Research and Applications* (Heidelberg: Springer, 2012), 375–90.

7. Tillett, "RDA and the Semantic Web," 140.

8. Coyle, "Linked Data," 58.

laboration, metadata sets will become more accurate and comprehensive over time, meaning the value they hold for researchers, artists, students, and patrons can only increase. Linking metadata sets in a centralized hosting or dynamic referencing scheme will benefit institutions and users even more—for one, the redundancies in storage of descriptive data in multiple locations can be eliminated, reducing associated costs.[9]

## CORE CONCEPTS

The Open Knowledge Foundation defines a piece of data or content to be "open" if it is available online, freely, to anyone without special request.[10] While institutions may require a user to create an account in order to access data or content, the data or content themselves must come without restrictions on use, subject at most to a requirement to preserve the attribution and post a share-alike clause.[11] While a library might buy proprietary MARC records en masse for the majority of its collection, its arguably more valuable data will be what it creates for individual items in rare books, archives, and artifactual collections—these are the data most needed in an open format. A set of licensing options by the Open Data Commons allows for data to be labeled as "open."[12]

Linking one's metadata is a more vague term and encompasses several levels of reference or collaboration.[13] For example, imagine two opposing models: a centralized repository is created to host data of multiple institutions, or each local catalog pulls missing information from the others as required. The concept behind, for example, the LIBRIS cooperative catalog of the Swedish Union in 1972 was "that every library contribute what was unique to them, but still gain access [to] the sum of descriptions from all member libraries."[14] This catalog was linked, but not open: only participating libraries' staff had access to collections metadata within it until 1997.[15] Many institutions do not collaborate on catalogs such as this, but create lower levels of linking through providing manual references to other collections' materials. It can be as simple as inserting a Uniform Resource Identifier (URI) into a metadata record under a "Related Material" field, or using an established dictionary or vocabulary to ensure the open metadata will appear in relevant searches. These methods are all levels of ensuring retrievability on the web.[16] The extent to which metadata is inherently linked is variable; given a sophisticated database with intelligent query processing, any metadata can be "linked," even without use of a controlled vocabulary. Using a standardized metadata structure such as the Resource Description Framework

9. Tillett, "RDA and the Semantic Web," 140.

10. Open Knowledge Foundation, "Open Definition," 2013, http://opendefinition.org/.

11. Ibid. A share-alike clause states that if a user were to re-publish the content or use it in a new derivative work, the re-publishing or new work must come with the same lack of restrictions on its use.

12. Open Knowledge Foundation, "Open Data Commons," 2014, http://opendatacommons.org/licenses/odbl/.

13. Berners-Lee, "Linked Data Design Issues."

14. Martin Malmsten, "Cataloguing in the Open: The Disintegration and Distribution of the Record," *JLIS.it, Italian Journal of Library and Information Science* 4, no. 1 (2013): 417.

15. Ibid.

16. Berners-Lee, "Linked Data Design Issues."

(RDF), or the RDF-specific query-formatting language SPARQL, may not even be a requirement for machine parsing.[17]

Berners-Lee created a ranking system for all datasets online to achieve status as LOD, paraphrased here as states of increasing adherence:

1. Available on the web (in any format) with an open license
2. Available as machine-readable structured data
3. In a non-proprietary format (e.g., a CSV list instead of an Excel spreadsheet)
4. Using open standards from W3C (RDF and SPARQL) to identify things
5. Linked to other people's data to provide context.[18]

In parallel with this rating system, the American Art Collaborative recommends a series of steps to ensure one's data qualify for LOD status: prepare a complete set of data; relate it to an existing or emerging ontology; map it to an open machine-readable standard, preferably RDF; link it where possible to external hubs of data; and publish.[19]

As definitions go, these are both circular—in order to be called "linked data," one must link the data. It is relatively easy to automate the assignment of URIs to one's own data, and thus be linkable, but the process of pointing to resources in other LOD collections will be more arduous, and success will depend on the collections chosen, the complexity desired, and the level of manual involvement.

To be open is not necessarily to be linked, and to be linked is not necessarily to be open, but the ideal is to be both.[20] In contrast, a "data silo" is a newly-pejorative term for what was once a standard: metadata sets stored locally, in isolation, usually maintained and accessed internally in a given institution.[21] To keep one's data in a silo is to limit both reviewing and editing, and thus stand in the way of improvements and additions. What was once the standard in libraries exchanging Machine-Readable Cataloging (MARC) bibliographic records for local storage, or worse, physical card-catalog inserts, is now an outdated method of work.[22]

Various institutions describing similar objects can collaborate easily by employing standardized metadata formats and authoritative vocabulary sets, and publishing their records online for reference. For instance, a rare book will have its own physical descriptors of wear and tear or unique characteristics, but its edition-specific publication data will be identical for all of its physical manifestations in various collections. If a librarian with limited research resources cannot identify the binding materials or printing location with certainty, he or she can simply refer to the metadata set of a

17. See the W3C guides to RDF and SPARQL for more information: http://www.w3.org/TR/sparql11-overview/ and http://www.w3.org/RDF/.

18. Berners-Lee, "Linked Data Design Issues."

19. Diana Folsom, Rachel Allen, Shane Richey, Eleanor Fink, and Pedro Szekely, "The Blossoming of the Semantic Web: Linked Open Data and the American Art Collaborative" (presentation at the Museum Computer Network 2013 Annual Conference, Montreal, Quebec, November 22, 2013), http://americanartcollaborative.org/info/slides-the-blossoming-of-the-semantic-web/.

20. Yoose and Perkins, "The Linked Open Data Landscape," 198.

21. Tom Heath and Enrico Motta, "Revyu: Linking Reviews and Ratings into the Web of Data," *Web Semantics: Science, Services and Agents on the World Wide Web* 6, no. 4 (2008): 266–73; Folsom, "The Blossoming of the Semantic Web."

22. Tillett, "RDA and the Semantic Web," 139.

corollary item in another's collection. In this way, leaders have emerged as metadata authorities, often institutions with large budgets and resourceful catalogers.

"Tombstone data" is static, factual information about an object. In the example of a rare book, tombstone data could be author, publisher, date of publication, title, number of (intended) pages, or object dimensions. Without a major historical discovery (such as discovering an author was writing under a pseudonym) this data essentially is seen to be set in stone. This is the best sort of data to share with others: it will be consistent across all physical objects created in multiples. The alternative to tombstone data is dynamic information that changes over time or with cultural perceptions or individual biases. This would include categorization according to genre, mood, or style; user-contributed tagging; internal collections information, which is subject to change; ownership history, provenance, or donor information; fluctuating condition and preservation information; or market value.[23]

## EXAMPLES OF LOD INITIATIVES

Some authorities in the GLAM community have begun to make it easier for institutions to join the fold.[24] Many controlled vocabularies have been easy to push into linked data, including the *Library of Congress Subject Headings* (LCSH)[25] and the *Virtual International Authority File* (VIAF) hosted by the Online Computer Library Catalog (OCLC).[26] The Library of Congress has been working on its Linked Data Service since 2009, when it released the LCSH as linked data; the VIAF aggregates data from a number of institutions and provides links to *Wikipedia* articles to reinforce its relevance.

More recent examples include the opening and linking of both content and data by the Getty Research Institute,[27] the National Gallery of Art,[28] and the Rijksmuseum.[29] These institutions not only released materials on their websites but encouraged users and creators to develop mobile applications, software, and new artworks using the collections. The Getty in particular is partway through a staged release of its major controlled vocabulary sets as LOD. The *Thesaurus of Geographic Names* (TGN) came in August 2014, after the *Art and Architecture Thesaurus* (AAT) in February. James Cuno, in *The Getty Iris* blog, describes possible applications:

> To show how Linked Open Data from TGN can enhance research, let's take the single example of Ellora Caves in Maharashtra, India, a UNESCO World Heritage Site famed for its astonishing rock-cut architecture. TGN contains not only the caves' location but also their geographical hierarchy, variant names in multiple languages, and the religious traditions represented there. Now imagine

23. Coyle, "Linked Data," 53–61; Folsom et al., "The Blossoming of the Semantic Web."

24. Chris Bizer, Richard Cyganiak, and Tom Heath, "How to Publish Linked Data on the Web," 2007, http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/.

25. "Linked Data Service," *Library of Congress*, 2014, http://id.loc.gov/.

26. Coyle, "Linked Data," 58.

27. "Getty Vocabularies as LOD," 2014, *Getty Research Institute*, http://www.getty.edu/research/tools/vocabularies/lod/; "Open Content Program," 2014, *The J. Paul Getty Trust*, http://www.getty.edu/about/opencontent.html.

28. "NGA Images," *National Gallery of Art*, 2014, https://images.nga.gov/.

29. "Rijksstudio," *The Rijksmuseum*, 2014, https://www.rijksmuseum.nl/en/rijksstudio.

> that this data is linked to other data—such as maps, books and articles, and photographs depicting this location. A vast trove of interrelated resources, currently only findable individually through manual search using variant spellings, becomes [a] click away.
>
> Within the Getty alone, in a future Linked Open Data world multiple resources could be interlinked: a digitized volume from the early 1800s from the special collections of the Getty Research Institute; art historically significant early photographs of the site by English, French, and Indian photographers in the collection of the Getty Museum; and multiple publications from the Getty Conservation Institute including an update on conservation efforts.[30]

Eleanor Fink and Erik T. Mitchell provide excellent histories of LOD initiatives for art and culture collections, including the Getty's development of the *Union List of Artist Names* (ULAN), the AAT, and the TGN, essential tools for describing works of art.[31] Datahub (www.datahub.io), a listing for open and linked datasets run by the Open Knowledge Foundation, does not categorize its content but does provide an "art" tag where one can review datasets such as those mentioned above (Figure 1).

In many ways large-scale initiatives such as the Digital Public Library of America (DPLA) or Europeana are not just content-indexers but champions of LOD, as they create metadata schemas to which their contributing institutions must adhere and work to provide blanket interoperability and cross-collection research opportunities.[32] There are as yet no case studies detailing the costs and learning curves for participation in these initiatives, and further research is required to document the efforts by institutions of various sizes and under various constraints. Lisa Gregory and Stephanie Williams reported on their experiences with forming a service hub (an aggregate of local institutions' data for contribution to the DPLA) and underscored the need for simplified metadata requirements to ensure a robust level of contribution.[33] It appears that for many institutions, contemplating an LOD release will necessitate a review and clean-up of existing data to fill gaps and edit for compliance with controlled vocabularies.

Until standards are implemented that have clear plans for iterative improvements, it will be hard to argue for an LOD project that may undergo unforeseen changes as kinks are worked out of the system. Maintenance of a dataset will be reduced considerably with collaboration, but only if an institution trusts in the authority of its collaborators and can be certain of reliability and validity.[34] The limitations on materials and collections information are numerous—it is difficult to open up data that has

30. James Cuno, "Getty Thesaurus of Geographic Names Released as Linked Open Data," *The Getty Iris*, August 21, 2014, http://blogs.getty.edu/iris/getty-thesaurus-of-geographic-names-released-as-linked-open-data/.

31. Eleanor Fink, "Art Clouds: Reminiscences and Prospects for the Future," *Proceedings of the Digital World of Art History 2012* (Princeton University, July 12, 2012), http://ica.princeton.edu/digitalbooks/digitalday2/fink.pdf; Erik T. Mitchell, "Chapter 3: Three Case Studies in Linked Open Data," *Library Technology Reports* 49, no. 5 (2013): 26–43.

32. For more information, see the DPLA's metadata application framework at http://dp.la/info/developers/map/.

33. Lisa Gregory and Stephanie Williams, "On Being a Hub: Some Details behind Providing Metadata for the Digital Public Library of America," *D-Lib Magazine* 4, no. 7/8 (2014), doi:10.1045/july2014-gregory.

34. Malmsten, "Cataloguing in the Open," 420.

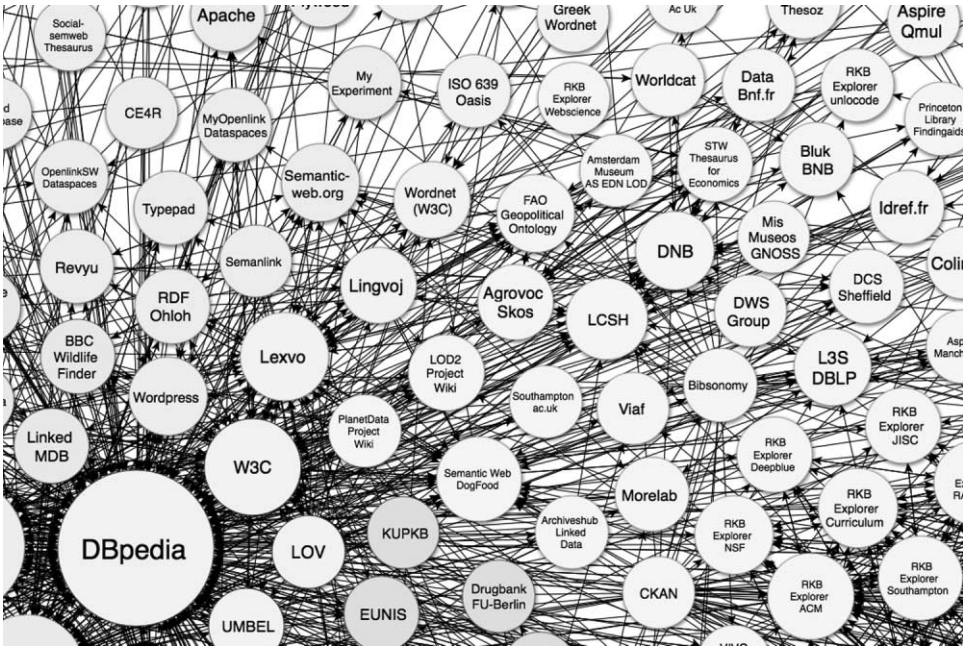


**Figure 1.** A visualization of some of the datasets available on Datahub.io, showing sources such as DBPedia, the LCSH, and Revyu. Linking Open Data cloud diagram, 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak (http://lod-cloud.net/). Used under a Creative Commons By-Attribution Share-Alike license. Please see the online edition of *Art Documentation* for a color version of this image.

restrictions from donor agreements, privacy, or copyright.[35] Implementing LOD in large collections with various licensing and access types will be a time-consuming process.

Using a controlled vocabulary with its own linking system for metadata values is half the battle; the other is using keys of established metadata languages that can crosswalk well into collaborative repositories.[36] In a case study of the Smithsonian American Art Museum (SAAM)'s LOD project, which is expanding into a multi-institutional repository under the name American Art Collaborative (AAC), Pedro Szekely et al. note: "First, there is the problem of mapping the underlying data sources with the metadata about the artwork into RDF. . . . There are often attributes or properties of this data that are unique to a particular museum, and the data is often inconsistent and noisy since it has typically been maintained over a long period of time by many individuals. In past work, this mapping process is typically done by manually writing rules or programs to define the mapping."[37]

The SAAM/AAC project is a model of where LOD is heading: it is not just a

35. Ibid.

36. Timothy W. Cole, Myung-Ja Han, William Fletcher Weathers, and Eric Joyner, "Library MARC Records into Linked Open Data: Challenges and Opportunities," *Journal of Library Metadata* 13, no. 2–3 (2013): 163–96.

37. Pedro Szekely et al., "Connecting the Smithsonian American Art Museum to the Linked Data Cloud," in *The Semantic Web: Semantics and Big Data* (Heidelberg: Springer, 2013), 593–94.

conceptual agreement between institutions to share information, but a practical tool and workflow implemented by a larger institution with a standing invitation to smaller collections to participate.[38] The SAAM project began as an internal repurposing of a software utility called Karma, created at the University of Southern California for the scientific community, for formatting 44,000 collection records into LOD.[39] It has since expanded into a multi-institutional project, allowing twelve other museums to use this tool and contribute data to an online hub.[40]

The open-source Karma tool pushes metadata into RDF from spreadsheets or XML. It is a semi-automated tool that proposes links, which are then verified by a human user. Karma is written to be self-learning and improve iteratively, given which proposed links are approved by the user. So far the SAAM reports links to the Getty, the Rijksmuseum, the *New York Times*, and DBPedia, to name a few.[41] Karma has a function to save unlinked data, then wait for other institutions to release their own data, at which point Karma will attempt to reconnect. Karma's remote interface allows any number of institutions to upload, link, and edit data to the shared database. As Szekely et al. relate:

> For the Smithsonian, the linked data provides access to information that was not previously available. The Museum currently has 1,123 artist biographies that it makes available on its website; through the linked data, we identified 2,807 matches to DBpedia. They can now embed the *Wikipedia* biographies into their collection information, increasing the biographies they offer by 60%. Via the links to DBpedia, they now have links to the *New York Times*, which includes obituaries, exhibition and publication reviews, auction results, and more. They can embed this additional rich information into their records, including 1,759 Getty ULAN identifiers, to benefit their scholarly and public constituents.[42]

The SAAM staff plan next steps that include improving *Wikipedia* articles with the proven authority of the institution, improving geography-based links, and linking to contemporary social media content.[43] The American Art Collaborative initiative that has resulted from the Karma project will eventually be able to offer project grants for participating institutions to initiate data conversion, and even do trial conversions of small datasets to help identify processes and issues.[44] The Smithsonian, in spearheading and guiding this project, has created a system that caters to its specific constituents, the twelve smaller institutions with fewer resources but important American art collections and data to share.[45]

---

38. Ibid.
39. Ibid.
40. "American Art Collaborative," *American Art Collaborative*, 2013, http://www.americanartcollaborative.org.
41. Animesh Manglik, "Karma: A Data Integration Tool," 2013, http://www.isi.edu/integration/karma.
42. Szekely et al., "Connecting the Smithsonian American Art Museum to the Linked Data Cloud," 605.
43. Folsom et al., "The Blossoming of the Semantic Web."
44. Ibid.
45. "American Art Collaborative."

## CHALLENGES

Patron demands have changed significantly over the past decade. Karen Coyle states that contemporary users "expect to do their research and interact with information without prior training, preferably using a single search box [and interacting] with the library through software and hardware that is not under the library's control. . . . To today's users 'access' means 'obtain a copy,' and 'obtain a copy' means that the resource is removed from the organizational context of the library or the database or the web site; every user has a hard drive full of documents that have no particular organizational context."[46]

Without undertaking surveys and user testing, it is impossible to say whether patrons will continue to prefer localized (i.e., downloaded) collections of content, or whether the trend of curated materials online (e.g., Tumblr and Pinterest) might signify a change in behavior towards trusting online repositories to keep content safe and available, preserving some degree of context. The question has greater implications in research, where one might prefer to save a dataset offline so that the accompanying analysis in articles can show fidelity to what was available at the time (even if the data have improved since). It would be possible to provide part of this through extensive date-stamping of data upon contribution, as long as previous values were also maintained. Being able to access obsolete information is an essential part of an LOD framework, but infrequently discussed in the literature.

Allowing users localized "dumps" of datasets will create not only the same redundancies of individual institutions holding siloed data, but also inconsistencies if a researcher refers to an obsolete version of a record while searching. The static forms of traditional publishing methods will create de facto obsolescences: the moment an article is printed, it poses a risk to researchers who reference it instead of the dynamic dataset.[47]

There are other technical issues with having a single source for data: the more people depend on it, the more it will be in demand, and it will be harder to provide remote editing permissions for multiple users simultaneously. Datasets cannot be downloaded for remote editing without risking multiple versions.[48] In collaborative-editing environments such as *Wikipedia*, multiple users can edit one article, but this can be a complicated process that involves many suggested edits and one overseeing editor to approve or combine suggestions into a final draft.

A repository must also avoid the bandwidth load of query/response environments wherein users "refresh" their view of a given set of metadata, in expectation of an eventual edit or improvement that may not come.[49] Downloading metadata repeatedly, looking for changes, is inefficient; notification systems instead can be created so that users or automated systems can be informed when an edit does occur. Fine-grained control must be offered so that edits of a single field or key-value pair can be done without affecting the access to or authority of the remainder of the data.[50]

46. Coyle, "Linked Data," 54–55.
47. Malmsten, "Cataloguing in the Open," 419–20.
48. Ibid.
49. Ibid.
50. Ibid.

As with any wiki, there is a fear of an aggregated, "democratic" information system taking precedence over empirical, established truths. If enough people have editing access, even tombstone data stands a chance of being manipulated or distorted. With the greater data complexities tied to simpler and more intuitive user interfaces that are characteristic of the Semantic Web, "control becomes a matter of trust, not technology."[51]

Meanwhile, the naming and addressing of data elements must become more complex. For example, a painting may be assigned a creation date of "circa 1500s" at first, but later updated to an exact year of 1506 to reflect more accurate research. Queries might by default look for the latest value of the "creation date" element, unless given a specific point in time to view (e.g., when the record was first created), but a user may also wish to see all values over time and what else was changed in conjunction, and who was responsible for the improvements. This is similar to the online availability of government regulations that allows a researcher to see what version of a law was in effect during a particular court case. The creation of such meta-metadata (information about what the metadata is and where it came from) may be incorporated into existing schemas or necessitate entirely separate records.

As many institutions create interfaces with data-curation functionality, more researchers will build virtual datasets within the host environment. An example is the Rijksmuseum's Rijksstudio environment, which allows users to create multiple sets of artworks and cultural objects. This reduces data duplication and has the added bonus of providing institutions with valuable usage metrics on their data.[52] Presumably, similar curated collections can be made for open data.

With the wide gap in linking methods, the environment is not yet at the point where one needs to control for a loss of authority in institutional data. Manual entry of links to related resources is still common, and it takes up much of a cataloger's time and energy. Unfortunately, machine-automated linking is still of variable reliability. The Karma data integration tool used by the SAAM and AAC (described above) functions in an iterative process of semi-automated link suggestions which are then manually verified by human editors, and purports to be self-learning, improving with each piece of feedback.[53]

The other obvious downside to manual linking is that some institutions may simply not have persistent-linking abilities on their websites. URLs often change with the implementation of new categorizing software or search interfaces. For many libraries in particular, the string-based authority control of MARC records is incompatible with the URIs more frequently in deployment today. As Timothy Cole et al. found, MARC records needed to be crosswalked to MODS (Metadata Object Description Schema) in order to be URI-compatible and thus LOD-friendly.[54]

Tom Heath and Enrico Motta, implementing LOD on a reviews website called

51. Ibid.

52. Ed Rodley, "What We Discovered at Museums and the Web," 2013, *Peabody Essex Museum,* http://connected.pem.org/exploring-museums-and-the-web/.

53. Knoblock et al., "Semi-Automatically Mapping Structured Sources into the Semantic Web," 378.

54. Cole et al., "Library MARC Records into Linked Open Data," 165.

Revyu, tackle the functionality of URIs, a functionality that is defined by their proper and frequent use:

> All things represented on Revyu are assigned URIs: reviews, people, reviewed things, tags assigned to things, and even the bundles that represent tags assigned by one person at one point in time. Providing URIs for all reviewers and reviewed things gives many items a presence on the Semantic Web which they would not have otherwise, and enables any third party to refer to these items in other RDF statements. This "linkable" data creates the potential for inward links to Revyu from other data sets. All URIs in the Revyu URI-space can be dereferenced. Attempts to dereference the URIs of non-information resources receive an HTTP303 "See Other" response containing the URI of a document that describes the resource. This . . . serves to reinforce the distinction between a resource and a description of that resource, as each has a distinct URI.[55]

An essential part in an implementation of LOD is the inclusion of Cross-Origin Resource Sharing (CORS) headers in client-side websites and APIs (application programming interfaces, an advanced way to automate data access).[56] Any dataset online has the potential to be exposed through external websites that link to (and often curate) that data, but many are restricted de facto by browser restrictions which assume that links between websites in this way are dangerous and unwanted (such as someone automating access to all content, which could overload a server). In the case of some security restrictions for institutions, this is true; other, truly open, data repositories must employ the CORS headers framework.[57] Like LOD, CORS implementation can also vary. A CORS statement can be open to anyone, or a specific list of trusted linking websites or else a CORS statement can be automatically generated for every webpage in a domain, or vary from page to page.[58]

Other concerns that cannot be enumerated completely within the scope of this article include the institutional desire for usage metrics. While the definition of "open" information provided by the Open Knowledge Foundation does not discuss collecting statistics on a patron's use of data, it is often implemented practically without issue.[59] Collecting basic information on a visitor to an institution's website, for example, is an accepted practice. While some institutions, such as the Getty, ask that patrons maintain attributive information on derivative forms and inform them of their uses of the materials that the Getty offers, this is not a formalized system of attribution. However, the Rijksmuseum has found that people freely attribute their copied materials back to the source. As a staff member from the Rijksmuseum expressed, it seems that explicit association with the Rijksmuseum is a mark of quality, and a point of pride, for patrons.[60]

---

55. Heath and Motta, "Revyu," 269.

56. Anne van Kesteren, ed., "Cross-Origin Resource Sharing," 2013, http://www.w3.org/TR/cors/.

57. Ibid.

58. Ibid.

59. Open Knowledge Foundation, "Open Definition."

60. Geertje Jacobs, "To Open or Not to Open? A Technical, Legal, or Philosophical Question" (presentation at the Museum Computer Network 2013 Annual Conference, Montreal, Quebec, November 22, 2013).

Return-on-investment assessments are harder to quantify; whether the increase in public awareness and materials use translates to a justification in an institutional budget has yet to be proven. Larger institutions certainly benefit, but smaller institutions and collections that are largely copyrighted (e.g., modern art or design materials) will have very little content that can be opened or linked, making the decision one worth adjourning. Any collaborative effort should offer institutions feedback on their contributions.

## SUGGESTIONS FOR FUTURE WORK

The above stated concerns about the ambiguities in LOD status must be addressed. Several summits and conferences, and the most notable partners and initiatives, are working hard to determine the practical implementations of LOD. These overseeing bodies must remain accountable to member institutions' needs and ideas as they develop streamlined frameworks for participation.

It is easy to envision a wiki-like environment with several levels of membership: for institutions which possess the physical collections and thus have authority over certain parts of records; for users who wish to suggest improvements to records or create customized collections or sets of data for their own research interests; editors, most likely subject experts, who oversee changes and regulate modifications (e.g., settling arguments over a value's relevance or factuality); and technicians who oversee formatting and perform webmaster duties.[61]

Perhaps a solution will look more like cloud-computing documents collaboratively edited. One user will be able to see other users' cursors as they type, watching them work and interacting with them in real time. In this scenario, instead of sourcing a given data change to a single user, some changes may be the product of discussion or negotiation between two or more parties interacting virtually.

It is, of course, possible to have a collection entirely protected by restrictions of donation or privacy. In archives it is common to receive an accession which must remain closed for a certain period of time after the accession date, or after an unspecified date such as the eventual death of the donor. Often donor agreements focus on the content of the materials, not the metadata about them. In this case it would be perfectly acceptable to release finding aids online knowing full well that access to the materials themselves would be out of the question. It does become tricky ensuring that biographical sketches or scope and content information do not inadvertently offer restricted information—performing batch processing on finding aids and metadata sets must be sure to take these subtleties into account. In these cases donor agreements must be actively improved to contain explicit instructions about metadata accessibility.[62] Being proactive and vigilant about restrictive demands may decrease accessions or exhibition opportunities but will vastly simplify the data- or content-sharing process—and it will improve an institution's standing in the eyes of its users.

61. Fink, "Art Clouds."

62. Dinah Cardin, "Policing Pictures in the Gallery," 2013, *Peabody Essex Museum*, http://connected.pem.org/policing-pictures-in-the-gallery.

These sorts of concerns, in terms of openness and access of institutional materials, are long overdue.

## CONCLUSION

With conversion to LOD systems seemingly inevitable, there needs to be more information about the process and its costs, how to assess one's collections for LOD viability, and how to value, use, and determine a return on investment.

With LOD still in its infancy, the DPLA, DBPedia, and the American Art Collaborative have not yet released public reports of patron use or feedback, or studies on usability for their datasets or materials. The functions of an LOD framework for multiple institutions must answer the questions of multi-author editing, how to reference data over time, the complexity of URI assignation, the use of CORS and other security measures, the availability of a centralized data-curation method for researchers, and the level of standards required for contributing data.

Discussions about these plans need to happen in an open environment where, much like with LOD, institutions can learn from one another. Institutions with the most resources must lead the way and clearly document case studies for others to take into account.

## ADDITIONAL SOURCES CONSULTED

de Boer, Victor, Jan Wielemaker, Judith Van Gent, Michiel Hildebrand, Antoine Isaac, Jacco Van Ossenbruggen, and Guus Schreiber. "Supporting Linked Data Production for Cultural Heritage Institutes: The Amsterdam Museum Vase Study." In *The Semantic Web: Research and Applications*, 733–47. Heidelberg: Springer, 2012.

Guibault, Lucie, and Christina Angelopoulos, eds. *Open Content Licensing: From Theory to Practice*. Amsterdam: Amsterdam University Press, 2011.

Matsumura, Fuyuko, Iwao Kobayashi, Fumihiro Kato, Tetsuro Kamura, Ikki Ohmukai, and Hideaki Takeda. "Producing and Consuming Linked Open Data on Art with a Local Community." In *COLD*. 2012. http://ceur-ws.org/Vol-905/MatsumuraEtAl_COLD2012.pdf.

Méndez, Eva, and Jane Greenberg. "Linked Data for Open Vocabularies and HIVE's Global Framework." *El profesional de la información* 21, no. 3 (2012): 236–44.