

Cultural Heritage objects, linking their context by ontologies and the FAIR³ data principles

What have the DH to offer, based on the CIDOC-CRM and TEI?

(Keynote)

Christian-Emil Ore
 Department of Linguistic and Nordic Studies
 University of Oslo
 Oslo, Norway
 c.e.s.ore@iln.uio.no

Cultural heritage is a wide and somewhat diffuse concept. UNESCO differentiates between tangible cultural heritage like paintings, sculptures, coins, manuscripts, monuments, archeological sites etc. and intangible cultural heritage like oral traditions, performing arts, rituals etc. The latter cannot be stored in museums or collections. The documentation of intangible cultural heritage is, however, tangible and can be stored.

In museums, archives and libraries, the physical objects are at the centre of attention; hence most of the information in the systems is organized as attributes describing the physical objects. The more interesting information of the provenance and cultural context of the objects is found in archives, grey literature, exhibition catalogues and (academic) publications. Today, at least the most recent, texts are available in digital form, but rarely include 'live' links to the records for the artefacts and objects described. For non-digital texts, the finding aids usually do not give a detailed formalized description of the content. One should also note that the context of objects could be historical events like expeditions, conflicts etc. These are often referred to in the literature and information system, but they are intangible in nature. Unfortunately, there are few authority registers for such events.

In 2001 Berners-Lee, Hendler and Lassila [1] foresaw a second web, the semantic web, readable for computers and based on the RDF-technology. Compared with the traditional web it has not become an undisputable success. Five years later Berners-Lee [2] suggested a more concrete and practical solution called Linked (Open) Data, also based on RDF(S). Linked Data offers a simple and easy way to publish data in an open and uniform interface enabling others to link scholarly data resources. Thus Link Data should be ideal for building resources in the Digital Humanities.

The programmatic slogan of Semantic Web and Linked Data community is "Anyone can say anything about anything". That is, anything can be linked. From a scholarly and scientific point of view, this is not satisfactory. Information is generated through exclusion using meaningful distinctions according to a common conceptual model or formal ontology. Thus meaningful information integration in a scholarly field using the Linked Data mechanism requires a common conceptual model for the context in question.

The Finnish project WarSampo [3] links a large number of datasets concerning Finland in the Winter War and Second World War. In WarSampo, CIDOC-CRM is used as the harmonizing basis for modelling data, with events providing the semantic glue for data linking. This is an elegant example of an advanced LOD application scalable using a common conceptual model designed for data integration. It can be argued that a solution with the 5 LOD stars and the two additional stars (see [4]) supports not only the FAIR (Findable, Accessible, Interoperable, Reusable) [5] but the extended principles FAIR³, that is, with the additional "Relevant and Reliable".

Reproducibility of results is a core concept in text-based research as in all research. The content in information systems and virtual reconstructions in the cultural heritage sector are to a large degree directly based on information deduced from text studies. In many cases, the links from the information system back to the texts are not available, and such links may be difficult to re-establish. How should LOD prepared information, based on a reading of a text, be linked to the text itself? It is important to base such linking on data standards evolved in the fields of text encoding and conceptual modelling. Thus, the understanding of text encoding represented by the TEI [6] guidelines and the understanding of conceptual models represented by initiatives like the CIDOC-CRM [7] and FRBRoo [8] (now LRM[8]) should be combined.

The examples in the lecture will mostly be taken from the event oriented conceptual model, CIDOC-CRM (ISO211/27) and the development of this model.

References

1. Berners-Lee, T., Hendler, J., and Lassila, O. (2001) 'The Semantic Web', Scientific American, 284 (May), pp. 34-43. doi:10.1038/scientificamerican0501-34, last accessed 2019/05/21
2. Berners-Lee, T. (2009) Linked Data. Available at: <https://www.w3.org/DesignIssues/LinkedData.html>, last accessed 2019/05/23

3. War Sampo <https://www.sotasampo.fi/en/>, last accessed 2019/05/23
4. Hyvönen, E., Tuominen, J., Alonen, M., and Mäkelä, E. (2014) 'Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets', in Presutti, V., Blomqvist, E., Troncy, R., Sack, H., Papadakis, I., and Tordai, A. (eds) *The Semantic Web: ESWC 2014 Satellite Events*. Anissaras, Crete, Greece, May 25–29, 2014. Revised Selected Papers. Springer International Publishing Switzerland, pp. 226–230. Series; Lecture Notes in Computer Science 8798. doi: 10.1007/978-3-319-11955-7_24. (https://link.springer.com/chapter/10.1007/978-3-319-11955-7_24)
5. FAIR principles Available at: <https://www.go-fair.org/fair-principles/>, last accessed 2019/05/23
6. TEI = Text Encoding Initiative P5: www.tei-c.org, last accessed 2019/05/23
7. CIDOC-CRM Available at: <http://www.cidoc-crm.org>, last accessed 2019/05/23
8. FRBRoo Available at: <http://www.cidoc-crm.org>, last accessed 2019/05/23
9. LRM = Library Reference Model, Available at <https://www.ifla.org/publications/node/11412>, last accessed 2019/05/23