

Carlos H. Marcondes, Maria Luiza de Almeida Campos

Searching for a methodology to define culturally relevant relationships between digital collections in archives, libraries and museums

Abstract

The emergence of Semantic Web and LOD - Linked Open Data - technologies enable that digital objects representing the holdings of archives, libraries and museums collections be semantically interlinked throughout the Web. What are the different types of cultural relevant relationships that may exist between digital objects of collections in archives, libraries and museums throughout the Web? How discover, organize and formalize these relationships to be used by curators in LOD applications? A methodology to analyze the holdings of archives, libraries and museums is proposed based on onomasiologic perspective. Such methodology is applied to a hypothetical competency question that might be proposed by a digital curator. Results indicated that conceptual models such as FRBR, CIDOC CRM and EDM may provide a rich repertoire of semantic relationships that may be used in LOD applications to interlink collections in heritage institutions.

1. Introduction

For centuries institutions as archives, libraries and museums have the mission of curators of the memory and cultural heritage of societies where they are inserted in. Although their common mission each of these institutions are specialized in different facets of this legacy, thus creating specific methodological procedures and value added criteria concerning the memory and cultural heritage of their holdings each kind of institution is responsible for. The emergence of Semantic Web and LOD - Linked Open Data - technologies enable that digital objects representing the holdings of archives, libraries and museums collections be semantically interlinked throughout the Web thus creating unexpected meaning and contextual networks, empowering their synergies, their complementarities, their educational and curatorial potentialities.

A digital curator, with the aim of a formalized vocabulary of such relationships, could produce culturally rich virtual archive, libraries, museums, expositions and classes, accessible by anybody from anywhere, which explores exploring the increasing number of memory and cultural heritage collections now available throughout the Web. Such technologies enable that a digital curator discovers and makes sense of, or propose new, unforeseen, cross-domain semantic relationships between digital cultural heritage objects.

A curator (from Latin *curare* meaning "take care") is a manager or overseer. Traditionally, a curator or keeper of a cultural heritage institution (i.e., gallery, museum or archive) is a content specialist responsible for an institution's collections and involved with the interpretation of heritage material. The object of a traditional curator's concern necessarily involves tangible objects of some sort, whether it is artwork, collectibles, historic items or scientific collections. (CURATOR, Wikipedia).

Cultural heritage objects hold different types of relationships. A film may be inspired in a literary work, a work of art illustrates an edition of a literary work famous painters who created the scenario and costumes of ballets or assembly plays. There are different versions of Da Vinci's Mona Lisa made by artists such as Marcel Duchamp, Andy Warhol and Fernando Botero. In KO literature such relationships are the associative relationships.

However, due to a long time tradition of independent, self-contained collections, the adoption of different standards, the possibilities of interoperability between such diverse collections are beyond technological issues. In recent years the Documentation as a domain has used conceptual models to identify, make explicit, standardize, and semantically integrate its objects. This paper aims at investigating methodological approaches that enables the discovery of such culturally relevant semantic relationships. Such semantic relationships are no more within the scope of a specific collection's domain but between the content of the digital representations of archives, libraries and museum objects, certainly with holdings in different subject domains.

LOD technologies enable cross-searching such collections. Thus such technologies enable digital objects of different collections to be mobilized by curators in specific domains as Art, Culture, Literature, History, Journalism, Education, Scientific Divulgateion, etc., in order to create a new, unique, curated, digital resource, as a virtual exhibition. Curatorial work is multidisciplinary, hard to delimit, personal, authorial. The exploitation of curatorial potentialities of such LOD resources can be considered domain specific or problem oriented. Consider, for example, exhibitions as "Leonardo Da Vinci: The Mechanics of a Genius"¹, or "Human Bodies The Exhibition"².

This paper communicates some initial findings and insights of an ongoing research that aims at addressing the following questions. From a curator standpoint what are the different types of cultural relevant relationships that may exists between digital objects of collections in archives, libraries and museums throughout the Web? How to discover, organize and formalize these relationships in a vocabulary to be used by curators in LOD applications in Culture? What should be the role of curatorial activities, what should be the role of knowledge organization activities, related to publishing LOD datasets? What Methodology should lead to such aims?

The specific aim of this paper is to discuss methodological approaches to knowledge discovery, organization and representation in order to identify, define and formalize cultural relevant relationships that may exists between digital objects of collections in archives, libraries and museums aimed to be used in LOD applications.

This paper is organized as follows: after this introductory section, section 2 revises theoretical approaches in KO to classify relations and make methodological assumptions about how to identify, define and formalize cultural relevant relationships. Section 3 proposes a case in which a digital curator with the task of organize a thematic virtual exhibition needs to find throughout the Web relations between themes relevant to develop such a virtual exhibition. Section 4 discuss how and where - from what sources - he/she can retrieve and reuse a repertoire of semantic relationships and

¹ Available at http://www.sciencemuseum.org.uk/visitmuseum/plan_your_visit/exhibitions/leonardo

² Available at <http://www.humanbodies.eu/en/the-exhibition/>.

proposes new curated relationships. Finally, section 5 evaluates the results and draws some initial conclusions.

2. Theoretical bases

Information Science, Knowledge organization and Terminology literature exhaustively discusses the nature and meaning of associative relations. Although the vast amount of literature these relations are still controversial.

Within Information Science the so called associative relations rise within the development of thesaurus in 1970 decade. Unlike hierarchical relations, who are logical and abstraction relations, associative relations, including whole-part relations, are *ontic* since they occur between objects. Dalhberg (1978) claims that the establishment of associative relations may depend on specific contexts. Although not adding further details she suggests bases to identify such relations.

Terminological Theory (WÜSTER, 1977) classifies relation in logical and ontic. Within the further there are two subclasses, contact and causality relations. Wüster deeply analyzes these relations. Within the former are whole-part and associative relations. Contact relations are the most important subclass of ontic relations. They are self explanatory from their species, coordination and concatenation relations. The main coordination relation is whole-part. This type of relation may occur between a whole and its parts or between the parts of a whole. It is considered as a relation occurring within a specific spatial location, stressing the fact that a whole and its parts exist simultaneously in a spatial (and time) location. Coordination relations as mentioned by Wüster are the inclusion and integration relations. Concatenation relations are conditioned by time and have as its subclasses precedence and succession relations.

Causality relations are parent relations and have two subclasses: ascending or descending parent relations between two different generations; and phase relations expressing different phases of an individual or substance life. Accordingly phase relations are classified in phylogenic, ontogenic and substance-substance relations (Sales, Campos & Gomes, 2008).

Within Terminology Sager (1990) proposes another classificatory schema to relations, namely: generic, whole-part, polyvalent and complex relations. Generic relations are equivalent to logic relations. Whole-part relations are the same as those proposed by Wüster. Polyvalent relations are equivalent to polyhierarchical relations found in ISO 2788 (1986) and in Aitchison (1987) when discussing hierarchical relations.

Been (2008, p. 156) claims that “Associative relations come into a variety of flavors”. Peters and Weller (2008, 101) claim that “They are unspecified connections of concepts that can have any kind of relation”, “Thesauri make use of (entirely undifferentiated) associative relations”, and that “In addition, associative relations can be split into a diversity of domain-dependent, specified paradigmatic relations”.

Also associative relations are sometimes defined by exclusion of hierarchical or paradigmatic relationships. According to the EuroVoc Thesaurus: “The associative relationship is a relationship between two concepts which do not belong to the same hierarchical structure, although they have semantic or contextual similarities”. Marcia Lei Zeng (2005) also defines them in a similar way: “This relationship covers associations

between terms that are neither equivalent nor hierarchical, yet the terms are semantically or conceptually associated”.

Associative relationships in Knowledge Organization literature are thus dubious and semantically inaccurate (imprecise). They are also highly context dependents. In order to be useful and enable computational inferences these relationships must be specialized and have a clear, unequivocal and formal semantics.

3. Methodological assumptions

According to ICOMOS (2002) “Cultural Heritage is an expression of the ways of living developed by a community and passed on from generation to generation, including customs, practices, places, objects, artistic expressions and values. Cultural Heritage is often expressed as either Intangible or Tangible Cultural Heritage”.

UNESCO, within the scope of Cultural Heritage sites, defines “World Heritage is the designation for places on Earth that are of outstanding universal value to humanity.

A tentative definition of culturally relevant relationships claims that those are relationships holding between representation of cultural heritage objects - digital objects - in archives, libraries and museum collections that are supported, cited, mentioned, discussed, exposed by a socio-cultural event, that is, cited in a publication, in a conference, in a lecture, in a film, in a law, or in any other socio-cultural event.

Heritage objects from collections in archives, libraries and museums all have intrinsic cultural value recognized and attributed by curatorial activities developed within the scope of these institutions. Thus the cultural value of such objects is not an essential property, instead, is socially attributed by these institutions. Socially attributed properties turn out to be incorporated to what such objects *are*, to their essence. Searle (1995) presents an exhaustive discussion about the process of social attribution of properties to objects, especially to artifacts, that is insightful to the understanding of how heritage institutions as archives, libraries and museums attribute cultural values to objects, to what is the social and cultural value of an archive, library or museum object.

Thus the socially attribution of social, cultural and heritage values to objects is a mission and a mandate of these institutions. These facts have consequences on the way these objects are represented as digital and information objects. This deserves further discussion on the values attributed by the society to archival documents, publications and museum objects; however this is beyond the limits of this paper.

Identify culturally relevant relationships is within the scope of activities a cultural curator. Their job is to interpret, attribute new meanings, re-contextualize, etc., the different kinds of cultural manifestations. If a specific interpretation of a cultural manifestation will become culturally relevant or not, will depend upon its recognition and acceptance as one of the accepted interpretation theories for that cultural manifestation. It seems to be a phenomenon similar to citation in scientific communication.

The search for a methodology to identify, define and formalize such relationships is based on the assumptions presented in sequel.

1- Cultural heritage objects belonging to collections in archives, libraries and museums have an “intrinsic” cultural value, attributed by (Searle, 1995) local curators.

2- Due to the subjective character of cultural interpretations, which are individual and authorial we opted not to act within the scope of cultural curatorial activities, which are specific of domains as previously mentioned.

3- Any possible relationship holding between two cultural heritage objects is a consequence of what these object are or were, of its creation, i.e., of its (social) life.

These assumptions suggest the adoption of a methodology which enables the discovery of what a cultural object is or was, of all its possible attributes and relationships, in all the contexts in which such an object has existed during all its life span. Following the previous proposal of Wüster (1978) there may be two types of relationships, logic and ontic. Logic relationship are epistemic, they depend on how someone knows and classifies an object; therefore they are abstract, occurring between concepts. Instead ontic relationships depend on what an object is or was, they occur between objects that are contiguous in space and/or time, ie. have had contact during their life span. These last relationships are fundamental to the methodological approach we are proposing.

A methodology with such a focus is based on the onomasiologic perspective. According to this perspective Language has as one of its functions to build the referent object to a speech community. Language terms reflect and agreed conception of a social object. Accordingly, Language terms reflect the building of a concept, that is, by selecting and highlighting, but also by hiding, different aspects of such a social object. The onomasiologic perspective aims at capturing the “version” of an object that corresponds to how a speech community interacts with such an object.

This proposal aims at enlarging and maximizing the potential cultural value of digital heritage objects. Despite the fact that onomasiologic perspective was developed to other aims it may be useful in describing cultural objects in order to enable possible links between.

Thus, following the methodological perspective outlined, ccultural objects must be exhaustively described, discovering and/or making explicit all its attributes and the ontic relationships between them. If, in doing this, we make explicit and standardize the descriptive items of each object, we are thus potentializing the use of these descriptive items as anchors to semantic links using LOD technologies This may enlarge interoperability between different digital collections in order to answer the competency question posed by curators and have their cultural potential realized.

4. Case analysis

In sequel is presented a hypothetical competency question that serves to guide the inquiry proposed by this paper.

Suppose a curator is in charge of the development of a virtual exhibition dedicate to Brazilian 19th and early 20th century writer, Machado de Assis, and possible relations between his literary style and European literature. He/she needs to be aware of every digital resource about the writer, of his life, of social and literary, historical and social context in which Machado wrote his romances, available throughout the Web and formatted according to LOD technologies. Suppose this curator is using a LOD web browser and virtual exhibition editor that enables the navigation and recovery of records from different LOD information resources and the recording of semantic links established (“curated”) by the curator.

What should be the relationships between the digital representations of two cultural heritage objects beyond associative relationships, mere “semantically or conceptually associated” relationships?

As domain specific, the culturally relevant relationships, as defined, are excluded of the analysis, according to the assumption 1, ontic relationships must be examined. In KO literature these relations are also known as syntagmatic relations, as they occur in space-time realms, independent of how it is organized according to specific paradigms (paradigmatic relations).

Conceptual models such as the FRBR (1998), CRM CIDOC (2013) or Europeana Data Model (2016) exhaustively analyze digital objects from collections in archives, libraries and museums according to onomasiologic perspective and provide a semantically rich and exhaustive inventory of such relationships. The following is a list of some relationships drawn from these conceptual models.

Digital representations of two cultural heritage objects are related if, for ex.:

- both are assigned the same subject, *frbr:has_as_subject*;
- one has as subject the other, *frbr:has_as_subject*;
- both have the same creator, *frbr:is_created_by*;
- both have the same producer, *frbr:is_produced_by*;
- both are or were owned by the same person or corporate body, *frbr:is_owned_by*;
- both share the same context, *frbr:context* for the work;
- one was influenced by the other, *crm:P15* was influenced by;
- one object is illustrated by the other, *crm:P65* shows visual item (*is shown by*);
- one object is inspired on the other, *edm:isDerivativeOf*.

If relationships between digital objects in heritage institutions collections and other Web resources are also considered, as for example, between a digital object and an entry in an Authority file or in Wikipedia/DBpedia, additional relationships must be considered as, for example:

- one object *edm:Has Met* an authority, meaning for example that some object was used by a person in some event.

A search for Machado de Assis in LOD datasets with content provided by Brazilian heritage institutions may retrieve records and full-text of different books written by Machado de Assis, or books of different authors analyzing his works, from library collections. Also may retrieve records and digital images of different photos of Machado or his portraits drawn by different artists and published in newspapers, from archives collections. He/she may also retrieve records and digital images from a writing desk with an ink-glass and a pen that belonged to Machado, and with which he wrote some of his works; these objects belong to the collection of Brazilian Literary Academy’s museum. Also a video may be retrieved from a digital film library, of an adaptation of Machado masterpiece, *Capitu*, to Brazilian television.

The authority and bibliographic records retrieved, about Machado de Assis as author or about his works from library catalogs may inform he is a Brazilian 19th and early 20th century author associated with Realism literary movement, or may *frbr:has_as_subject* Realism. Also Machado de Assis’s entry in the DBpedia informs he is a Brazilian author associated with Realism literary movement; there is an additional entry in DBpedia, “Realism in Brazil” that outlines the role of Machado as

the main expression of this literary movement in Brazil. These entries refer to another, “Literary Realism”, which list the main authors associated with this literary movement in the Americas and Europe.

The relationships founded by the curator are provided by LOD information resources from Brazilian heritage institutions in association with LOD information provided by DBpedia.

Lehmann, Schüppel & Auer (2007) describe a Web application, the RelFinder³, which enables graphic visualization of the direct and indirect relationships, found in RDF/LOD records, between two concepts typed by a user. This application uses DBpedia RDF/LOD records. Relfinder application enables one to foresee the operation of a Web virtual exhibition editor to be used by digital curators.

As a digital curator use such a virtual exhibition editor, he/she can establish the relationship (and thus records it), creating a semantic link with the semantic crm:P15 was influenced by, between an authority record of Machado de Assis and the DBpedia entry “Literary Realism”, thus establishing that Machado de Assis as author was influenced by European Realism literary movement. He/she can also establish a new, *culturally relevant relationship* between Machado and Magic realism⁴ or between Machado and the Portuguese authors as Eça de Queiroz and José Saramago.

5. Concluding remarks

Conceptual models such as FRBR, CIDOC CRM and EDM provide a semantically rich repertoire of relationships that can be used in LOD applications to connect digital objects from collections in archives, libraries and museums. These relationships can be the starting point to more specialized, curated, *culturally relevant relationships*. The competency question proposed could be answered by LOD records from collections in archives, libraries and museums, complemented by records from DBpedia.

The curatorial work in Culture is interested in creators and their works, in cultural movements, tendencies and influences. However archives, libraries and museum hold objects that are partial facets of, or relate to, a creator work or an artistic movement, as the different works of a writer or a painter, personal letters of an individual, objects that belong to him/her. Within this context authority files seems to be relevant. But authority control as traditionally exercised by libraries must be complemented and/or integrated with information provided by entries in Wikipedia/DBpedia.

The methodological approach outlined, the onomasiologic perspective - suggests that instead of searching for, discovering, formalizing and organizing culturally relevant relationships to make sense of the amount of cultural data now available throughout the Web of Data, it is more useful to make explicit and available the features and relationships that exist in and between cultural objects from collections in archives, libraries and museums as a starting point and so these features and relationships can be used by digital curators in doing their job.

References

³ <http://www.visualdataweb.org/refinder/refinder.php>

⁴ https://en.wikipedia.org/wiki/Magic_realism

- Aitchison, J. (1987). *Thesaurus construction: a practical manual*. 2. ed. London: ASLIB.
- THE CIDOC CONCEPTUAL REFERENCE MODEL. (2013). Version 5.1.12, January 2014. [http://www.cidoc-crm.org/docs/cidoc_crm_version_5.1.2.pdf]. Accessed 15 Nov. 2014.
- Dalhberg, I. (1978). A referent-oriented analytical concept theory for Interconcept. *International Classification*, Frankfurt, 5(3), 142-150.
- DEFINITION OF THE EUROPEANA DATA MODEL v5.2.7. (2016). [<http://pro.europeana.eu/documents/900548/0d0f6ec3-1905-4c4f-96c8-1d817c03123c>]. Accessed 15 Nov. 2014.
- ICOMOS, International Cultural Tourism Charter (2002). *Principles And Guidelines For Managing Tourism At Places Of Cultural And Heritage Significance*. ICOMOS International Cultural Tourism Committee.
- IFLA. (1998). *Study Group on Functional Requirements for Bibliographic Records: final report*. München: K. G. Saur. (UBCIM Publications New Series).
- ISO 2788-1986: documentation guidelines for the establishment and development of monolingual thesauri. (1986). 2.ed. [S.l.]: ISO, 32p.
- Lehmann, J., Schüppel, J., & Auer, S. (2007). Discovering Unknown Connections-the DBpedia Relationship Finder. *CSSW, 113*, 99-110. [<http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-301/proceedings.pdf#page=99>]. Accessed 10 May 2016.
- Peters, Isabella; Weller, Katrin. (2008). Paradigmatic and syntagmatic relations in knowledge organization systems. *Information Wissenschaft und Praxis* 59(2): 100-107. <https://www.phil-fak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Informationswissenschaft/peters/1204547334paradigmat.pdf>. Accessed 20 abr. 2015.
- Sager, J. C. (1990). *A practical course in terminology processing*. London: John Benjamins Publishing Company.
- Sales, L. F. ; Campos, M. L. A. & Gomes, H. E. (2008). Ontologias de domínio: um estudo das relações conceituais. *Perspectivas em Ciência da Informação*, 13, 62-76.
- Searle, J. (1995). *The construction of social reality*. New York, United States of America: The Free Press.
- Wüster, E. (1974). *L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, L'informatique et les sciences des choses*. [1974]. French version in RONDEAU, G. & FELBER, F. (org.) *Textes Choisis de Terminologie: I. Fondements théoriques de la terminologie*. Québec: GIRSTERM. p. 57-114.
- Zeng, Marcia Lei. (2005). *Construction of controlled vocabularies: a primer (based on Z39.19)*. NISO, 2005. [<http://marciazeng.slis.kent.edu/Z3919/index.htm>]. Accessed 20 Abr. 2015.