

“The research is happening in the text fields” – Are Linked Open Data and Art History a good match?

Nanni, Giacomo

giacomo.nanni@fh-potsdam.de
UCLAB, University of Applied Sciences Potsdam

Freyberg, Linda

l.freyberg@dipf.de
DIPF Leibniz-Institut für Bildungsforschung und
Bildungsinformation, Germany

de Günther, Sabine

sabine.de.guenther@fh-potsdam.de
UCLAB, University of Applied Sciences Potsdam

Dörk, Marian

marian.doerk@fh-potsdam.de
UCLAB, University of Applied Sciences Potsdam

Introduction

The scientific discourse in object-focused and historical disciplines follows certain methods and builds a narration with objects, events, evidence and conclusions.

In Digital Humanities (DH) research this process and its specific structure has to be represented on a data level. Linked Open Data, Semantic Web ontologies and, in particular, reference models like CIDOC CRM¹ (2022) offer a precise structural vocabulary to describe cultural heritage objects and their relations. These frameworks can support GLAM² institutions to make the data on their objects findable, accessible and interoperable.

However, in research projects the course of developing arguments and describing findings is crucial. We use Art History research, in particular, the project “Restaging Fashion - Digital contextualization of vestimentary sources” (Refa), as one example to show how emerging discourses within the research process can be mapped in a knowledge organization system and which aspects may not be formally represented. A recurring question is: *What should be modeled as linked open data (LOD) and what should be entered as free-form text?* Despite the proven utility of data standards, vocabularies, and authority files providing for accuracy, context, and unambiguity, there is also a need for an open space to formulate and present tentative observations. With this contribution we present how we bridge LOD and full text descriptions through information visualization.

Semantic data modeling in Museums

Semantic data modeling of a collection during an early stage of the research prompts a trade-off to be made between data standardization and semantic accuracy. Sophisticated ontologies like CIDOC CRM encourage the detailed description of an object through an event-driven structure, which makes it a very powerful tool for the cultural heritage sector by providing for the possibility of adding meaning to the original data.

Since 2006, when the term *Linked Data* was coined (see Berners-Lee 2006), according to Daquino et al. “only a few pioneers have abandoned legacy cataloguing and archiving systems to fully embrace the Semantic Web paradigm and manage their catalogues through LOD-native management systems” (2022).

When considering the LOD datasets published by The British Museum and the Rijksmuseum, who also create tools such as ResearchSpace (Oldman / Tanase 2018) and dedicated platforms to engage with their collections, it appears that these museums only share relatively shallow data models for their artworks on the web. The Rijksmuseum provides basic metadata on objects structured in DC³, EDM⁴ and LIDO⁵ schemas. Only one example is modeled deeply: *Linked Art van Gogh object metadata download*. But the creators clearly state that: “both the implementation of the model as well as the use of identifiers will be subject to change” (Rijks-Data 2022). The British Museum, on the other hand, structured its entire collection using CIDOC CRM, but was unable to preserve its database in a consistent and accessible manner (Lincoln, 2015). Data that should be linked together is no longer accessible, making the research of scholars no longer available (see Programming-historian 2017).

The potential of these applications is obvious: LOD allows data to be reused and redistributed, revealing relationships with other data on the Web (Berners-Lee 2006). The resources involved, however, are somewhat excessive both from a technological perspective (see De Decker 2015) and regarding the cataloging of information. Manual tagging can be performed by domain experts in a very precise manner, but it cannot be performed manually to large volumes of already existing content due to the lack of resources (Simou et al. 2017) or may not be conducted automatically due to inconsistencies and/or the lack of standardization in the datasets. This leads institutions to publish only a subset of the available information, exposing minimal metadata (Doerr et al. 2010), which might prevent researchers from using the data and constrains them to conventional research methods.

Data-driven art historical discourse

Art historical research on and with collections is deeply concerned with the history of objects and their discourse such as provenance history, history of interpretation, dating or attribution based on comparable works. This can be expressed with the event structure of CIDOC CRM including descriptive information on the objects like dimensions, inscriptions, and alternative titles also linked to relevant vocabularies like ICONCLASS⁶ and AAT⁷. But a crucial part of humanities research consists of annotating that basic formal information, by elaborating and contextualizing the history of the artworks and their interpretation, often concluding in the form of a paper or monograph.

Undertaking this part of research in the form of LOD is very intricate and laborious. Complex ontologies will never be completely exhaustive since no cultural artifact or experience could be fully and finally contextualized and vocabularies are not equally accurate and detailed for every subtopic. In addition, contexts and meanings shift over time, even during the timeframe of a research project. According to Galloway "there are some things that are unrepresentable. And the computer is our guide into that realm" (2011: 91). In terms of representing the research process it is not worth making this part *machine*-readable as a basis for formal queries. It is more important to present it in a way that is accessible and understandable to other researchers, which does not require this high level of formalization, because it would end up in overcomplicated queries.

Visualizing context

One way to equally accommodate formal LOD structures as well as full text descriptions is through custom interfaces. In particular, information visualization can be used for the contextualization of images, sources, texts and their relations.

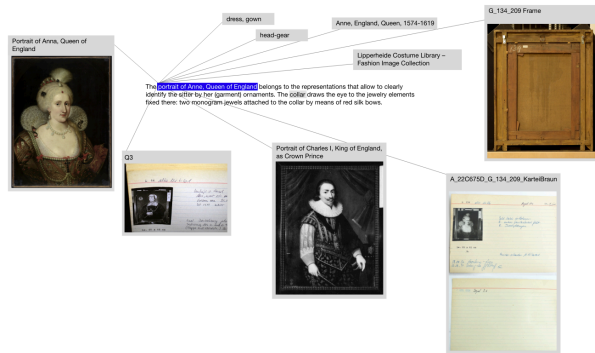


Fig. 1: Draft of Refa interface

As part of the Restaging Fashion project, we devised different experimental prototypes, linked specific parts of the textual description to artworks, to classification entries, to archival sources as well as to other cultural heritage objects (see Fig. 1.) with the aim to establish an overview of the context.

Instead of hiding the full text in unstructured note fields, we approach this as a starting point for the semantic contextualization of the research process.

In conclusion, certain aspects of humanities research can be properly expressed by semantic web structures. In particular quantitative information, for instance, on objects and their relations, which can be queried and expressed statistically, are worth to be modeled in triple structures. Other, more elaborated, reflexive, or contextualized information may require different forms of representations like visual interfaces.

Notes

1. CIDOC Conceptual Reference Model
2. Galleries, Libraries, Archives and Museums
3. Dublin Core
4. Europeana Data Model
5. Lightweight Information Describing Objects

6. <https://iconclass.org/>

7. <https://www.getty.edu/research/tools/vocabularies/aat/>

Bibliography

Berners-Lee, Tim (2006): *Linked data* . <http://www.w3.org/DesignIssues/LinkedData.html> [26.04.2023].

CIDOC CRM (2022): Definition of the CIDOC Conceptual Reference Model version 7.2.2. https://cidoc-crm.org/sites/default/files/cidoc_crm_version_7.2.2%5B20%20Oct%5D.pdf

Daquino, Marilena / Wigham, Mari / Daga, Enrico / Gianolini, Lucia / Tomasi, Francesca (2022) : *CLEF: A Linked Open Data native system for Crowdsourcing* . <https://arxiv.org/abs/2206.08259> [26.04.2023].

Doerr, Martin / Gradmann, Stefan / Hennicke, Steffen / Isaac, Antoine / Meghini, Carlo / van de Sompel, Herbert (2010) : " The E uropeana D ata M odel (EDM)", i n: IFLA (ed.): *World Library and Information Congress: 76th IFLA general conference and assembly*, 1 0: 10-15.

De Decker, Kris (2015): "Why We Need a Speed Limit for the Internet: The energy use of the internet can only stop growing when energy sources run out, unless we impose self-chosen limits." in: Lowtech Magazine, 10, 2015 <https://solar.lowtechmagazine.com/2015/10/can-the-internet-run-on-renewable-energy.html> [26.04.2023].

Galloway, Alexander R. (2011): *Theory, Culture & Society*. Los Angeles: SAGE, 28, 7- 8: 85- 1 02. DOI: 10.1177/0263276411423038.

Lincoln, Matthew (2015): "Using SPARQL to access Linked Open Data" in: The Programming Historian.

Oldman, Dominic / Tanase, Diana (2018): "Reshaping the knowledge graph by connecting researchers, data and practices in R esearch S pace", i n: *International Semantic Web Conference. Lecture Notes in Computer Science book series (LNISA)* 11137: 325-340. Cham: Springer. https://dx.doi.org/10.1007/978-3-030-00668-6_20.

Programminghistorian (2017): *British Museum SPARQL endpoint failing again #569* . <https://github.com/programminghistorian/jekyll/issues/569> [26.04.2023].

RijksData (2022): <https://data.rijksmuseum.nl/object-meta-data/download/> [26.04.2023].

Simou, Nikolaos / Chortaras, Alexandros / Stamou, Giorgos / Kollias, Stefanos (2017): "Enriching and Publishing Cultural Heritage as Linked Open Data", i n: Ioannides, M arinos / Magneat-Thalmann, N adia / Papagiannakis, G eorge (eds): *Mixed Reality and Gamification for Cultural Heritage* . Cham: Springer, 201-223. DOI: 10.1007/978-3-319-49607-8_7 .