A Demonstration of Tools for Building Linked Data for the American Art Collaborative

Craig A. Knoblock[1], Pedro Szekely[1], Eleanor Fink[2], Duane Degler[3], David Newbury[4], Robert Sanderson[4], Kate Blanch[5], Sara Snyder[6], Nilay Chheda[1], Nimesh Jain[1], Ravi Raju Krishna[1], Nikhila Begur Sreekanth[1], and Yixiang Yao[1]

[1] University of Southern California
[2] American Art Collaborative
[3] Design for Context
[4] J Paul Getty Trust
[5] The Walters Art Museum
[6] Smithsonian American Art Museum

Abstract. This is the demonstration for the paper titled, "Lessons Learned in Building the Linked Data for the American Art Collaborative" by Knoblock et al. We will demonstrate the complete set of tools that were used to build the Linked Data, including the Karma software for mapping data to the CRM ontology, the Mapping Validator to evaluate whether the data was correctly and consistently mapped, the Link Review Tool to evaluate the links to other resources, and the Browse application to review and explore the integrated Linked Data repository. This demonstration complements the research paper by providing a live demonstration of the approach and tools discussed in the paper.

1 Introduction

There is growing interest in Linked Open Data (LOD) among museums and the cultural heritage sector. In recent years it has gained traction because most museums are interested in using technology to reach new audiences, collaborate with other museums, deepen research, and help audiences of all ages experience, learn about, appreciate, and enjoy art. These concepts and others that characterize features of LOD inspired 14 art museums to form a collaborative to learn about and implement LOD within their respective museums and set the stage for the broader art-museum community to explore LOD.

One of the key goals of the American Art Collaborative (AAC)[7] is to create and publish a critical mass of LOD drawn from the collections of the 14 museums that will be made available on the Internet for researchers, educators, developers, and the general public. Towards this goal, we built 5-star Linked Data for the museums by applying existing tools and developing new tools where needed to map and link the data. We will demonstrate the end-to-end process for creating the Linked Data for the museums.

[7] http://americanartcollaborative.org/

2 Mapping the Data

In previous work we developed the Karma information integration system,[8] a semi-automated tool for mapping data sources to a domain ontology. Karma has a machine learning capability to provide recommendations on the mappings to an ontology and has an intuitive graphical interface for visualizing and refining mappings. Figure 1 provides a fragment of a screen shot of the use of Karma to map one of the datasets to the CRM ontology. In this project, we made many improvements to Karma to support the extensive modeling effort. In this presentation, we will demonstrate how Karma is applied to map an example dataset to the CRM ontology and highlight the new capabilities in Karma.

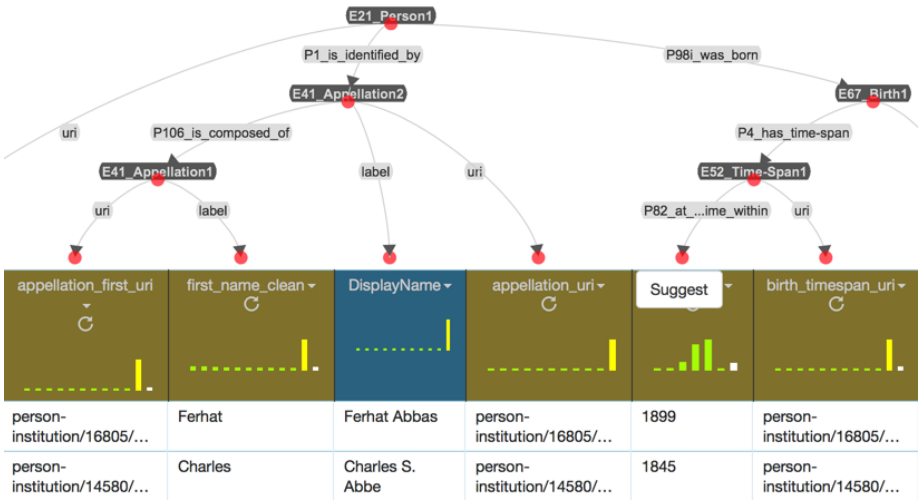

Fig. 1. Screen shot of Karma building a mapping of the National Portrait Gallery data

3 Validating the mappings

To address the challenges in creating consistent and correct mappings, we developed the AAC Mapping Validator,[9] which is shown in Figure 2. We will demonstrate how the tool provides a target mapping for each type of information and a corresponding query that will return a set of data if the data has been correctly mapped to the domain ontology.

4 Linking and Reviewing the Data

Museums take enormous pride on the quality of their data, so they want 100% correct links. They were willing to manually review *every* link before publication, so we developed a workflow where an automated algorithm first proposes

[8] karma.isi.edu
[9] review.americanartcollaborative.org

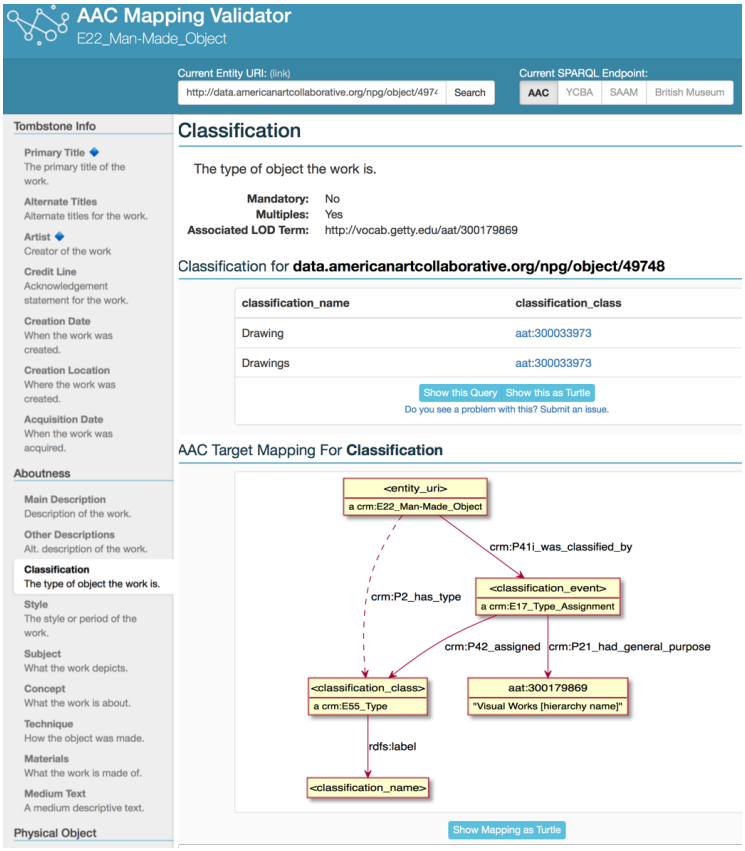

Fig. 2. The Mapping Validator specifies the target mapping and queries the triplestore based on the mapping to verify that the mapping is done correctly

links and a human curator verifies each link. The automated linking algorithm produced 24,733 links that needed to be reviewed by museum personnel. We will demonstrate the generic Link Review Tool[10] shown in Figure 3, which is optimized to support efficient and accurate comparison of pairs of records.

5 Using the Data

As part of this effort, we developed the Browse application, which allows the museums to review their data and for other users to explore the data by institution, artists, and categories. We will demonstrate the Browse application.[11]

[10] linking.americanartcollaborative.org
[11] browse.americanartcollaborative.org

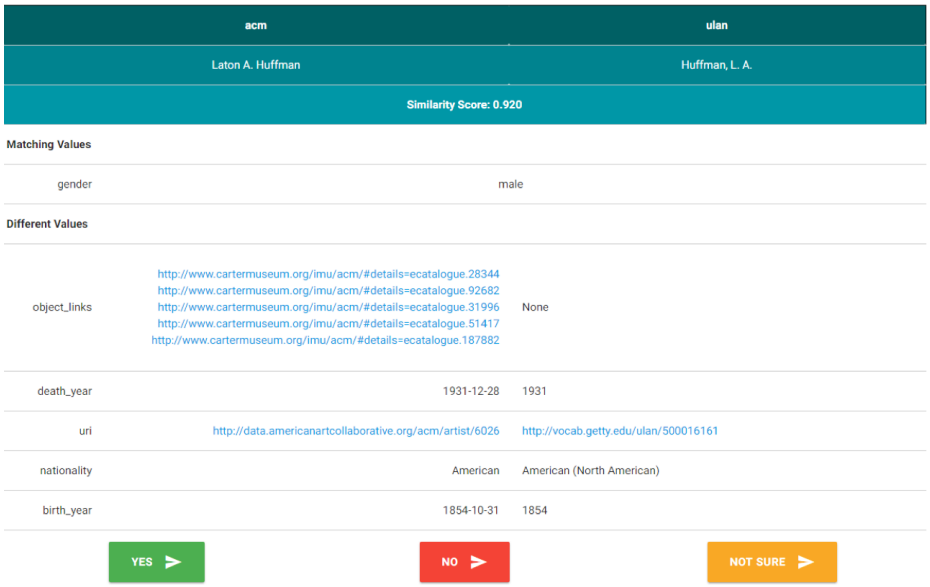

Fig. 3. Screenshot of the Link Review Tool

6 Discussion

Building a consistent and accurate set of Linked Data for the 14 American art museums was a challenging task. To support this task, we developed and employed an effective set of tools that allowed us to successfully complete the project. First, we used Github as the infrastructure to manage the raw data, the mappings to the CRM ontology, the links of the artists to the Getty ULAN, and the published RDF data. Second, we extended our existing Karma software to simplify the mapping process and integrate directly with Github, automatically publishing the R2RML mappings, a visualization of the mappings, and the RDF data directly to Github. Third, we developed the Mapping Validator, which provides a target model for the various types of information provided by each museum and executes SPARQL queries based on the target model against the triplestore to ensure that the data is correctly mapped. Fourth, we developed the Link Review Tool, which allowed the museum personnel to review each of the links of their artist data to the Getty ULAN to ensure that the artists were correctly linked. Finally, we developed the Browse application, which provides a variety of useful views of the data and allows the museum personnel to review the final set of data to ensure it is correct and complete.

Acknowledgements

The American Art Collaborative was made possible by grants from the Andrew W. Mellon Foundation and the Institute of Museum and Library Services.