

Linking to Linked Data

Natalia Ermolaev, Princeton University,
Michael Lesk, Rutgers University,
Danielle Reay, Princeton University,

Abstract

Does very precise description of objects and semantic-web linking truly enable new applications for catalog data? This poster presents research testing this questions using data from a variety of library, museum and archive collections.

Keywords: linked open data, interoperability, museums

Citation: Ermolaev, N., Lesk, M., Reay, D. (2015). Linking to Linked Data. In *iConference 2015 Proceedings*.

Copyright: Copyright is held by the author(s).

Acknowledgements: Our thanks to the British Museum and to Dominic Oldman of their staff for providing data and assistance.

Research Data: In case you want to publish research data please contact the editor.

Contact: nataliae@princeton.edu, lesk@acm.org, dreay@princeton.edu

1 Introduction

The "linked open data" movement encourages the expression of information in the form of triples: rather than hierarchical or relational models, assertions about an object are expressed in a predicate-subject-object form. For example, "Jane Austen wrote *Pride and Prejudice*" is expressed in the British National Bibliography (with respect to one specific edition of the book) as:

```
person/AustenJane1775-1817  
blterms:hasCreated  
http://bnb.data.bl.uk/id/resource/015594626
```

Triples provide a more flexible way to model complex data, which helps resolve ambiguities in descriptions.

2 Mapping

For example, consider the task of comparing the origins of objects acquired in a certain museum by century. We used the same British Museum data expressed both in the traditional catalog format as well as linked open data. When comparing the choropleth maps we produced (using place names as found in the catalog descriptions vs. geographic names in the semantic web ontology), we found that the formal representation did a noticeably better job. For example, there are several places named "Memphis" in the world, and also several named "Rochester". The largest city with each of those names is in the United States, but objects in the British Museum from those places are usually from Egypt and England, respectively. The linked open data catalog has a geographic hierarchy and disambiguates the names for each object.

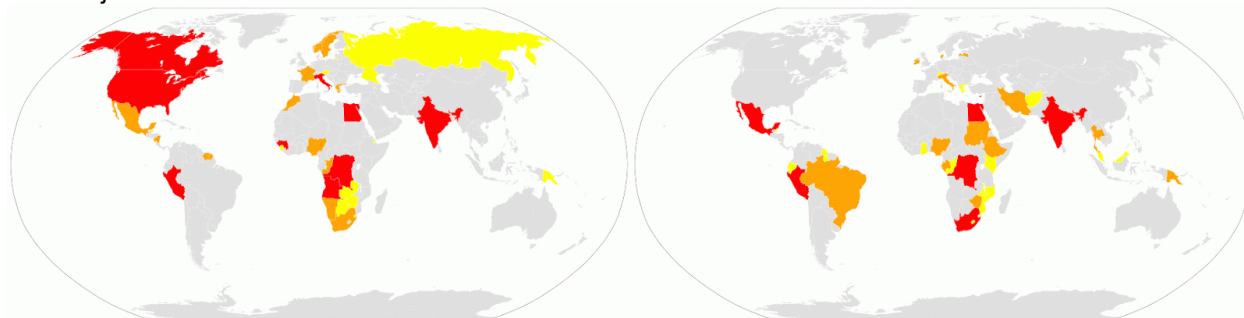


Figure 1. This choropleth map compares the 18th century (left) and 19th century (right); the countries from which most objects came are in red, the next most important are in orange, and the countries from which still fewer objects came are in yellow.

3 Roles and Linkages

The formal data representation can also identify the *role* of a particular person or object in a specific context, and do so very precisely [Doerr]. Figure 2 shows two pairs of images. On the left is a portrait of Joseph Pennell, drawn by Whistler, and then a print drawn by Joseph Pennell and depicting the Statue of Liberty. One can easily imagine a user needing to search or browse separately for one or the other. Similarly the right pair of images is a drawing of Salisbury published in London combined with a view of Wiltshire made in Salisbury. Although some of these distinctions would be made in conventional cataloging, think of a coin which has the name CAESAR inscribed and a portrait; is Julius Caesar the person under whose authority the coin was struck or the person depicted? In both cases the name will appear in the same form and would be part of the same catalog field in a short catalog record. The British Museum has relations for “Associated Person (Depicted)” and “Production (Moneyer)” to separate these meanings.



Figure 2. Left to right: “Firelight: Joseph Pennell, No. 2” by James McNeill Whistler, 1896. “Hail America - the Statue of Liberty,” by Joseph Pennell, 1910. “Salisbury, this view of the cathedral,” published by Francis Jukes in London, 1798. “Cottage beside a bridge near Stratford-sub-Castle...,” drawn by John Constable, 1821. All images from the British Museum.

An important goal of linked open data is interoperability: making connections between different information resources. As an example, we looked at a number of modernist art magazines in the Blue Mountain Project at the Princeton University Library [Brooke]. Figure 3 is a sample page, on which Joseph Pennell’s name appears.

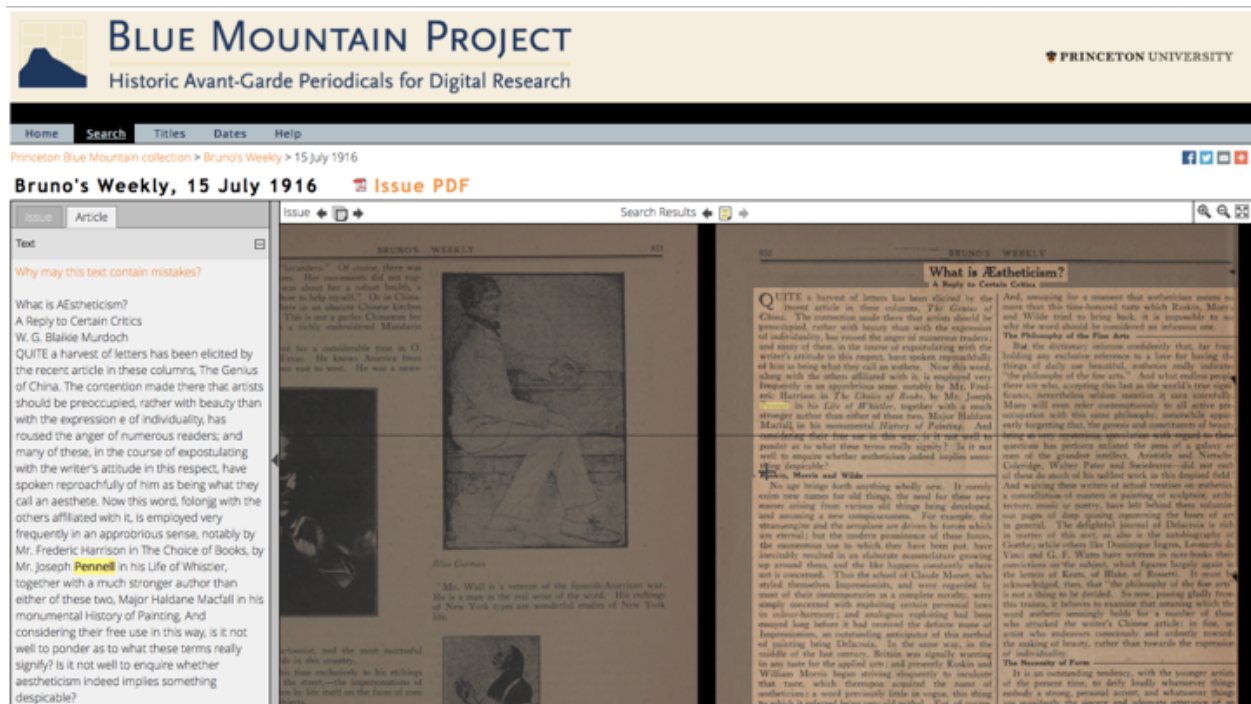


Figure 3. A page from *Bruno's Weekly*. OCR text on left, page images on right.

We located names of artists from this issue and produced a network graph that represents their co-authorship in the same issue. We then downloaded catalog data from several museum collections - the British Museum, the Museum of Modern Art, the Tate Gallery, the Guggenheim, and the Cooper-Hewitt - and found examples of art produced by these artists. Adding images allowed us to enhance our network graphs, and to develop a visual argument about the influences and collaborations within this creative community. Figures 4 and 5 illustrate the same exercise with artists who wrote for the modernist journal *Broom*. The ability to connect open catalog data between library and museum collections can help establish a more comprehensive and nuanced picture of cultural, political and social movements, and also reveal how objects have been collected and maintained by cultural memory institutions.

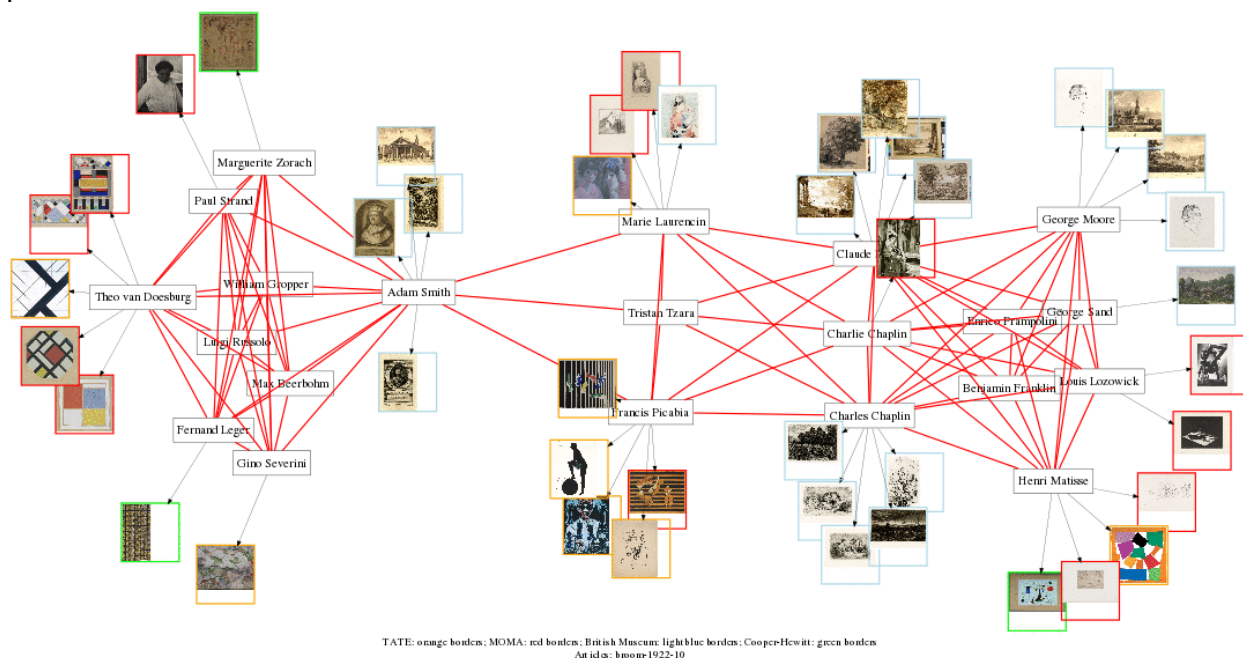


Figure 4. In this illustration, artists are linked if they are mentioned within a span of $\frac{1}{3}$ of an issue of *Broom*; this issue is October 1922. Illustrations of their works are retrieved from different museum websites. The color of the outline of the images signals the museum from which the image came.

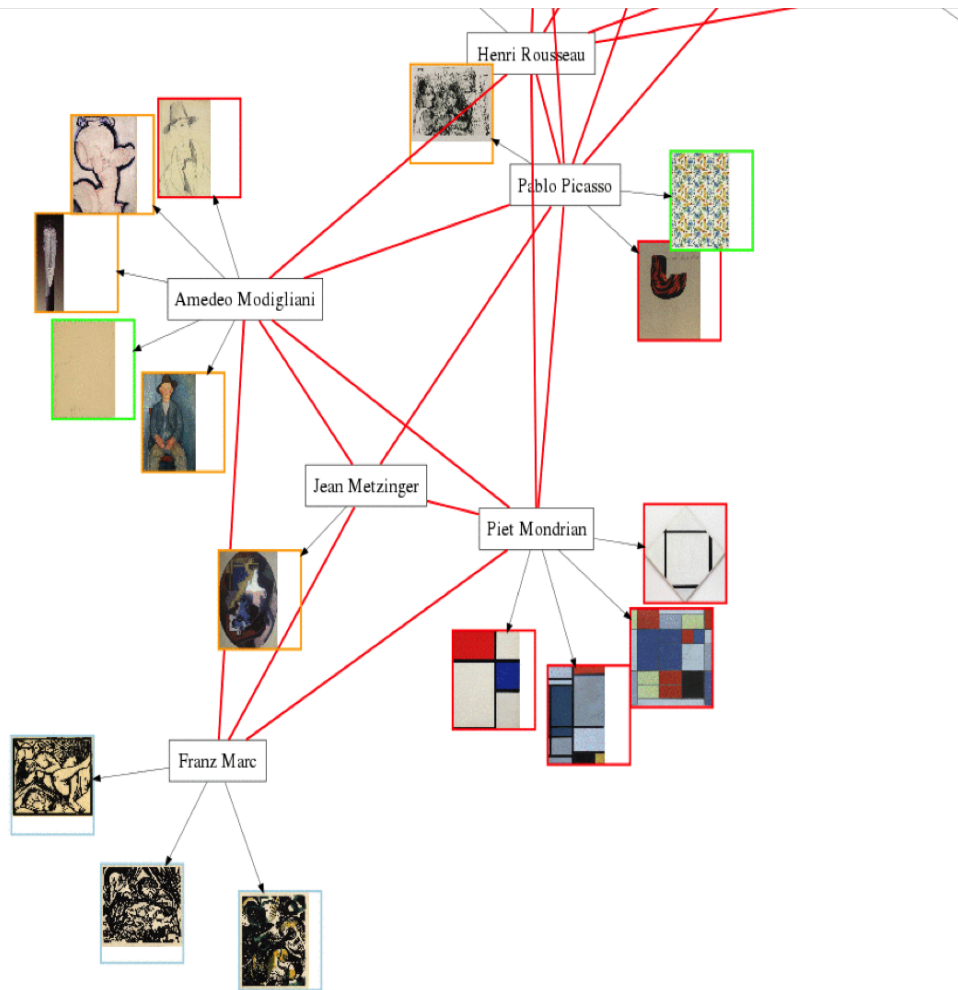


Figure 5. Partial linkups between artists from overlaps within a catalog published by the Guggenheim Museum.

The formal knowledge representation facilitates more complex research inquiries because it disambiguates roles such as draftsman, engraver, or printer. It enables other kinds of sophisticated exploration, such as seeking objects from a common source.

4 Complexities

Unfortunately, the level of standardization and collaboration that would make such interoperability a reality has not yet been reached. The use of formal knowledge representation may even act at cross purposes. Different systems or institutions use different ontologies, or have conflicting conceptual frameworks, that make it hard (or impossible) to translate from one catalog to another. For example, LCSH (Library of Congress Subject Headings) places “Haida Art” underneath “Canadian Art” (the Haida are a First Nations group who live in British Columbia). The Getty Art and Architecture Thesaurus places “Haida (culture or style)” underneath first “Northwest Coast Native American styles” and then “Native North American styles.” . A “broader” search from Haida objects using LCSH would first go to Ontario or

Quebec painters from European culture, while using the Getty ontology you might first go to Navajo weavings.

Moreover, the more complicated a knowledge representation is, the more likely it is that such conflicts will arise. Note, for example, that the formal ontologies in the British Museum catalog, which follows the CIDOC-CRM standard, are different from the ones in the British Library catalog. In the British National Bibliography, James Abbott McNeill Whistler is

bnb.data.bl.uk/id/person/WhistlerJamesMcNeill1834-1903

while in the British Museum he is

collection.britishmuseum.org/id/person-institution/50934

and neither record connects to

viaf.org/viaf/46804212

which is his Virtual International Authority File entry. These are all exact specifications but need to be linked for effective interoperability. They do, however, distinguish this painter from others (the unrelated artist Rex Whistler, for example).

5 Conclusion

We are continuing to explore the possibilities and limitation of linked open data. We see advantages in

- name resolution through authority lists: which *Memphis*? which *Whistler*?
- finding broader terms through ontologies: look for *Egypt* instead of *Memphis*, or *stone* instead of *granite*.
- knowing a role: painter vs. subject, or printer vs. engraver.
- distinguishing the stages of creation: composer vs. arranger vs. performer.
- interoperability: finding the same concept in different collections

However, what is both the greatest advantage of linked open data may also be its biggest challenge: full and rigorous knowledge representation is both time-consuming and difficult to create. With such complexity, will different organizations faced with the same object reliably describe it the same way? Will they commit to following established standards and ontologies that are very precise and detailed? Linked open data, like many technologies, has rewards and costs. The balance is still being weighed.

6 Table of Figures

Figure 1. This choropleth map compares the 18th century (left) and 19th century (right); the countries from which most objects came are in red, the next most important are in orange, and the countries from which still fewer objects came are in yellow.

Figure 2. Left to right: “Firelight: Joseph Pennell, No. 2” by James McNeill Whistler, 1896. “Hail America - the Statue of Liberty,” by Joseph Pennell, 1910. “Salisbury, this view of the cathedral,” published by Francis Jukes in London, 1798. “Cottage beside a bridge near Stratford-sub-Castle...,” drawn by John Constable, 1821. All images from the British Museum.

Figure 3. A page from *Bruno’s Weekly*. OCR text on left, page images on right.

Figure 4. In this illustration, artists are linked if they are mentioned within a span of $\frac{1}{3}$ of an issue of *Broom*; this issue is October 1922. Illustrations of their works are retrieved from different museum websites. The color of the outline of the images signals the museum from which the image came.

Figure 5. Partial linkups between artists from overlaps within a catalog published by the Guggenheim Museum.

7 Acknowledgments

We appreciate the help of the British Museum, specifically Dominic Oldman, and of the Princeton University Library.

8 References

Brooke, Sandra. 2013. "Marquand Library and the Blue Mountain Project," *Art Libraries Journal*, vol. 38, no. 1, pp. 38-42.

Doerr, Martin. 2003. "The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata," *AI Magazine*, vol. 24, no. 3, pp. 75-92.