Full paper

Context is all: Guidelines for context characterization in knowledge modeling and data formats

Gioele Barabucci (*), Fabio Vitali (**)

(*) NTNU — Norwegian University of Science and Technology
(**) Alma Mater Studiorum - University of Bologna

Longer abstract

Every recorded piece of data (both analog or digital) exists after having gone through some type of human interpretation. This interpretation is more or less evident, but it is always there, directly or indirectly. In fact, there are always many levels of interpretations, layered one on top of the other. On the bottom of one such stack there is somebody writing a treaty and on top of it there is somebody else writing a commentary about it. On the bottom of another stack of interpretations there is somebody transcribing the wrong word (the famous "soiled fish" in Moby Dick), and on top of it there is somebody writing scholarly articles about dirty fish. A sizable amount of scholarship takes place in and between these layers of interpretations: commentaries, corrections, references, but also commentaries on corrections, commentaries on commentaries on corrections, and so on.

To better illustrate this concept, we present two examples of stratified levels of interpretations, one that focuses on mistakes, and one that deals with how a re-contextualization of a work can influence its scholarly and cultural reception. Both examples show the need for the characterization and preservation of various pieces of "context" and interpretations in the digital representations of these works, in order to faithfully illustrate the scholarly and cultural process around the digital works.

Mistakes. The creation of a cultural work is rightfully regarded as an insightful act worth of commentaries. The act of making or correcting a mistake is, on the other hand, not that valued. However, the act of making a mistake, as well as the identification and the correction of a mistake are all interpretation acts, carried out by a human just like the creation of the work in the first place. In turn, some of these mistakes and some of these corrections will become the subject of other cultural explorations and artifacts. In the case of the "soiled fish" mistake, that mistake has given birth to many books and pop-science articles written about that mistake and about all the scholarly articles that saw in that typo the actual intent of the author. This makes "soiled fish" not just a transcription error that has been made, identified and then fixed, but an interesting (and now integral) part of the cultural discourse. It is not just the mistake itself that is part of the cultural discourse, but also its correction. Each of these pieces of data (the original work, the mistake, the correction) has the opportunity to

become the source of other cultural artifacts, sometimes even more relevant than the original work itself.

Re-contextualization. Not only mistakes and corrections have the ability to sprout deep cultural debates. Every act of interpretation or commentary on an artifact can change its cultural relevance. One example of "impact through commentary" is that of the Mona Lisa, a well-known painting by Leonardo da Vinci, considered just one of the many works by Leonardo until 1913, when it was stolen by Vincenzo Peruggia and later recovered; the chronicle of these events and the echo it had on the readers of mass media outlets of the times turned one of the many paintings of Leonardo into the most iconic painting of the world. This shift in notoriety and allure is not due to any change in the painting itself (the painting was returned undamaged), nor to the interpretation of the artistic value of the work (scholars did not appreciate Leonardo's technique more after the thief). What has changed is the fact that this artifact ended up inside a newly minted network of cultural references built on top of these events that dramatically changed our perception of it.

In this paper we argue that It is important that these acts of modification, curation or reference (e.g. identification of mistakes, their correction, commentaries and commentaries on the commentaries) are not seen as trivial acts: they are acts with the same level of dignity and cultural relevance associated to the underlying cultural works they refer to. To this end, it is fundamental that cultural artifact stored digitally are represented in a way that preserves not only the data itself, but that also the expression of a number of contextual framing characterizations of the work itself that better allow to express the complexity of cultural and historical relevance, impact and relevance that we associate to it (both positively and negatively). We have identified a number of contextual characterizations of artefacts as assertions on the data used to represent them digitally:

- Temporal relationships: when has this artifact been created? when has it been digitized? when has this interpretation of the artifact been given? In other words, the identification of temporal boundaries of the artefact and, in general, of anything we bother expressing in direct or indirect reference to the artifact. This may be precise or not. For instance, we have some clear boundaries for the existence of the Mona Lisa (it did not exist before 1503, it most probably existed in 1516) but we do not have any specific date for its creation.
- Spatial/jurisdictional relationships: where has this artifact been created? in which geographical and juristitional context? where has it been digitized or transformed? Like for temporal relationships, the identification of geographical and jurisdictional context and, in general, of anything we bother expressing in direct or indirect reference to the artifact. For instance, can we state that Leonardo Da Vinci produced the Mona Lisa in Italy, considering that Italy did not exist at the time and he was born within the city-state of Florence?
- Part-whole relationships: is the artefact part of a larger work? In this case, the description proposed belongs to the individual artefact or the larger work?
- Object-subject relationships: Is the artefact talking about another artefact? Is this statement about the principal artefact or about the one being talked about?

- Authorship relationship: who made this assertion? Where? For what purpose? To what audience?
- Derivation relationships: which other artifact allows me to assert the following assertion?
- Confidence relationships: how confident am I that this assertion is correct? Are there other competing assertions to compare this one with?

The aim of these characterizations is to allow authors, scholars and users in general, to make contextualized assertions that reference, and rely on, other assertions.

In addition, it is of utmost importance to have the ability to express and store multiple conflicting views on each of these kinds of assertion (author A, based on document X, states that this coin is from 764 CE; author B, based on document Y, states that this coin is from 890 CE).

In our paper we present a series of guidelines for the evaluation and the development of knowledge modelling techniques and data formats. These guidelines ensure that models and formats are able to express all these kinds of contextual information. We based these guidelines on our experience in dealing with existing data and metadata formats for cultural heritage and public history. [1] In our paper we also show links between these guidelines and related efforts in the fields of museum archives [2], textual criticism [3], law [4] and digital prosopography [5].

References

[1] Gioele Barabucci, Francesca Tomasi, Fabio Vitali. *Modeling data complexity in public history and cultural heritage*. In: Serge Noiret, Mark Tebeau, Gerben Zaagsma (eds) Handbook Digital Public History, De Gruyter. 2021. 9783110439229

[2] Franco Niccolucci. *Documenting archaeological science with CIDOC CRM*. International Journal on Digital Libraries 18(3). 2017 DOI: 10.1007/s00799-016-0199-x

[3] Øyvind Eide, et al. Encoding cultural heritage information for the Semantic Web: Procedures for data integration through CIDOC-CRM mapping. In Open Digital Cultural Heritage Systems Conference (Vol. 25). 2008.

[4] Gioele Barabucci, Luca Cervone, Monica Palmirani, Silvio Peroni, Fabio Vitali. *Multi-layer Markup and Ontological Structures in Akoma Ntoso*. In: Casanovas P., Pagallo U., Sartor G., Ajani G. (eds) AI Approaches to the Complexity of Legal Systems. Complex Systems, the Semantic Web, Ontologies, Argumentation, and Dialogue. AICOL 2009. Lecture Notes in Computer Science, vol 6237. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-16524-5_9

[5] Georg Vogeler, Gunter Vasold, Matthias Schlögl. *Von IIIF zu IPIF? Ein Vorschlag für den Datenaustausch über Personen*. In: Patrick Sahle (Ed.): DHd 2019 Digital Humanities: multimedial & multimodal. Frankfurt / Mainz. DHd. 2019. DOI: 10.5281/zenodo.2600812.