Gioele Barabucci, Francesca Tomasi, and Fabio Vitali

Modeling Data Complexity in Public History and Cultural Heritage

Abstract: The publication by Galleries, Libraries, Archives and Museums of metadata about their collections is fundamental for the creation of our shared digital cultural heritage. Yet, we argue, these digital collections are, on one hand, of little use to scholars (because of the inconsistent quality of the published records), and, on the other hand, they fail to attract the interest of the general public (because of their dry content). These problems are exacerbated by the current move towards public history, where citizens are no longer just passive actors, but play an active role in contributing, maintaining and curating historical records, leading some to question the trustworthiness of collections in which non-scholars have the ability to contribute. The core issue behind all these problems is, we believe, a (doomed) search for objectivity, often caused by the fact that data models ignore the derivative and stratified nature of cultural objects, and allow only one point of view to be expressed. In turn this forces the publication of bowdlerized records and removes any venue for the expression of disagreement and different opinions. We propose an approach named "contexts" to solve these issues. The adoption of contexts makes it possible to support multiple points of view inside the same dataset, not only allowing multiple scholars to provide their own possibly contrasting points of view, but also making it possible to incorporate additions, corrections and more complex kinds of commentaries from citizens without compromising the trustworthiness of the whole dataset.

Keywords: trustworthiness, multiplicity, contexts, disagreement, cultural heritage metadata

Introduction

One of the core tenets of public history is that all levels of society must play an active role in the construction of its cultural identity, in particular by participating in the valorization of its culture and of the artefacts that its culture has produced through its history. It is thus the responsibility of cultural institutions to not just make cultural artefacts, or in the digital case, raw metadata about such cultural artefacts available, but also provide critiques and reflections on said artefacts. The ability to contribute reflections should, however, not be limited to renownedscholars, but extended to amateurs and the general population. Doing so, however, poses a series of issues, common to all crowdsourcing activities and exacerbated by the use of digital technologies: e.g.,

https://doi.org/10.1515/9783110430295-041

how can trustworthiness be maintained? In which way can rich and complex reflections be faithfully expressed? How can one attract new active contributors?

Let's take a step back. What do we use digital technologies for when dealing with cultural heritage? A quick overview in the field of public history as well as of GLAMs (Galleries, Libraries, Archives and Museums) identifies two principal aims: first, preservation of knowledge, i.e., the need to associate to our physical or cultural artefacts what is known about them, should they get destroyed or forgotten about; second, circulation of knowledge, i.e., the desire to allow larger audiences (scholars, students, general citizens) to improve themselves by gaining access to information about our cultural artefacts that otherwise (for fragility, physical distance, obscurity) would be hard or impossible to reach.

Yet, we notice, we are failing on all the abovementioned counts: on the one hand, we frequently find poor digital records, as an extremely limited quantity of information ends up being associated with our physical and cultural artefacts; on the other, we are more or less failing to interest a substantially larger audience to our digital collections.

In many ways, we believe, both failures stem out of a single cause: a (doomed) search for objectivity, often caused by the fact that models ignore the derivative and stratified nature of cultural objects, and allow only one point of view to be expressed, bringing forth unbalanced and incomplete descriptions of our artefacts. In turn, this forced objectivity eliminates conflict and disagreement, the very matters that have the best chance to create and maintain interest in lay audiences.

We suggest, on the contrary, that we should explicitly aim at representing competing points of view and opinions, and make sure that we fully document their existence and strengths, as well as the supporting ideas and providing backing, so that our audiences can finally perceive representations that are truer and more interesting than the sterilized and boring renditions forced by so-called objectivity.

From its inception, digital public history has been well aware of the challenges and potentialities of the use of digital tools to narrate multiple viewpoints, even if competing between them (Noiret 2018).[1] Whether these points of views represent the distance between two similarly accredited scholars with different opinions, or between a recognized expert and an Everyman submitting his/her own private records, the possibility for the data structure to allow, accept and represent faithfully the multiplicity of points of view over our past and our cultural heritage. So far, this complexity has been unavailable in our digital tools. It is our desire and plan to amend this limitation.

1 Serge Noiret. *Digital Public History*. In: *A Companion to Public History*, David Dean (Ed). 2018. Wiley doi:10.1002/9781118508930.ch7.

Starting off with an example of the record of an image in Europeana (see figure # 1), we derive a classification of the issues and a methodological approach that we name "contexts" to express multiple points of view in data models. Our approach draws from a number of existing data models and modelling techniques, and can be implemented using current Semantic Web technologies, although with limitations due to the current state of the standards.

Scholarship, Truth, Disputes: An Example

In many disciplines of the humanities, truths are hard to come by and facts are rare. In most cases, we use words such as *facts* and *truths* just to mean "statements for which there is an acceptable trail of supporting sources," or "statements that are more or less accepted by the majority of the relevant scholars" or even "statements that so far haven't been disproven."

This is not a surprise, as precise knowledge in many cases is impossible or outside our reach. Yet scholars can easily deal with incomplete or conjectural knowledge, and even when they are personally convinced of the truthfulness of some information, they are aware and able to accept that different viewpoints, dissent and speculations may exist about them.

Unfortunately, the ability to contemplate and handle different or opposing interpretations over the same piece of information is not embodied in the software and in the digital data structures that we use to represent them: single points of view and unique data items are usually expected in our digital collections. The inevitable conflicts and disagreements cannot be expressed explicitly, and need to be simplified, resolved and eliminated before committing information to the digital realm.

Yet, unknowns, disputes and dissenting opinions are often what in the first place attracted, fascinated and still keeps fascinating scholars into their respective fields of expertise. Neutered, undisputable and unequivocal data as expressible in our digital world have the twin problems of a) being a poor representation of what we know and think and b) being fundamentally boring and unable to attract anyone, especially lay people and younger students.

Even the ever important dialogue between experts and collective memory must recognize that this dialogue does not always . . . in fact, almost never ends up converging into an agreed, shared, resolved piece of knowledge that is also, at the same time, engaging and fascinating to all. In public history, just like everywhere else, increasing the number of voices participating in the cultural debate about something tends to increase divergence and diatribe rather than solve it, and our software simply cannot keep up with this additional complexity.

Allow us to make an egregious example by considering figure #1, taken from a record in Europeana, and itself coming from the Bildarchiv Foto Marburg (cfr. Peroni,

Tomasi, Vitali 2012).[2] The image was present in 2012 in two separate and dissimilar records in Europeana (See Fig. 2), and is now (2020) present in only one, the second having been removed probably due to being a duplicate and because of the number of issues in its description.

Fig. 1: An image taken from Europeana.

The existing record[3] (Fig. 2a) is accompanied by metadata stating that the item being described is a 53.9 x 41.3 cm print by G.B. Piranesi, dated 1756 (and/or 1787), titled "Veduta dell'Anfiteatro Flavio detto il Colosseo," showing a table number and a signature, and describing as subjects James Caulfield and King Gustaf III of Sweden. The deleted record (Fig. 2b) described the (very same) item as "Amphitheatrum Flavium / Colosseum", a 70 to 80 A.D. building by Vespasianus, and, in a plain text description, represented via a 1960/70 photo by Konrad Helbig of a print by G.B. Piranesi.

Both records are clearly incomplete and wrong. They contain objective and factual information (e.g., 53.9 x 41.3 cm being the size of the item), some more or less

2 Silvio Peroni, Francesca Tomasi, Fabio Vitali. 2012. *Reflecting on the Europeana Data Model*. IRCDL 2012: 228–240.

3 https://www.europeana.eu/en/item/2064137/Museu_ProvidedCHO_Bildarchiv_Foto_Marburg_obj20089555_T_001_T_071.

acceptable interpretations (e.g. the nature of the work, the title, and the author) and some much less acceptable assertions, in fact errors (James Caulfield and King Gustav III are clearly NOT the subject of the image, this item is NOT a building and its creation date is NOT 70 to 80 A.D.).

At the same time, questions abound: what is the record about, a first century building, an eighteenth-century print or a twentieth-century photograph? Who is its author, Vespasianus, G.B. Piranesi or Konrad Helbig? Why two dates: 1756 and 1787? What was the role of James Caulfield and Gustav III?

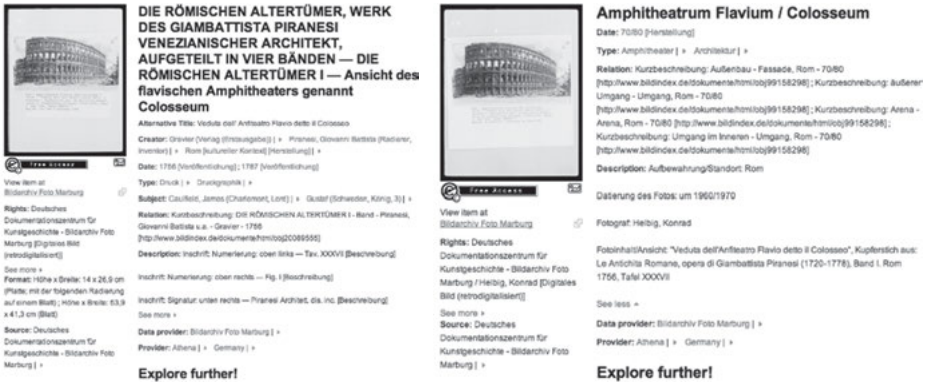

Fig. 2: Two records for the same image from Europeana in 2012.

As we see, the facts are few, and actually constitute the most boring and uninspiring part of the data we are given, while many additional details, although wrongly characterized, paint a much more complex and interesting story than the physical dimensions, once we take pains to understand and explain them. Most interestingly, the more incorrect of the two records (i.e., the one that ended up being deleted) is also the one that contains the most details that can help clarify the whole story.

The following could be a better story: in the 1990s, the Bildarchiv Foto Marburg organized a retrospective of the works of Konrad Helbig (1917–1986), a famous German photographer and art historian, known for his groundbreaking early photographs of Italian classical works of art, and of naked, tanned, underage Sicilian boys. Among the pictures selected, there was one (53.9 × 41.3 cm) of Piranesi's print of the Colosseum. The dimensions as posted are of the exhibition's cardboard, as demonstrated by the typewritten label and the fact that Piranesi's book has a completely different size.

"Antichità romane" by Giovan Battista Piranesi (1720–1778) is a four-volume book of high quality etchings of various scenery of Rome and its countryside, a famous bestseller of the eighteenth century, often brought back home by the young

nobles and high society bourgeois of the age upon their return from the so-called "Grand Tour" in Italy. In particular, table thirty-seven of volume I is titled "Veduta dell'Anfiteatro Flavio detto il Colosseo."

The first edition of the book (1756) was dedicated to James Caulfeild, Lord Charlemont (1728–1799), a 28 year old Irish patron of the arts on his third year of a grand tour in Italy. At that time, in 1756, Gustav III (1746–1792) was only 10 years old, not a king yet, and likely uninterested in the arts. Yet, some years after Piranesi's death, his son Francesco had problems financing a new printing of the book, and found another patron in King Gustav III of Sweden, to whom he dedicated the new edition (1787). This explains both the two dates and the double dedication of the volume in the Europeana records.

This may be a better story, but can we now suggest fixing the Europeana record accordingly?

Well. First, ours is just one among many possible explanations: it should not be promoted as truth only because it is richer or more plausible than the current one. Second, and most importantly, this explanation deals with many different levels of abstraction of the artefact (the chain of reproductions, the physicality of these reproductions, the people associated with them, etc.), and the data model employed by Europeana (Doerr 2010)[4] does not possess the necessary concepts, nor does it allow for the expression of level-specific annotations. Unfortunately, Europeana's problems are not unique as these limitations are common to most data models used in the humanities.

In public digital history, the role of the final users in adding complexity and variety to the description of artefacts is a key requirement. What can the readers add to the interpretation of the observed reality? How can we allow and manage information coming from crowdsourced initiatives for data enrichment? Europeana, with the CrowdHeritage project,[5] is in fact moving toward this direction, but unfortunately still with the unexpressed expectation that only one reading of reality can be preserved, and dissimilar points of view must be reconciled and made to converge outside the data model and before committing them to the digital world.

Common Issues in Representing Metadata

The example in the previous section has shown various typical problems in representing metadata (digitally or otherwise) about our history and cultural heritage. We can classify these issues in a few broad categories:

4 Martin Doerr et al. The Europeana Data Model (EDM). World Library and Information Congress: 76th IFLA general conference and assembly, 2010, 10. http://www.ifla.org/en/ifla76.

5 https://pro.europeana.eu/post/crowdheritage-a-crowdsourcing-platform-for-enriching-europeana-metadata.

- *reticence*: information that was probably known at the time of the digitization was not recorded due to haste, lack of skill, or, more probably, lack of policies and software support for this kind of information. The role of Helbig, the retrospective held about his work, the relationship between table thirty-seven of volume I and the whole of Piranesi's four-volume work, etc. are missing.
- *flattening*: the fact that an artefact is derived from another artefact and that other artefact is derived from yet another artefact is never made explicit. A single set of metadata is used to describe the content of the image, the physical support, and the long chain of entities represented, leading to inconsistent and meaningless information.
- *coercion*: information that was felt to be important, but for which no appropriate field was found, was forced into inappropriate fields (e.g., the dedication of the book to James Caulfeild and Gustav III ending up as the subject of a single page of the book), leaving to the puzzled reader the task of making things straight, and forever baffling any automatic tool tasked with indexing and searching collections by subject.
- *dumping*: some important information, for which no appropriate field was found, was placed as plain text inside a descriptive field, easy for humans to read but forever lost to any automatic management of the data. In particular, the existence of Helbig, the date of the photograph and the placement of the image as an individual page within a four-volume book are narrated in the description field, rather than being formalized in any specific field.

And, a few remaining problems we should also consider:
- not all statements of the story we told have the same reliability;
- all statements are clearly authorial (we are expressing our own personal interpretation); and
- we ourselves have fairly diverse levels of confidence about the events described: while the position of the image within the 1787 edition of the work has been verified *de visu*, we know that the roles of Francesco Piranesi and Gustav III are plausible but unverified, and, worse still, that the 1990 retrospective about Helbig is an interesting but purely hypothetical invention.

In the following section we propose an approach called "contexts." Contexts cohesively address these issues and enable data models to express not only finer details about the artefacts, as they do now, but also a wide variety of opinions, points of views and conjectures that constitute the largest part of our knowledge about cultural heritage.

Contexts for Qualification of Metadata

We can find many data models to represent metadata, facts and information about our history and cultural heritage. These data models are able to encode many fine details about artefacts and historical events, but cannot handle situations like the one presented above and cannot describe the circumstances within which these details find their place, their role, their correctness, their plausibility. The approach we suggest, in order to achieve these goals, is that of *contexts*.

Contexts provide boundaries to opinions and make them comparable to facts. Identifying and expressing the context of all statements is fundamental for their correct interpretation and use. Without the proper context, it is easy to draw false conclusions from data.

In general, we define contexts as sets of statements meant to characterize the metadata about an entity, rather than the entity itself (the entity can be an artefact, an event, a person, a concept or, in general, anything worth describing). For instance, assigning a provenance attribution to the creation date of a print is not a statement about the print, but a statement about a statement about the print.

The following are a few contexts that we have identified:

- *Temporal relationships*: facts and assertions are rarely absolute, and more often constrained by a temporal interval. For instance, Gustav III was in fact the King of Sweden, but only between 1771 and 1792; although he was alive in 1756, he certainly was not the King of Sweden then, and hardly the addressee of a dedication from a Roman printer.
- *Spatial relationships* (or, better, *jurisdictional*): geopolitical entities are evolving concepts and statements that refer to them must be qualified. For example, was Caulfeild Irish? James Caulfeild, 1st Earl of Charlemont (1726–1799),[6] was a noble of the Kingdom of Ireland, then under the rule of the Crown of England, with little or no direct connection with the current Republic of Ireland. Describing him as Irish is therefore just a handy simplification for a much more nuanced characterization of his true affiliation.
- *Part – whole relationships*: the item being described in Helbig's photograph is only one page in a four-volume book, itself published in at least two editions. It is important to be able to distinguish between the statements regarding the individual page (the heading of the image, the subject, etc.) and those regarding the volume as a whole (the author, the publication date[s], the dedications, etc.)
- *Object-subject relationships*: Object-subject chains can be particularly deep, intricate and fascinating. In our example, what we are describing is, in fact, a JPEG image created in the 2000s, derived from a high-resolution TIFF scan dated

6 The authors would like to express their gratitude to Daniel Kiss for his clarification about the name and title of James Caulfeild.

somewhere in the 1990, about a photograph by Konrad Helbig dated 1956, about an etching dated 1756 by Giovan Battista Piranesi, about a 70 to 80 AD building called the "Colosseum" in Rome. Each one of these entities deserves descriptive metadata about them, but they must be correctly associated with the entity actually being described.
- *Provenance*: All statements present in the metadata come from a source that should be identified: an individual, a text, a direct analysis of the artefact, etc. Provenance information is needed, not only to give backing and responsibility to the statement itself, but, most importantly, to allow multiple different and competing statements, possibly in contrast with each other, to coexist in the same metadata collection. Without provenance there is no complexity, there is no dissent, there is no public history.
- *Confidence*: the sources themselves expose varying degrees of confidence in expressing this or that fact. Recording a confidence level allows for conjectural, hypothetical and even whacky statements to correctly coexist with established and settled information.

The use of contexts brings back the objectivity and truthfulness that software needs and that computer scientists crave for, without giving up the richness of information favored by scholars in the humanities. Consider the statement "at the end of twentieth century the Marburg Foto Archiv organized a retrospective about Konrad Helbig." It is clearly speculative, non-objective and conjectural. Adding a few contexts, the previous statement becomes "Barabucci, Tomasi and Vitali (2020) speculate that it is possible that at the end of twentieth century the Marburg Foto Archiv organized a retrospective on Konrad Helbig." By flatly stating the conjectural nature of the hypothesis, we made the statement as a whole more objective and easily verifiable: the addition of the contexts around the statement made it stronger and usable in a scientific discourse.

In summary, contexts have the undeniable advantage of being able to accept a much larger quantity of information than simply the official and established data, allowing for multiple conflicting views over the same items, and, ultimately, allowing a much more interesting and nuanced representation of our cultural past.

Representing Contexts in the Semantic Web

Models adopted for the description of cultural objects are various and heterogeneous. Traditionally, metadata in cultural heritage are classified depending on their role (e.g. descriptive, administrative/technical and structural). Riley (2009–10)[7]

7 Riley Seeing Standards: A Visualization of the Metadata Universe 2009–10, http://jennriley.com/metadatamap/.

identifies instead four macro-categories: community, purpose, function and domain. This classification allows us to deal, in an effective and explicit way, with the complications that arise when metadata descriptions are created by different communities (e.g. libraries, archives and museums), for different purposes (e.g. description, preservation, technical features, structure or rights), with different functions (e.g. structure standards, content standard, conceptual model, controlled vocabularies, markup languages, record format) and while dealing with different domains (e.g. cultural objects, moving images, datasets, geospatial data, music materials, scholarly texts, visual resources).

Originally, data models were basically content standards, created as the result of reflections about theories on data description elaborated within different scholarly communities. The AACR2 rules[8] for libraries, or ISAD[9] and ISAAR-CPF[10] for archives are examples of this type of content standards.

These early theoretical models, fairly distant from actual implementations, were later rethought as structural standards, especially after XML started providing an adequate syntax to formalize existing vocabularies (DTDs first, XSD Schemas later). Examples of such structural standards, designed to support the description of data through markup languages, are EAD[11] and EAC-CPF[12] for archives, TEI[13] for literary texts and MODS[14] for libraries.

We now live in a yet more sophisticated world where we talk about ontologies as a new form of formalization methodology for enriching the expressivity of Schemas. Ontologies such as OAD,[15] EAC-CPF,[16] CIDOC-CRM,[17] RDA,[18] Bibframe[19] are behind the conceptual modelling of the new Linked Open Data cloud.[20] Add to the mix controlled vocabularies for managing the value of the attribute for person, places, subjects, concepts and objects (e.g. VIAF, DDC, UDC, LCSH, Getty vocabularies,

8 Anglo-American Cataloguing Rules (AACR): http://www.aacr2.org/.

9 General International Standard Archival Description (ISAD(G)): https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition.

10 International Standard Archival Authority Record for Corporate Bodies, Persons and Families, 2nd Edition (ISAAR (CPF)): https://www.ica.org/en/isaar-cpf-international-standard-archival-authority-record-corporate-bodies-persons-and-families-2nd.

11 Encoded Archival Description (EAD): https://www.loc.gov/ead/.

12 Encoded Archival Context for Corporate Bodies, Persons, and Families (EAC-CPF): https://eac.staatsbibliothek-berlin.de/.

13 Text Encoding Initiative (TEI): https://tei-c.org/.

14 Metadata Object Description Schema (MODS): http://www.loc.gov/standards/mods/.

15 Ontology for Archival Description (OAD): http://culturalis.org/oad/.

16 Encoded Archival Context – Corporate bodies, Person and Families (EAC-CPF) Ontology: http://culturalis.org/eac-cpf/.

17 CIDOC CRM: http://www.cidoc-crm.org/.

18 Resource Description and Access (RDA): see the registry, https://www.rdaregistry.info/.

19 Bibframe: https://www.loc.gov/bibframe/.

20 The Linked Open Data Cloud: https://lod-cloud.net/.

Geonames, Dbpedia and Wikidata[21]) and you obtain a fairly complete and sophisticated set of domain models for the description of a large part of cultural heritage artefacts and historical sources and knowledge.

To summarize, metadata element sets, controlled vocabularies, schemas and ontologies – and in general any relevant standard such as those proposed by domain associations such as IFLA,[22] ICA[23] and ICOM[24] – clearly show us how rich and detailed the description of cultural objects can become (Isaac et al. 2011).[25] Yet, although each of these models in fact exposes a rather complex, multidimensional and interconnected landscape, none of them gets close to the very issues we are discussing here, because of the underlying and implicit assumption that data description should be neutral and objective.

Many past and current reflections of Digital Public History as a discipline bring forth the importance of allowing and handling crowdsourced contributions from all sectors of the society, not just scholars, thus reinforcing the need for multiplicity of points of view on data. These contributions must be documented in order to generate more expressive and complex descriptions.

Already the notion of the neutrality of data models is being challenged by proposals such as HICO[26] (Daquino and Tomasi 2015),[27] in order to deal explicitly with interpretation acts (hico:InterpretationAct as well as classes such as Criterion and Type) as fundamental tools for expressing provenance of semantic interpretations; similarly Mauth[28] (Daquino 2019)[29] is useful to express the authoritativeness of existing statements with explicit paternity, and to let final users become active parts of the description process.

21 See in particular from Isaac et al. 2011 the section devoted to "Value vocabularies": https://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/#Value_vocabularies.

22 International Federation of Library Associations (IFLA) standards: https://www.ifla.org/standards.

23 International Conuncil on Arcives (ICA) standards: https://www.ica.org/en/standards-and-tools.

24 International Council of Museums (ICOM): https://icom.museum/en/resources/standards-guidelines/standards/.

25 Isaac Antoine [et al.], Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets, W3C Incubator Group Report. October 25 2011, <https://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/>. A complete overview of ontologies could be read in Linked Open Vocabularies: https://lov.linkeddata.es/dataset/lov/.

26 Historical Context Ontology (HiCO): http://hico.sourceforge.net/.

27 Marilena Daquino, Francesca Tomasi. 2015. *Historical Context Ontology (HiCO): a conceptual model for describing context information of cultural heritage objects*. MTSR 2015: 424–436.

28 mining Authorithativeness in Art History (Mauth), http://purl.org/emmedi/mauth.

29 Marilena Daquino, *Mining Authoritativeness in Art Historical Photo Archives: Semantic Web Applications for Connoisseurship*, IOS press 2019.

Other ontologies such as PROV-O[30] for describing the provenance, SPAR[31] for publishing data, FaBiO[32] (as an FRBR-aligned ontology), PRO[33] for managing roles, or CiTO,[34] for representing the citations are able to add other types of contexts to data description. Similarly, the CMV+P (Barabucci 2019)[35] document model is based on the fact that most cultural artefacts reference or embed other artefacts explicitly and integratably into other data models.

Finally, the recent RiC-O[36] model uses the similarly named notion of "contexts" to indicate a plurality of paratextual information used to translate the classical siloed approach to data descriptions into a graph of connections between vocabularies and ontologies.

Contexts can also be seen as a generalization of the Factoid model (Bradley and Short 2005)[37] used in prosopography, where it is common practice to treat information found in old records not as objective truths, but as utterances of partially trusted sources. This mistrust of sources, and the consequent need for contextualization, is reflected in the design of modern APIs for querying historical datasets (Vögeler 2019).[38]

Some recent projects (Daquino et al. 2017;[39] Daquino, Giovannetti and Tomasi 2019[40]) demonstrated the possibility of a semantic enrichment through a contexts-aware approach, especially in the Linked Open Data workflow. This is the reason why Semantic Web, and RDF/OWL, are our starting point for the following reasoning on contexts.

30 Prov-O: https://www.w3.org/TR/prov-o/.

31 Semantic Publishing and Referencing Ontologies (SPAR): http://www.sparontologies.net/.

32 FRBR-aligned Bibliographic Ontology (FaBiO): http://purl.org/spar/fabio.

33 Publishing roles Ontology (PRO): https://sparontologies.github.io/pro/current/pro.html.

34 CiTO, the Citation Typing Ontology: http://purl.org/spar/cito.

35 Gioele Barabucci: *The CMV+P document model, linear version*. In *Versioning cultural objects*. IDE, 2019. urn:nbn:de:hbz:38–106539.

36 Records in Contexts Ontology (RiC-O): https://www.ica.org/standards/RiC/RiC-O_v0-1.html.

37 J. Bradley, H. Short: *Texts into Databases: The Evolving Field of New-style Prosopography*. Literary and linguistic computing 20: 3–24. 2005.

38 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.

39 Zeri & LODe project: http://data.fondazionezeri.unibo.it/.

40 Semantic Digital Edition of Bufalini Notebook: http://projects.dharc.unibo.it/bufalini-notebook/. Si veda anche Marilena Daquino, Francesca Giovannetti, Francesca Tomasi, *Linked Data per le edizioni scientifiche digitali. Il workflow di pubblicazione dell'edizione semantica del quaderno di appunti di Paolo Bufalini*. Umanistica Digitale 7, 2019. https://umanisticadigitale.unibo.it/article/view/9091.

Accommodating Contexts

The natural habitat for contexts as presented in the previous sections is in datasets expressed with Semantic Web standards, frequently used to represent metadata for cultural heritage artefacts. RDF[41] and OWL[42] are the two such technologies.

RDF is used to express statements about well-identified entities. In the RDF model each statement is expressed using a so-called triple, composed of a subject, a predicate and an object, the subject being the entity being described. For example: "Antichità Romane" (subject) has author (predicate) "G.B. Piranesi" (object).

Expressing metadata using RDF is easy, and most of the existing metadata models have an RDF representation. Modelling contexts means expressing statements whose subject (the entity being identified) is not the artefact being described, but another statement in RDF that expresses some quality about the artefact. For instance, in example #1, *_s* is a statement that expresses the fact that the book "Antichità Romane" was authored by G.B. Piranesi (numbers 1 and 2). The context *_c* added at the end (numbers 3, 4 and 5) affirms that statement *_s* was created by GBarabucci, FVitali and FTomasi by introducing a ctx:Context class that assigns a clear and unambiguous provenance to it.

```
:AntichitàRomane rdf:type ex:Book .                          ①
:_s rdf:type rdf:Statement;                                  ②
rdf:subject:AntichitàRomane;
rdf:predicate dc:author;
rdf:object:GBPiranesi .
   :_c rdf:type ctx:Context .                                ③
   :_c ctx:forStatement:_s .                                 ④
   :_c ctx:assertedBy [:GBarabucci,:FVitali,:FTomasi] .      ⑤
```

Example #1: A contextualized statement.

We cannot, however, simply add a couple of RDF statements to existing RDF collections. For instance, although these statements are meant to represent the sentence "Barabucci, Vitali and Tomasi assert that G.B. Piranesi is the author of the book 'Antichità Romane,'" they would actually be understood by RDF processors (reasoners) as expressing two slightly different, independent sentences: "G.B. Piranesi is the author of the book 'Antichità Romane' and Barabucci, Vitali and Tomasi assert this." The authorship of the book is placed on the same level of truth as the provenance of the statement. This is unfortunate, because the provenance of context *_c*

41 Resource Description Format: http://www.w3.org/TR/rdf11-concepts/.
42 OWL 2 Web Ontology Language: http://www.w3.org/TR/owl2-overview/.

does not factor in, nor restrains, the statement of authorship, as we were hoping to obtain.

Extending current data models with the ability to assert statements as true only within and depending on the truth of a given context is problematic due to shortcomings of the underlying RDF meta-model. The same issues arise not only in the Semantic Web, where RDF and OWL are used, but also in traditional databases, where the ER model (Chen 1976)[43] is used.

We could, of course, overcome this limit by introducing ad-hoc interpretation rules in our own metadata processors, e.g., by having them ignore assertions outside of explicitly activated contexts. This approach would work in practice, but would be project-specific and would prevent the sharing of our data with the rest of the world, e.g. in the Linked Open Data: we cannot expect all participants in the LoD to use of our modified rules instead of standard RDF reasoners to process this and other RDF datasets.

Or, we could adopt the RDF extension called "nested named graphs" (Gandon and Corby 2010),[44] that introduces the concept of "local truth" and changes what RDF reasoners are allowed to infer from a dataset. Example #2 provides the same example using nested named graphs.

```
:AntichitàRomane rdf:type:Book .
:C {
  :S {
    :AntichitàRomane dc:author:GBPiranesei .
  } ctx:assertedBy [:GBarabucci,:FVitali,:FTomasi] .
}
```

Example #2: A nested named graph of the same contextualized statement.

In this reformulated example we are nesting the authorship attribution *:S* inside graph *:C*, and explicitly specify that the graph *:S* is asserted by "GBarabucci," "FVitali" and "FTomasi." In other words, the outermost graph guards the content of the subgraph and blocks reasoners from considering the inner content as true independently of the truth of the outer context.

This approach is correct, working and in line with the best practices of the W3C, the authority behind Semantic Web technologies. Nonetheless, as of 2020, nested named graphs are not yet fully standardized, nor supported by reasoners.

---

43 Peter Chen. *The Entity-Relationship Model: Toward a Unified View of Data*. ACM Transactions on Database Systems. 1(1): 9–36. doi:10.1145/320434.320440.

44 Fabien Gandon, Olivier Corby: *Name That Graph, or the need to provide a model and syntax extension to specify the provenance of RDF graphs*. 2010. https://www.w3.org/2009/12/rdf-ws/papers/ws06/.

Adopting them or not constitutes a foundational problem that must be addressed by scholars and data modelers before they are able to apply our approach, or any similar one that looks towards allowing multiple reciprocally inconsistent datasets to be associated to the same entities.

With this model it becomes easy to express not only different opinions by established scholars, but also statements that are the result of crowdsourced activities, while providing, at the same time, the consistency and the truthfulness needed to support the complexity of heterogeneous point of views.

Conclusions

The push towards a single and objective description of cultural artefacts in digital systems is causing an impoverishment of our data collections, to the point that they end up as neither useful to scholars nor captivating for lay audiences. We believe that this happens because current data models and their underlying meta-models lack the ability to express, first, conflicting interpretations and, second, the stratified relations that exists between artefacts (e.g., a JPEG image being derived from a photo, that in turn depicts a painting, that represents a building). These problems stem from a root issue: the lack of support for multiple points of views in data models. To address this issue, we propose an approach in which all assertions are contextualized by associating various facts about them such as tempo-spatial relations, object-subject relations, provenance, etc.

Nevertheless, as Digital Public History recognizes, the general public must be allowed to have a voice in enriching our cultural tradition. Context information is key to this much needed participation, because it allows the role of contributors to be recognized, while at the same time, preventing the dilution of trust that may arise in such crowdsourced activities.

The adoption of this approach makes it possible to support multiple points of view inside the same dataset, allowing not only multiple scholars to provide their own possibly contrasting points of view, but also addressing key issues related to the public crowdsourcing initiative, such as the fear of contamination of curated datasets with unreliable information. Most importantly, with this model it is possible to incorporate additions, corrections and more complex kinds of commentaries from citizens without compromising the trustworthiness of the whole dataset.

With respect to sharing metadata with other institutions, our notion of contexts is in line with the ethos and the direction towards which the Semantic Web, the Linked Open Data cloud and modern data models are moving, but its practical adoption is at the moment hindered by its reliance on technologies that are not yet standardized nor widely implemented.

Still, the core principles behind our approach (in particular, the ability to express conflicting opinions) are necessary steps for our future digital collections to truly become useful, trustworthy and engaging for scholars and citizens alike.

Bibliography

Barabucci, Gioele. "The CMV+P document model, linear version." *Versioning Cultural Objects: Digital Approaches*. Eds. Roman Bleier, and Sean M. Winslow. IDE, 2019. urn:nbn:de:hbz:38–106539.

Daquino, Marilena, Francesca Mambelli, Silvio Peroni, Francesca Tomasi, and Fabio Vitali. "Enhancing Semantic Expressivity in the Cultural Heritage Domain: Exposing the Zeri Photo Archive as Linked Open Data." *ACM Journal of Computer Cultural Heritage* 10.4 (July 2017). http://dx.doi.org/10.1145/3051487.

Peroni, Silvio, Francesca Tomasi, Fabio Vitali. "Reflecting on the Europeana Data Model." *IRCDL* (2012): 228–240.

Vitali, Fabio. "Beyond Three Dimensions: Managing Space, Time and Subjectivity in your Data." *SUMAC '19: Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia heritAge Contents. October* 2019. https://doi.org/10.1145/3347317.3352728. 3–4.