# Supporting complexity and conjectures in cultural heritage descriptions

Gioele Barabucci[1], Francesca Tomasi[2], and Fabio Vitali[2]

[1] NTNU – Norwegian University of Science and Technology, Trondheim, Norway
gioele.barabucci@ntnu.no
[2] Alma Mater Studiorum – University of Bologna, Bologna, Italy
{francesca.tomasi,fabio.vitali}@unibo.it

**Abstract.** Data and metadata once hidden in dusty card cabinets of thousands of galleries, libraries, archives and museums worldwide are now available online in digital formats. An incredible explosion of metadata has been expanding in the quantity of digitized data, the richness and sophistication of data models, the number of institutions and private citizens that contribute and their interconnection.
A fundamental issue, however, limits this undeniable success: current data models force the expression of a single point of view. For example, the field *author* is either set to a value or to another one. Any disagreement about the content of a field is resolved before the publication of the data and forever lost.
Yet, we argue, the expression of different and contrasting points of views is a keystone of scholarship, as well as one of the most engaging aspects for everyone. Bowdlerized, sterile, conflict-free data records fails to capture the core of important scholarly debates and thus fails to attract the interest of the general public. The root cause of this issue is technical rather than cultural: current standards for data models (e.g. RDF, OWL) do not simply support the expression of contrasting statements.
In this paper we propose both a methodological approach to address this problem, and a proof of concept of how this problem could be fully and cleanly overcome with modest extensions to the existing standards. We name this approach "contexts and conjectures".

**Keywords:** multiple points of view, contrasting assertions, RDF serialization, scholarly conjectures.

## 1 Introduction

Why do Galleries, Libraries, Archives and Museums digitize collections? We can list a few reasons: *preservation of knowledge*, i.e., associating to physical or cultural artefacts what is known about them, should they get destroyed or forgotten about; *interlinking*, i.e., historically situating these artefact by showing the connections between them; *circulation of knowledge*, i.e., allowing larger audiences to gain access to information about cultural artefacts that otherwise (because of fragility, physical distance, obscurity) would be hard or impossible to reach.

Yet, we notice, we are failing on all counts. We fail at preserving knowledge because an extremely limited quantity of information ends up being associated with our physical and cultural artefacts; we fail at interlinking collections because technological constraints force us to choose between consistent or connected datasets; and, most importantly, we fail at circulating knowledge because we cannot interest a substantially large audience to our digital collections.

We believe that all these failures stem from a single cause: a (*doomed*) search for objectivity, forced by digital models that mostly ignore the derivative and stratified nature of cultural objects, and allow for only one point of view to be expressed, leading to unbalanced and incomplete descriptions of our artefacts. This eliminates debates and disagreements, exactly what has a chance to create and maintain interest in scholars and general audiences alike.

We rather believe that we should explicitly aim at representing competing points of view and opinions, and make sure that we fully document their existence, their strengths and the ideas behind them, 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.

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 have not 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 of contemplating and handling 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 are what we expect 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 keeps fascinating scholars. Neutered, indisputable and unequivocal digital data 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 laypeople and younger students.

Our approach to making datasets able to express the richness of the non-digital scholarly debates is composed of two parts: *contexts* (Section 3) to allow the expression of conflicting opinions on an artefact, and *conjectural graphs* (Section 4) to encode scholarly statements as qualified arguments, rather than as absolute truths, and a mechanism, *collapse to reality*, to turn some conjectural graphs into plain RDF statements that can be interlinked and get them to interoperate with other RDF-based datasets.

## 2 Common issues in representing metadata

Consider the example in Fig. 1, taken from a record in Europeana, coming from the Bildarchiv Foto Marburg [13].

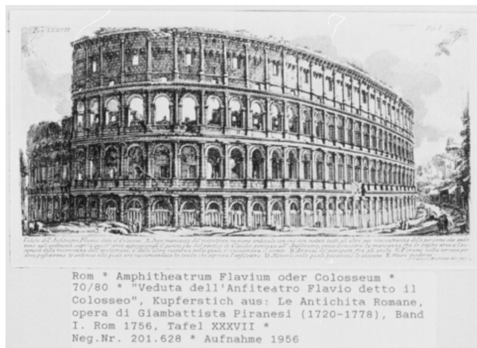


**Fig. 1.** An image found in an Europeana record

The existing record[3] 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", with subject James Caulfield and King Gustaf III of Sweden. Another (now deleted) record described the (very same) item as the Colosseum, a 70-80 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 (see Fig. 2).

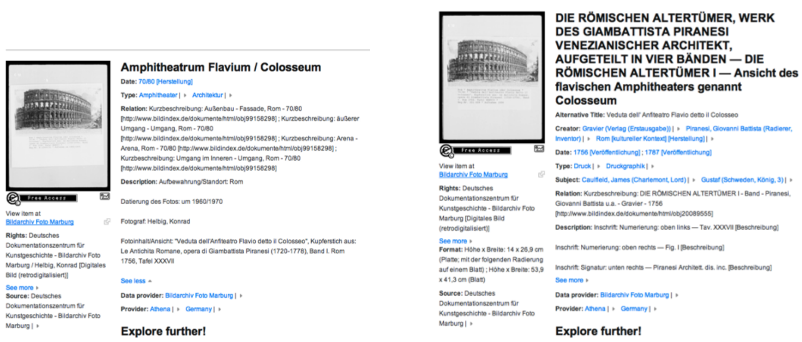


**Fig. 2.** Two records for the same image from Europeana in 2012

These records are both incomplete and wrong. They contain, at the same time, objective and factual information (e.g., the size of 53,9 x 41,3 cm), some

[3] https://www.europeana.eu/en/item/2064137/Museu_ProvidedCHO_Bildarchiv_Foto_Marburg_obj20089555_T_001_T_071 The image was present in 2012 in two separate and dissimilar records in Europeana, 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.

reasonable interpretations (e.g., the author) and many factual errors (e.g., 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* 70-80 A.D.).

So, the facts are few and actually constitute the most boring and uninspiring part of the data items we are shown, and many details, although wrongly characterized, paint a much more complex and interesting story, once we take it upon ourselves to understand and explain them.

We have a better story: this is a black/white photograph from a retrospective of the works by Konrad Helbig (1917-86), a famous German art photographer, known for his groundbreaking photographs of Italian classical works of art, and of tanned, underage, naked Sicilian boys. The photograph depicts a print of the Colosseum by Giovan Battista Piranesi (1720-78) taken from "Antichità romane", a four-volume book of high quality etchings of various sceneries of Rome and its countryside, a bestseller of the XVIII Century, often brought back home by the young nobles and high society bourgeois upon their return from their "Grand Tour" in Italy.

The first edition of the book (1756) was dedicated to James Caulfeild, Lord Charlemont (1728–1799). In 1756 Gustav III (1746–1792) was 10 years old, not a king yet, and likely uninterested in the arts. He was King, on the other hand, after Piranesi's death, when his son Francesco (1758-1810) made Gustav III the new patron of the publishing and dedicated the new edition (1787) to him. This explains both the two dates and the double dedication of the volume.

This may be a better story, but can we now expect the Europeana record to be fixed accordingly? Uhm, well. First, ours is just one of many possible explanations: it should not be promoted to 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 [10] does not possess the necessary concepts. Unfortunately, Europeana's problems are not unique: these limitations are common to most data models used in the humanities.

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

**reticence:** information that was well known was not recorded, due to haste, lack of skill, or, more probably, lack of policies and software support. The role of Helbig, the relationship between the table and whole book are missing, etc.

**flattening:** the fact that an artefact depicts another artefact is never made explicit, and here it actually occurs twice (a photograph of an etching of a building). A single set of metadata describes the image, the physical support, and the chain of entities depicted, as an inconsistent and meaningless record.

**coercion:** important information for which no appropriate field was found, was forced into inappropriate fields (e.g., the dedication of the whole book to James Caulfeild and Gustav III ends up as the subject of a single table), leaving to the reader the job of making things straight, and forever baffling any automatic tool tasked with indexing and searching collections by subject.

**dumping:** important information for which no appropriate field was found, was forced as plain text inside a descriptive field, easy for humans to read but forever lost to any automatic tool. For instance, the existence of Helbig, the date of the photograph and the exact placement of the table within a four-volume book are text-only in the Description field, and never formalized.

In addition to these, there are a few remaining problems we should also consider: not all statements of our story have the same reliability; our statements are clearly authorial (we are expressing our own personal interpretations); and we ourselves have fairly diverse levels of confidence about the events described.

In the following section we propose an approach called contexts. Contexts cohesively address these issues, and their adoption enables data models to express not only fine 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.

## 3 Creating rigorous metadata by adding contexts

Plenty of data models exist in the digital humanities to represent metadata, facts and information about our cultural heritage. They encode fine details about our artefacts but cannot easily handle more complex 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 to overcome these limits is the use of *contexts*, presented in more details in [2] and summarized in this section.

In general, we define contexts as sets of statements meant to characterize the *metadata about an artefact*, rather than *the artefact* itself. For instance, determining the provenance of the creation date of an ancient vase is not a statement about the vase, but a statement about a statement about the vase.

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 them, we could draw false conclusions. The following are a few contexts that we have identified:

**Temporal relationships:** facts and assertions are often constrained inside a temporal interval. For instance, Gustav III was the King of Sweden only between 1771 and 1792; he was alive in 1756, but he was not King then, and hardly the addressee of a dedication from a Roman printer.

**Spatial relationships (or, better, jurisdictional):** geopolitical entities are evolving concepts and statements referring to them must be qualified. For example, was Caulfeild Irish? James Caulfeild, 1st Earl of Charlemont (1726-1799), 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. His nationality is a handy simplification for a much more nuanced characterization of his true origins.

**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.

We should be able to distinguish facts regarding the individual page (the heading of the image, the subject, etc.) from 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, derived from a high-resolution TIFF scan, of a photograph by Konrad Helbig dated 1956, of an etching dated 1756 by G.B. Piranesi, of a 70-80 a.D. building in Rome. Each of these items deserves descriptive metadata that must be associated with the correct entity.

**Provenance:** Statements of the metadata come from sources that should be identified: an individual, a text, the direct observation of the item, etc. Provenance not only gives backing and traceability to the stated facts, but most importantly it allows multiple different and competing statements to coexist, even if in contrast with each other.

**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 wacky statements to correctly coexist with established and settled information.

The use of contexts provides the objectivity and truthfulness that software needs and computer scientists crave for, without giving up on the richness of information favored by domain scholars. E.g., the statement "the Marburg Foto Archiv organized a retrospective about Konrad Helbig" is clearly speculative, non-objective and conjectural. Adding a little context, the statement "Barabucci et al. (2020) speculate that the Marburg Foto Archiv organized a retrospective about Konrad Helbig" becomes objective and easily verifiable. The addition of provenance created a statement that is undeniable and usable in a scientific discourse.

Contexts allow expressing more than just settled and boring data: their ability to enunciate conflicting views allows the creation of a much more interesting and nuanced representation of our cultural past.

### 3.1 Expressing contexts in RDF

The natural habitat for contexts as presented in the previous sections is in datasets expressed with Semantic Web standards such as RDF[4] and OWL[5]. Expressing metadata using RDF is easy, and most of the existing models used in cultural heritage already have an RDF representation.

RDF is used to express statements about entities. In the RDF model each statement is expressed using a 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*).

Modelling contexts means expressing statements whose subject is not the artefact being described, but another statement in RDF that expresses some quality about the artefact. The simplest approach is called *reification*, which

[4] Resource Description Format: http://www.w3.org/TR/rdf11-concepts/

[5] OWL 2 Web Ontology Language: http://www.w3.org/TR/owl2-overview/

requires treating the statement itself as an entity, which becomes the subject of the contextual statement. For example, in Fig. 3, `:S` is a statement that asserts that the book "Antichità Romane" was authored by G.B. Piranesi (1). The statement at the end (2) affirms that statement `:S` was attributed to GBarabucci, FTomasi and FVitali, thus assigning a clear and unambiguous provenance to it.

```
:S rdf:type rdf:Statement ;       ①
    rdf:subject :AntichitàRomane ;
    rdf:predicate dc:author ;
    rdf:object :GBPiranesi .
:S prov:wasAttributedTo [ :GBarabucci ,          ②
                          :FTomasi , :FVitali] .
```

**Fig. 3.** A contextualized statement

We cannot, however, simply add these RDF statements to existing RDF collections. These statements would be understood by RDF processors and reasoners as expressing at the same time two related but independent sentences: 1) that "G. B. Piranesi is the author of the book 'Antichità Romane'", and 2) that "Barabucci, Tomasi and Vitali assert this". The truth of the first statement is not under discussion: it is flatly asserted. The provenance of this statement is provided in addition to it and is at the same level. Contexts expressed in this way do not 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 [6] 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. as Linked Open Data: we cannot expect all LoD participants to use our modified rules instead of standard RDF reasoners to process this and other RDF datasets.

Or, we could adopt a RDF extension called nested named graphs [11], that introduces the concept of "local truth" and changes what RDF reasoners are allowed to infer from a dataset. Nonetheless, as of 2020, nested named graphs are not yet fully standardized, nor supported by reasoners.

Adopting contexts would make it possible to express different opinions while providing the consistency and the truthfulness needed to support the coexistence of heterogeneous point of views.

## 4 Conjectures: machine-readable scholarly assertions grounded on contexts

Contexts provide a clear and systematic way to record possibly contrasting statements about the same object from multiple points of views, but they do not address how contrasting statements may be processed, combined and presented by a piece of software.

Our approach to make contextual metadata coexist and work with other RDF-based datasets goes under the name of *conjectures*. Conjectures are scholarly statements that are formally expressed without being asserted, i.e., without an associated truth value. Statements expressed in conjectures are as in a limbo: their truth state is unknown, and they can, if desired, be subjected to conditions, called *guards*. Now and then, for example when evaluating a what-if scenario, conjectures can be programmatically turned into usual assertions through a specific operation called *collapse to reality* (or *collapse to truth*).

Multiple statements can be collected in *conjectural graphs*, a set of statements that may or may not be asserted collectively and that must be either all asserted or none asserted at the same time. When a conjecture is collapsed to reality, the corresponding conjectural graph becomes equivalent to a plain RDF graph.

In the following sections we will illustrate through a running example how conjectures, conjectural sets and contexts fit together, discussing basic semantics, notation and compatibility with RDF. The basis of our example are two incompatible conjectures: "Francesca Tomasi states that Giovanni Battista Piranesi is the author of Antichità Romane" and "Fabio Vitali states that Elvis Presley is the author of Antichità romane".

### 4.1 Semantics and notation

The core semantics of a conjecture, in contrast to classic RDF statements, is that, by entering a conjecture, we do not assert the corresponding statements, but only assert that 1) these conjectures exist, 2) they are available for examination, and 3) they can be collapsed to reality.

We use the a simple extension to the TriG [5] syntax to denote a conjecture graph "Cj" that expresses that the statements "S1 P1 O1" and "S2 P2 O2" are possibly but unknowingly true. A guarding statement can be added, for instance an attribution to a specific agent A that has provided the conjecture:

```
CONJECTURE Cj {
    S1 P1 O1 .
    S2 P2 O2 .
}
Cj prov:wasAttributedTo A.
```

Conjectural graphs cannot be linearized as RDF graphs without losing their semantic expressivity. Nonetheless, there exists a simple and straightforward representation of conjectural graphs in RDF as named graphs, and it can be generated algorithmically with a very limited procedural effort.

By construction, the conjecture "Cj" is defined as equivalent to a new graph "CTXj", called the *weak context graph* of Cj, in which the predicates in the statements have been replaced by weakened predicates that are only conjecturally equivalent to their original counterparts:

```
CTXj {
    S1 PG1 O1 .
    S2 PG2 O2 .
```

```
    PG1 isConjecturallyEquivalentTo P1 .
    PG2 isConjecturallyEquivalentTo P2 .
}
CTXj prov:wasAttributedTo A.
```

The graph is a weak form of the conjecture because instead of using the expected predicates `P1` and `P2`, we use *ad hoc* predicates PG1 and PG2, which are different and independent from `P1` and `P2`, but are actually asserted, and connect the same subjects to the same objects. Weak predicates are associated to their respective strong predicates through the new predicate `isConjecturallyEquivalent` which provides the semantic closeness and structural weakness that we need to establish the connection between them.

In this way we can assert multiple clashing statements at the same time: conjectures can be expressed side by side and their assertion of truth is absent or postponed.

Our running example can now be expressed as follows:

```
CONJECTURE :T { :AntichitàRomane dc:creator :GBPiranesi . }
:T prov:wasAttributedTo :FTomasi.

CONJECTURE :V { :AntichitàRomane dc:creator :EPresley. }
:V prov:wasAttributedTo :FVitali.
```

As said, these conjectural graphs are equivalent to the following weak contextual graphs in plain RDF:

```
:T {
    :AntichitàRomane T:creator :GBPiranesi .
    T:creator conj:isConjecturallyEquivalent dc:creator .
}
:T prov:wasAttributedTo :FTomasi.
:V {
    :AntichitàRomane V:creator :EPresley.
    V:creator conj:isConjecturallyEquivalent dc:creator .
}
:V prov:wasAttributedTo :FVitali.
```

where the syntax `T:creator` and `V:creator` denotes two weak predicates that live in separate namespaces, differ from each other and differ also from the original Dublin Core `dc:creator`, but are explicitly represented as conjecturally equivalent to it.

### 4.2 Collapse to reality

The process one can use to actually assert the content of conjectural graphs is that of *collapse to reality*, in which the statements belonging to a conjecture are reasserted in their full and strong form, while leaving other conjectures in their weak, non-asserted form. The collapse does not modify the nature of the conjecture, but re-empowers it. The term refers to the reduction of the superposition of multiple independent (and, possibly, incompatible) conjectures to a single one which is then asserted as true, and it is inspired by a similar concept in quantum mechanics, where the wave function collapse is the effect of an observation from the external world that reduces multiple states of a quantum system to a single one.

In practical terms, collapsing a graph means programmatically adding extra statements that short-circuit the guarded predicates. They re-represent the guarded predicates as actual facts. This is done either via the introduction of predicates such as `owl:equivalentProperty` to connect weak and strong predicates, or by re-asserting in their full form the statements expressed in weak form. Since the two approaches are redundant but not reciprocally incompatible, we can do both so as not to depend on OWL tools to obtain the desired result.

Collapses to reality are, themselves, graphs that can be contextualized, for instance with provenance. For example, if we want to express the fact that Gioele Barabucci believes in the attribution by Francesca Tomasi, we would augment "T" with a collapsing predicate and attribute it to him, as follows:

```
:collapseOfT prov:wasAttributedTo :GBarabucci .
:collapseOfT conj:collapses :T .
:collapseOfT {
   :GBPiranesi dc:creator :AntichitàRomane .
   T:creator owl:equivalentProperty dc:creator .
}
```

The system sees and processes the new statement as any other statement, without needing specific support for conjectural graphs. The collapse to reality is well contextualized and can be made temporary, since we can cleanly undo it by removing the collapsed graph as a switch in what-if scenarios. At the same time, all conflicting conjectures are still in place and available for further processing by conjecture-aware tools.

To summarize, the collapsing process is powerful but simple to implement, requiring only the use of well-defined and contained pre- and post-processing procedures during the ingestion of conjecture-based datasets. This allows conjecture-based datasets to be used together with more classic RDF- and OWL-based datasets currently published as Linked Open Data.

## 5 Representing contexts in the Semantic Web

Much of the past and current data modelling efforts towards the description of cultural objects are focused on being able to represent an increasing amount of details. In [2] we provide an extensive survey of the depth and richness of many data models for the description of cultural heritage and historical sources.

Although these models expose a complex, multidimensional and interconnected landscape, none of them addresses the issues discussed here, because of their implicit assumption that data description should be neutral and objective.

Challenges to the notion of neutrality of data models can be seen in HiCO[6], where interpretation acts are used to express the provenance of interpretations. The Mauth tool [7] expresses the authoritativeness of existing statements with explicit paternity, let users become an active of the description process. More recent projects [9,8] demonstrate the possibility of a semantic enrichment through a contexts-aware approach, especially in the Linked Open Data workflow.

[6] Historical Context Ontology: http://purl.org/emmedi/hico

Similarly, ontologies such as PROV-O[7], for describing the provenance, those in the SPAR family[8] (FaBiO for publishing data, PRO for managing roles, CiTO for representing citations) or GAF[9], for linking mentions, can all be used to add pieces of context to data descriptions. Further, the CMV+P document model [1] makes the very fact that most cultural artefacts reference or embed other artefacts explicit and integrable into other data models. The use of graphs to contextualize specific statements is employed also by nano-publications [12], although not to block the assertion of the statement, but just to enrich them.

Furthermore, the recent RiC-O[10] model uses the similarly named notion of "contexts" to indicate a plurality of paratextual information that translate the classical siloed approach into a graph of connections between vocabularies and ontologies. Additionally, contexts and conjectures can be seen as a generalization of the Factoid model [4] 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 for sources, and the consequent need for contextualization, is reflected in the design of modern APIs for querying historical datasets [14,3].

## 6 Conclusions

The inability of current technologies such as RDF to express multiple contrasting points of views over the same cultural artefact is causing an impoverishment of the datasets published by GLAMs. In addition, the fact that each collection can state only a single point of view is making it harder (and often impossible) to interlink datasets that use, voluntarily or involuntarily, different points of view.

To address this issue, we propose a combination of two approaches. First, *contexts*: a method to allow the expression of conflicting opinions on all aspects of an artefact. Second, *conjectural graphs*: a way to encode scholarly statements as qualified arguments, and not as absolute universal truths, as required by RDF.

In addition, we present a mechanism, named *collapse to reality*, that makes it possible, in a limited and controlled way, to turn specific conjectural graphs into plain RDF statements that can be linked to and made interoperable with other RDF-based datasets.

In contrast to other techniques, our approach does not require the publishers of datasets to settle on a single point of view upfront. Instead, institutions can publish rich datasets containing multiple interpretations of the same data, leaving to the users or to the scholars the ability to accept certain conjectures while refuting others.

The ability to express conflicting opinions, different hypotheses and, in general, all that is part of a lively and meaningful scholarly debate, should represent a key requirement for all data formats used to describe cultural artefacts.

---

[7] W3C PROVenance Interchange Ontology: https://www.w3.org/TR/prov-o

[8] Semantic Publishing and Referencing Ontologies: http://purl.org/spar

[9] Grounded Annotation Framework: http://groundedannotationframework.org

[10] ICA Records in Contexts Ontology: https://www.ica.org/standards/RiC/ontology

This is not currently the case. We hope that the proposed contexts and conjectural graphs will provide a productive start point for a theoretical, practical and methodological reflection on how to beyond the current state of the art and address what is probably the biggest challenge that cultural institutions face when it comes to reaching and engaging scholars and the citizens alike.

## References


1. Barabucci, G.: The CMV+P document model, linear version. In: Bleier, R., Winslow, S.M. (eds.) Versioning Cultural Objects: Digital Approaches. IDE (2019), ISBN 978-3750427020
2. Barabucci, G., Tomasi, F., Vitali, F.: Modeling complexity in public history and cultural heritage. In: Noiret, S., Tebeau, M., Zaagsma, G. (eds.) Handbook Digital Public History. De Gruyter (2021), ISBN 978-3-11-043922-9
3. Barabucci, G., Zingoni, J.: PROSO: prosopographic records. In: Tomasi, F., Vitali, F. (eds.) Proceedings of the 1st International ACM Workshop on Collaborative Annotations in Shared Environment: Metadata, vocabularies and techniques in the Digital Humanities, DH-CASE@DocEng 2013, Florence, Italy, September 10, 2013. pp. 3:1–3:7. ACM (2013). https://doi.org/10.1145/2517978.2517982
4. Bradley, J., Short, H.: Texts into databases: The evolving field of new-style prosopography. Literary and linguistic computing **20**(Suppl), 3–24 (2005)
5. Carothers, G., Seaborne, A.: RDF 1.1 TriG. W3C recommendation, W3C (Feb 2014), https://www.w3.org/TR/trig/
6. Chen, P.P.: The entity-relationship model - toward a unified view of data. ACM Trans. Database Syst. **1**(1), 9–36 (1976). https://doi.org/10.1145/320434.320440
7. Daquino, M.: Mining Authoritativeness in Art Historical Photo Archives. Semantic Web Applications for Connoisseurship. Ph.D. thesis, University of Bologna (2019)
8. Daquino, M., Giovannetti, F., Tomasi, F.: Linked data per le edizioni scientifiche digitali. il workflow di pubblicazione dell'edizione semantica del quaderno di appunti di Paolo Bufalini. Umanistica Digitale **3**(7) (Dec 2019). https://doi.org/10.6092/issn.2532-8816/9091
9. Daquino, M., Mambelli, F., Peroni, S., Tomasi, F., Vitali, F.: Enhancing semantic expressivity in the cultural heritage domain: Exposing the zeri photo archive as linked open data. ACM Journal on Computing and Cultural Heritage **10**(4), 21:1–21:21 (2017). https://doi.org/10.1145/3051487
10. Doerr, M., Gradmann, S., Hennicke, S., Isaac, A., Meghini, C., Van de Sompel, H.: The europeana data model (EDM). In: World Library and Information Congress: 76th IFLA general conference and assembly. vol. 10, p. 15 (2010)
11. Gandon, F., Corby, O.: Name That Graph or the need to provide a model and syntax extension to specify the provenance of RDF graphs. In: W3C Workshop - RDF Next Steps. W3C, Palo Alto, United States (Jun 2010)
12. Groth, P., Gibson, A., Velterop, J.: The anatomy of a nanopublication. Inf. Serv. Use **30**(1-2), 51–56 (2010). https://doi.org/10.3233/ISU-2010-0613
13. Peroni, S., Tomasi, F., Vitali, F.: Reflecting on the europeana data model. In: Agosti, M., Esposito, F., Ferilli, S., Ferro, N. (eds.) IRCDL 2012, Bari, Italy, February 9-10, 2012, Revised Selected Papers. Communications in Computer and Information Science, vol. 354. Springer (2012)
14. Vogeler, G., Vasold, G., Schlögl, M.: Von IIIF zu IPIF? ein Vorschlag für den Datenaustausch über Personen. In: Sahle, P. (ed.) DHd 2019 Digital Humanities: multimedial & multimodal. (2019)