Models of Data Extraction and Architecture in Relational Databases of Early Modern Private Political Archives
edited by Dorit Raines

The Representation of Historical Uncertainties as the Outcome of Competing and Incompatible Certainties

Fabio Vitali
Alma Mater Studiorum - Università di Bologna, Italy

Valentina Pasqual
Alma Mater Studiorum - Università di Bologna, Italy

Abstract This study delves into the ontological implications associated with 'data' and 'capta', highlighting the profound importance of contextual factors and interpretation within the realm of critical discourse. More specifically, 'data', encapsulating the notion of objective observations of facts, differ from 'capta', which encompass selections, opinions, controversies, and debates. These capta elements contribute interest and added value to scholarly knowledge. The formalised separation of data and capta proves to be pivotal in shaping research practices conducive to a critical approach to knowledge production. Upon an evaluation of the efficacy of existing technologies in addressing this issue, the Conjectures model is introduced as a potential solution. This model aims to explicitly represent capta while concurrently facilitating the evolution of critical discourse within a one Knowledge Graph designed to capture complex information within a particular domain. To underscore our point, an illustrative example is drawn from the field of historiography.

Keywords Conjectures. GLAM. Provenance. Uncertainty. RDF.

Edizioni Ca'Foscari

Studi di archivistica, bibliografia, paleografia 8
e-ISSN 2610-9093 | ISSN 2610-9875
ISBN [ebook] 978-88-6969-919-1 | ISBN [print] 978-88-6969-920-7

Peer review | Open access
Submitted 2023-10-03 | Accepted 2024-01-18 | Published 2025-05-22

DOI 10.30687/978-88-6969-919-1//003

1 Introduction

The motivation that propels the scholar is not solely confined to the pursuit of distinguishing true from false. Rather, it extends to the discernment, within the realm of veracity, of aspects that elicit interest, intrigue, the capacity to narrate a compelling storyline, and the potential to yield significant implications for scholarly work. This implies that our best scholarly stories are not simply the final results of our studies, but also, and most importantly, the narration of how we have reached them. Furthermore, in order to narrate these stories, we must represent the scholarly activity in its complexity, including the points of view and opinions we have worked on, discussed, objected to, found lacking or false, or, possibly, rescued from obscurity and discredit.

In this paper our discourse revolves around methodologies and models, all with the aim of not solely encapsulating the final result of a scholarly investigation, but potentially encompassing the building blocks that substantiate it. This ultimate aim is intended to render the result accepted and appreciated in the full strength of its scholarly complexity. We therefore deem it highly important to correctly represent the two activities of collecting and interpreting data that scholars regularly perform. To this end, we turn to Checkland and Holwell (2005), who distinguish between two distinct approaches to data collection, named 'data' and 'capta'.

'Data' refers to the mass of facts that are given and observed. The term comes from the Latin word *dare*, meaning 'to give'. Data is often collected through observation and measurement of external reality (e.g., the material, the dimensions, the current location of a painting), and while individual data points may not be particularly interesting on their own, it is their collection and analysis that may become important. This approach to data collection is often associated with a realist perspective, where data is seen as an objective representation of the world.

On the other hand, 'capta' refers to the small fraction of the available data that we actively take, filter, select, and interpret. The term comes from the Latin word *capere*, meaning 'to take'. Capta is often associated with a constructivist perspective, where knowledge is viewed as constructed and subjective (e.g., the attribution of the same painting to a specific artist or artistic school). Capta involves searching for relevant information, filtering out irrelevant data, selecting the most important information, and interpreting it based on our situated, partial, and constitutive knowledge. This process of knowledge construction is often seen as a more humanistic approach to research, as it acknowledges the role of the researcher in shaping the knowledge that is produced.

Johanna Drucker (2011) further expands on this distinction, arguing that data and capta have different ontological implications.

Data is seen as representing pre-existing facts, while capta is seen as representing situated, partial, and constitutive knowledge of a constructed nature involving selections, interpretations, and expressions of opinions and points of view. This distinction has important implications for knowledge production, as it challenges the notion of data as an objective representation of the world and highlights the importance of context and interpretation in shaping our understanding of the world.

Despite these differences, to this date both data and capta are inserted and described in the same digital collections without differentiation. This approach neglects the correct representation of complexities in knowledge construction and limits the ability to engage in more nuanced and critical analysis. We argue that it is possible to develop a better critical approach to knowledge production by recognising the differences between data and capta and by incorporating this understanding into research practices. This approach can involve paying attention to one's own situated knowledge and perspectives, recognising the role of interpretation and context in shaping our understanding of the world, and being open to alternative perspectives and interpretations. Ultimately, this approach can lead to a computable, and a more robust model of the world and of the issues that we seek to understand.

Consider for instance the history of the evolving attributions of the (now known as) Hadrian's Wall as our guiding example. Flavius Eutropius, a historian who lived in Rome and Constantinople between 363 and 387 AD, wrote about the building of a wall in the northern part of Britain, under the orders of Septimius Severus, the Roman emperor who reigned between 193 and 211 AD. However, Eutropius had never been to Britain, and had no direct knowledge of the wall: he simply attributed the wall to Severus because he knew that he had been to Britain and died there. The venerable Bede, the English monk and historian who lived between 672 and 735 AD, tried to make sense of Eutropius' claim of a wall built by Severus, and since he knew the one built much further north by Antoninus Pius (142 AD), he inferred that the wall by Severus mentioned by Eutropius was the one situated further south. Given Bede's reputation and weight, his words and attribution were accepted for centuries, even though here and there a few historians raised some doubts.

The turn of the tide arrives in 1840, when the priest and historian John Hodgson, while writing his *magnus opus, History of Northumberland*, challenged the long-standing attribution of the wall to Severus. Given that his book held only peripheral relevance to the topic of the wall, he allocated a solitary footnote in the third volume to express his doubts. Being a thorough scholar, this footnote is a remarkable 173 pages long ground-breaking evidence supporting the attribution of the wall to Hadrian rather than Severus. Hodgson's ideas were initially misread or ignored, but eventually became accepted and adopted as the correct attribution of the wall, until today.

The data in this story are few, and possibly uninteresting: the Hadrian's Wall is 117 km long, about 2.4 m high, and runs from Wallsend to Bowness-on-Solway. But here the capta is the real story. It concerns ancient superficial reporting, imprecise deduction from erroneous information, deference to historical (holy, even) authority preventing the obvious truth to emerge, the stubbornness of one man against centuries of accepted facts, the heterogony of ends in scholarly works, and much more. Reporting just the facts unearths a fraction of the story, and the uninteresting bits of it anyway.

At the same time, uncertainty is a common challenge that scholars face in dealing with capta. Uncertainty can arise from various causes, such as temporary ignorance or evolving information. Scholars must also contend with disagreements and controversies, carry out carefully thought experiments where likely and unlikely scenarios are examined, and many other scholarly challenges that introduce uncertainties in their scientific hypothesis. In such situations, avoiding being reticent is crucial (Barabucci et al. 2021).

Competing datings, attributions, and interpretations are examples of the uncertainties that scholars often face. Competing datings refer to instances where multiple timelines or chronological sequences are proposed for a particular event or object. The implications of such doubts can be significant, as they can impact the understanding and implications of related facts and captas. Similarly, competing attributions refer to instances where multiple authors or creators are proposed for a particular work of art or literature under examination. Their impact and ramifications can extend beyond the question of originality, even impacting the monetary value of the item. Finally, competing interpretations refer to instances where multiple meanings or implications are proposed for a particular piece of information or capta. Their implications and ramifications can be significant, as they can impact the understanding of the nature and relevance of the item.

In this paper we intend to introduce a formalisation of capta utilising Semantic Web technologies. In section 2, several existing approaches have been assessed. In section 3, these approaches have been analysed and their advantages and disadvantages highlighted. In section 3, our own formalisation, called Conjectures, is proposed. In section 4, our findings are outlined, and our conclusions are drawn with some insights for further developments.

2 Related Works

Uncertainty ontologies focus on addressing uncertainties that arise from competing and incompatible certainties in datasets. They operate under the assumption that they are dealing with single data items, whose factuality may not be fully known, with the purpose of measuring and reasoning about quantified uncertainty. Conversely, when facing capta, our underlying premise is that we are dealing with a plurality of competing assertions, each accompanied by its distinct degree of factuality, rendering their simultaneous acceptance untenable. The uncertainty inherent in capta arises from the acknowledgment that we are obliged to contend with multiple points of view, some of which may not be factual or true.

The issue of uncertainty is a unique situation that has recently received significant attention. The nature of uncertainty has been expressed by the W3C Incubator Group on Uncertainty Reasoning (Laskey et al. 2008) in terms of epistemic versus aleatory, objective versus subjective, and contingent versus generic. This uncertainty emerges due to ambiguous, inconsistent, vague, incomplete, or empiric information.

To address this challenge, various ontologies have been developed, such as URREF (Blash et al. 2019), which provide a means to express and reason upon uncertainty. These ontologies offer a way to represent uncertainty in a structured and organised manner, allowing for more effective analysis and decision-making. By utilising these tools, researchers can better navigate the complexities of uncertain data and draw meaningful insights from them. The ontologies have been developed based on the assumption that datasets accurately identify the level, nature, and cause of uncertainty. Consequently, they target data and are designed to measure and reason about quantified uncertainty in single data items, even if their factuality is not known.

However, this approach falls short when dealing with capta, as capta represent multiple competing assertions whose factuality cannot all be accepted. As such, our target is not data but rather capta, and our assumption is that we must deal with the complexity of these multiple competing assertions. Our purpose is to represent this complexity and to acknowledge that uncertainty arises when dealing with multiple points of view that cannot all be accepted as factual and true.

Through the years, several models have been developed to express competing claims in the same Knowledge Base as Reification (Hayes 2004), N-ary relations (Noy, Rector 2006) used for example by CIDOC CRM (Doerr et al. 2007), Singleton Properties (Nguyen et al. 2015), and more recently RDF-star (Hartig 2017), but no common strategy has been defined by the community yet. The way of best expressing capta with either of these models heavily depends on the complexity of the information to be rendered, and the kind of queries on large datasets of uncertain assertions needed to be performed.

In the next section we shall examine how five different syntaxes can express the Hadrian Wall's controversy, and the relative advantages and disadvantages of each.

3 Expressing Capta with Semantic Web Technologies

In this section, the concurring claims about Hadrian's Wall concurring attributions are formalised on behalf of several models, namely reification, n-ary relations, Wikidata statements, named graphs and RDF-star quoted triples.

These statements all express two (competing) sets of statements:

1. According to the Venerable Bede, in his *Historia Ecclesiastica* (702 AD), the wall was built by emperor Septimius Severus in 189 AD.
2. According to John Hodgson, in his *History of Northumbria* (1840), the wall was built by emperor Hadrian in 122 AD.

3.1 RDF Reification

```
:S1 a rdf:Statement.                             :S2 a rdf:Statement.
:S1 rdf:Subject :Wall.                           :S2 rdf:Subject :Wall.
:S1 rdf:Predicate p:creator.                     :S2 rdf:Predicate dc:creator.
:S1 rdf:Object :Hadrian.                         :S2 rdf:Object :Severus.
:S1 prov:wasAttributedTo :Hodgson.               :S2 prov:wasAttributedTo :Bede.
:S1 prov:wasDerivedFrom :HistOfNorthumbria.      :S2 prov:wasDerivedFrom :HistEcclesiastica.
:S1 prov:atTime"1840"^^xsd:Year.                 :S2 prov:atTime"702"^^xsd:Year.

:S3 a rdf:Statement.                             :S4 a rdf:Statement.
:S3 rdf:Subject :Wall.                           :S4 rdf:Subject :Wall.
:S3 rdf:Predicate p:inception.                   :S4 rdf:Predicate p:inception.
:S3 rdf:Object "122"^^xsd:Year.                  :S4 rdf:Object "189"^^xsd:Year..
:S3 prov:wasAttributedTo :Hodgson.               :S4 prov:wasAttributedTo :Bede.
:S3 prov:wasDerivedFrom :HistOfNorthumbria.      :S4 prov:wasDerivedFrom :HistEcclesiastica.
:S3 prov:atTime"1840"^^xsd:Year.                 :S4 prov:atTime"702"^^xsd:Year.
```

Figure 1 Disputed attributions and dating of the wall expressed with reification

The concurring claims are represented in figure 1 through reification (Hayes 2004) **[fig. 1]**, where S1 is a statement attributed to John Hodgson, whose subject is the 'Wall', whose predicate is being 'created by' (p:creator), whose object is 'Hadrian'. S2 is a statement attributed to Bede, whose subject is the 'Wall', whose predicate is 'being created by', whose object is 'Severus'. The same patterns are followed to encode the concurring dating of the wall (S3 and S4), showing that a high number of triples is required to express such concurring claims; in particular each claim is represented by at least 4 triples each, for a grand total of 28 statements. Additionally, reified triples require the introduction of a 'fictitious' class (rdf:Statement) to express the claims, requiring the repetition of contextual information for each claim (e.g. who made the claim, when the claim was made and in which edition it has been published).

3.2 N-ary Relations and CIDOC CRM

In figure 2, the same concurring claims can be represented with n-ary relations (Noy, Rector, 2006) and the CIDOC CRM ontology (Doerr et al. 2007) **[fig. 2]**:

```
:Wall a crm:E24_Physical_Human-Made_Thing

:A1 a crm:E13_Attribute_Assignment;
   crm:P177_assigned_property_of_type
crm:P14_carried_out_by;
   crm:P140_assigned_attribute_to :P1;
   crm:P141_assigned :Hadrian
   crm:P14_carried_out_by :Hodgson.

:A3 a crm:E13_Attribute_Assignment;
   crm:P177_assigned_property_of_type
crm:P4_has_time_span;
   crm:P140_assigned_attribute_to :P1;
   crm:P141_assigned "122"^^xsd:Year;
   crm:P14_carried_out_by :Hodgson.

:P1 a crm:E12_Production;
   crm:P108_has_produced :Wall.

:A2 a crm:E13_Attribute_Assignment;
   crm:P177_assigned_property_of_type
crm:P14_carried_out_by;
   crm:P140_assigned_attribute_to :P1;
        crm:P141_assigned :Severus;
        crm:P14_carried_out_by :Bede.

:A4 a crm:E13_Attribute_Assignment;
   crm:P177_assigned_property_of_type
crm:P4_has_time_span;
   crm:P140_assigned_attribute_to :P1;
   crm:P141_assigned "189"^^xsd:Year;
   crm:P14_carried_out_by :Bede.
```

Figure 2 Disputed attributions and dating expressed with CIDOC CRM (using n-ary relations)

According to CIDOC CRM, as shown in figure 2, the wall is a man-made thing (crm:E24_Physical_Human-Made_Thing). A production event P1 that produced the wall exists. An attribution A1 exists and it regards the identity of the person who realised the event P1, which is 'Hadrian' according to John Hodgson. An attribution A2 exists, referring to the person who realised the event P1, which is 'Severus', according to Bede. An attribution A3 exists, recording the time when the event

P1 occurred, which is the year '112', according to Hodgson. An attribution A4 exists, recording the time when the event P1 occurred, which is the year '189', according to Bede. Similarly to reification, n-ary relations require the inclusion of a 'fictitious' class to express each claim (crm:E13_Attribute_Assignment), and a high number of triples (23). Furthermore, their provenance information needs to be repeated to each claim (e.g., A1 and A3 are both attributed to Hodgson while A2 and A4 are attributed to Bede).

3.3 Wikidata Statements and Ranking

Similarly to the previous cases (reification and n-ary), by using Wikidata syntax and model, as shown in figure 3, we represent each claim as a wikibase:Statement (Erxleben et al. 2014) **[fig. 3]**. For the sake of clarity, we replaced the numerical ids used by Wikidata with meaningful labels. Wikidata has a rich way of representing provenance, including individuals, sources and dates. Additionally, it has interesting ways of reporting both claims asserted as preferred and claims proposed as deprecated through the use of ranking over statements and other syntactic methods. Unfortunately, deprecated and preferred statements are less than 2% out of all statements in Wikidata. Additionally, many preferences simply represent corrections of typos, changes in location or of inventory numbers. Furthermore, the 'nature of statement' predicates use a rich vocabulary to characterise statements. Of the more than 260 different types of natures of statements present in Wikidata, we counted about 60 dealing with uncertainty, such as approximation, controversy, 'disputed', 'dubious', estimate, guess, hypothesis, 'presumably', 'possibly', speculation, 'unconfirmed', 'unknown', 'unofficial', etc. Yet, 'nature of statement' predicates are very rare; they were added to fewer than 0.1% of the collected sample. Their scarcity, the vastly overlapping semantics and labelling of their values does not help in increasing their presence (Di Pasquale et al. 2024).

The first claim (S1) assigns a creator to the wall using the property wdt:creator and the value :Hadrian. Additionally, it provides a reference to support this claim using the Wikidata property wikibase:Reference. The reference is assigned the identifier ref:REF1 and includes the following information: it was stated in the publication *History Of Northumbria*, authored by :Hodgson, and published in the year '1840'. The same pattern is implied with statements S2, S3 and S4. As with reification and n-ary relations, each statement is independent and each must be repeated independently and independently attributed to the corresponding source (respectively, ref:REF1 and ref:REF2).

```
ref:REF1 a wikibase:Reference ;
  pr:statedIn :HistOfNorthumbria;
  pr:author :Hodgson;
  pr:publicationDate "1840"^^xsd:Year .

:Wall wdt:creator :Hadrian.
:Wall p:creator s:S1.
s:S1 a wikibase:Statement;
  wikibase:rank wikibase:PreferredRank;
  ps:creator :Hadrian;
  pq:natureOfStatement :attribution;
  prov:wasDerivedFrom ref:REF1.


:Wall wdt:inception "122"^^xsd:Year.
:Wall p:inception s:S3.
s:S3 a wikibase:Statement;
  wikibase:rank wikibase:PreferredRank;
  ps:inception "122"^^xsd:Year;
  pq:natureOfStatement :attribution;
  prov:wasDerivedFrom ref:REF1.
```

```
ref:REF2 a wikibase:Reference ;
  pr:statedIn :HistEcclesiastica;
  pr:author :Bede
  pr:publicationDate "702"^^xsd:Year .

# no assertion since this is deprecated
:Wall p:creator s:S2.
s:S2 a wikibase:Statement;
  wikibase:rank wikibase:DeprecatedRank;
  ps:creator :Severus;

  prov:wasDerivedFrom ref:REF2.


# no assertion since this is deprecated
:Wall p:inception s:S4.
s:S4 a wikibase:Statement;
  wikibase:rank wikibase:DeprecatedRank;
  ps:inception "189"^^xsd:Year;

  prov:wasDerivedFrom ref:REF2.
```

Figure 3 Disputed attributions and datings of the wall expressed with Wikidata statements

In contrast to previous cases, Wikidata distinguishes between accepted claims (S1 and S3, by Hodgson) and discarded ones (S2 and S4, by Bede). Both claims regarding the Hodgson attribution are ranked as preferred (wikibase:PreferredRank), marking the claims as 'accepted', and therefore stating their logical status validity. Additionally, in order to assert S1 and S3 a new triple is added to both (respectively, :Wall wdt:creator:Hadrian and :Wall wdt:inception "122"^^xsd:Year), meaning that this information can be retrieved by using simple queries in Wikidata SPARQL endpoint. On the other hand, deprecated claims (S2 and S3, by Bede) are marked as wikibase:DeprecatedRank, denoting the claims as 'not accepted', and therefore 'not true anymore'. Additionally, no triple is added to S2 and S4 claims, making them non-asserted claims, meaning that this information cannot be retrieved by using simple queries in Wikidata SPARQL endpoint. In general, rank is not only used for accepted/non-accepted assertions, but also for, e.g., temporal statements (a painting used to be in museum A, with DeprecatedRank, but it is now in museum B (with PreferredRank).

3.4 Named Graphs

“Named Graphs” (Carroll et al. 2005) provides a succinct model compatible with RDF 1.1 (2014) to express claims and their provenance information [fig. 4]. Differently from previous cases, the two attributions can be grouped in just one graph so that each graph shows its provenance just once, thereby reducing the number of triples needed to express the claims. However, the claims in figure 4 can be seen as both equally asserted (read it as ‘the wall was created by Hadrian in 122 according to Hodgson and at the same time the wall was created by Severus in 189 according to Bede’), and the provenance is just some kind of additional information. Named graphs have a great potential of expressivity, but the lack of control over whether the content of the graph is asserted or not somehow prevents a full exploitation of their potentialities.

```
GRAPH :S1 {
  :Wall :creator :Hadrian;
      :inception "122"^^xsd:Year.
}
:S1 prov:wasAttributedTo :Hodgson;
  prov:wasDerivedFrom :HistOfNorthumbria.

GRAPH :S2 {
  :Wall :creator :Severus;
      :inception "189"^^xsd:Year.
}
:S2 prov:wasAttributedTo :Bede;
  prov:wasDerivedFrom :HistEcclesiastica.
```

Figure 4 Disputed attributions and datings of the wall expressed with named graphs

3.5 RDF-star

RDF-star (Hartig 2017) extends RDF 1.1 (2014) syntax in order to express statements without asserting them, representing claims in angle brackets. The statement << The wall was created by Hadrian >> is attributed to Hodgson. The statement << The wall was created in 122 >> is attributed to Hodgson. The statement << The wall was created by Severus >> is attributed to Bede. The statement << The wall was created in 189 >> is attributed to Bede. Neither quoted statement is asserted. In order to assert one of them, the claim is repeated outside of the quote (:Wall p:creator:Hadrian and :Wall p:inception “122”^^xsd:Year). Moreover, every single quoted assertion must be associated separately to each provenance statement implying that there is no simple way to avoid repetitions as with reification [fig. 3], n-ary relations [fig. 4] and Wikidata [fig. 5].

In summary, all the above methods for representing capta have interesting characteristics but also some limitations. One issue is the necessity to incorporate numerous additional fictitious entities such as statements, activities, and events, which can lead to an unnecessarily large number of statements (e.g., see rdf:Statement, crm:E13_Attribute_Assignment, wikibase:Statement). It is somewhat challenging to distinguish between asserted statements (predicates presented as true facts) and those that are only expressed (predicates not associated with a truth value).

```
 :Wall :creator :Hadrian.
<< :Wall :creator :Hadrian. >>                 << :Wall :creator :Severus. >>
  prov:wasAttributedTo :Hodgson;                 prov:wasAttributedTo :Bede;
  prov:wasDerivedFrom :HistOfNorthumbria.        prov:wasDerivedFrom :HistEcclesiastica.

 :Wall :inception "122"^^xsd:Year.
<< :Wall :inception "122"^^xsd:Year. >>        << :Wall :inception "189"^^xsd:Year. >>
  prov:wasAttributedTo :Hodgson;                 prov:wasAttributedTo :Bede;
  prov:wasDerivedFrom :HistOfNorthumbria.        prov:wasDerivedFrom :HistEcclesiastica.
```

Figure 5 Disputed attributions and datings of the wall expressed with RDF-star

Moreover, ontologies tend to include specialised entities for even the most straightforward scenarios, resulting in more complex and indirect representations that require a greater number of statements whose veracity is difficult to determine.

Each Wikidata statement uses 3 triples to express a claim, but differently from reification and n-ary relations, rankings allow one to distinguish between currently accepted and deprecated claims. While the named graphs method uses just one quadruple to express each claim, its semantics do not distinguish explicitly between currently accepted and deprecated claims. RDF-star uses one quoted triple to express a claim and allows for distinction between currently accepted claims (asserted claims) and deprecated claims (non-asserted).

Finally, controversial and debated propositions are often expressed individually (e.g., one reification block per proposition, one n-ary relation per proposition, one Wikidata ranked statement per proposition, one RDF-star per proposition) and no relation is explicitly provided to connect multiple claims belonging to a single theory (e.g., the attribution of the creator of the wall goes hand in hand with the inception date and they cannot be chosen independently). The endeavour of identifying uncertainties, ambiguities, and complex situations is thus exceptionally demanding. These challenges underscore the necessity for a more comprehensive and flexible approach to representing capta, one that can account for the numerous nuances and complexities inherent in these phenomena.

4 Expressing Capta with Conjectures

Our proposal, introduced in 2021 (Daquino et al. 2022), aims to provide a more nuanced approach to representing capta by using a specialisation of named graphs (Carroll et al. 2005). Unlike traditional named graphs, Conjectures do not assert the truth of their content by construction. Instead, we divide statements into three categories based on the level of agreement or disagreement surrounding them.

1. 'Undisputed statements' are those that have not been doubted by anyone so far, although this does not necessarily mean that the claim is true. Such statements are represented using plain RDF 1.1 named graphs.
2. 'Disputed statements' are those that have at least one known source casting doubts on the claim or providing incompatible and competing claims. To represent such statements, we use Conjectural Graphs, which allow for multiple competing claims to coexist within the same graph.
3. 'Settled statements' are those in which the author of the dataset has opted for a single claim over competing alternatives, even while acknowledging the existence of disagreement. These statements are represented using Collapsed Conjectural Graphs.

A complete formal model of Conjectures has been developed. It demonstrates the correctness of our approach both as an extension of RDF 1.1 in its strong form **[fig. 6]** and within plain RDF 1.1 in its weak form **[fig. 7]**.

In the strong form, undisputed statements are stored into plain named graphs. Each named graph is introduced by the keyword GRAPH and the content of each of them is asserted. Consider for example the claims in C1 (Wall:length "117km" and:Wall:height "2.4m") in figure 8, which was never questioned.

Disputed claims are stored into special named graphs introduced by the keyword CONJ, meaning that their content is not asserted. For example, the claim :C2 in figure 8 stores :Wall p:creator:Severus and p:inception "189"^^xsd:Year. In this case, both claims are being part of the now discarded attribution of the wall and therefore recorded as non-asserted. Additionally, with the use of named graphs, it is possible to group those claims which belong to the same theory (e.g., in this case, the wall attribution) and record the provenance referring to both the dating and the attribution of the wall (e.g., :C2 prov:wasAttributedTo:Bede).

Settled claims are stored into special named graphs introduced by the keyword SETTLED CONJECTURE, meaning that their content is asserted while being disputed in the past.

```
GRAPH :C1 {
  :Wall:length "117km".
  :Wall:height "2.4m".
}
CONJ :C2 {
  :Wall:creator:Severus.
  :Wall:inception "189"^^xsd:gYear.
}
:C2 prov:wasAttributedTo:Bede;
  prov:wasDerivedFrom:HistEcclesiastica.

SETTLED :C3 {
  :Wall:creator:Hadrian;
  :Wall:inception "122"^^xsd:gYear.
}
:C3 prov:wasAttributedTo:Hodgson;
  prov:wasDerivedFrom:HistOfNorthumbria.
```

Figure 6 Undisputed, disputed and settled attributions and datings of the wall expressed with Conjectures (strong form)

Figure 6 contains the representation of the claims about the wall dimensions (undisputed claim), the discarded attribution of the wall to Severus by Bede (disputed claim, definitely non-asserted) and the accepted attribution of the wall to Hadrian by Hodgson (disputed but settled claim, part of the debate but definitely asserted) with Conjectures in the strong form.

The weak form of Conjectures presents identical information as the strong form but adheres to the syntax of RDF 1.1.[1] As in the strong form, undisputed statements are stored into plain named graphs as shown in figure 7. Disputed claims on the other hand are formalised in the weak form according to the following definition (Daquino et al. 2022):

> A conjectural graph is a named graph where all triples (*s*, *p*, *o*) are represented with two triples, (*s*, *cp*, *o*) and (*cp*, *conj*:*isAConjecturalFormOf*, *p*), where *cp* is a unique newly minted predicate created specifically for the triple to conjecture.

Conjectures adopt newly-minted predicates used only once, which are mapped to their original predicate via the property *conj:isAConjecturalFormOf*. Similarly to :singletonPropertyOf (Nguyen et al. 2015), the property allows to easily retrieve original predicates. For example, as shown in figure 7, the attribution (:Wall:creator:Severus and :Wall:inception "189"^^xsd:Year) claimed by Bede (graph :C2) is now represented by four triples and each original predicate (:creator and :inception) is mapped to the newly minted predicates (C2:creator and C2:inception) by conj:isAConjecturalFormOf.

Settled claims are recorded according to the following definition (Daquino et al. 2022):

1 A parser of Conjectures in their weak and strong form is available at `http://conjectures.altervista.org/`.

A collapse graph *c*1 consists of two graphs: the first is the conjecture *c*1 and the second is a new graph *cc*1 including all the triples in *c*1 but with their original predicates, excluding *conj*: *isAConjectualFormOf*, and adding the triple (*cc*1, *conj*: *collapses*, *c*1).

As shown in figure 7, the claim by Hodgson is represented with two graphs **[fig. 7]**: :C3 represents the claim in its conjectural form, while :settlementOfC3 asserts the claim. The addition of the triple (:settlementOfC3 conj:settles:C3) explicitly links the two graphs.

```
GRAPH:C3 {
   :Wall C3:creator:Hadrian;
   :Wall C3:inception “122”^^xsd:gYear.
   C3:creator conj:isAConjecturalFormOf:creator.
   C3:inception conj:isAConjecturalFormOf:inception.
}
:C3 prov:wasAttributedTo:Hodgson;
   prov:wasDerivedFrom:HistOfNorthumbria.
GRAPH:settlementOfC3 {
   :Wall:creator:Hadrian;
   :Wall:inception “122”^^xsd:gYear.
   :settlementOfC3 conj:settles :C3.
}
GRAPH:C1 {
   :Wall:length “117km”.
   :Wall:height “2.4m”.
}
GRAPH:C2 {
   :Wall C2:creator:Severus.
   :Wall C2:inception “189”^^xsd:gYear.
   C2:creator conj:isAConjecturalFormOf:creator.
   C2:inception conj:isAConjecturalFormOf :inception.
}
:C2 prov:wasAttributedTo:Bede;
   prov:wasDerivedFrom:HistEcclesiastica.
```

Figure 7 Undisputed, disputed and settled attributions and datings of the wall expressed with Conjectures (weak form)

Since Conjectures in their weak form are compliant with RDF 1.1, they can be easily retrieved with plain and traditional SPARQL queries **[fig. 8]**.

Strong and weak forms aim to provide two (non-alternative) solutions to express critical discourse and theories, preventing technological barriers in their adoption. The full semantics of Conjectures is separately documented (Rolfini 2021), along with a longer dissertation on the structure of Conjectures.

```
SELECT DISTINCT ?conj
WHERE {
 GRAPH ?conj {
   ?item ?conjpredicate ?author .
   ?conjPredicate conj:isAConjecturalFormOf :creator
 }
}
```

Figure 8 SPARQL query retrieving all claims (?conj) and items (?item) referring to discarded attributions

5 Conclusions

In conclusion, facts are not the only important aspect of scholarly knowledge that deserve attention. Capta, which include selections, opinions, controversies, and debates, are what make scholarly knowledge interesting and valuable. However, formally encapsulating these diverse and multifaceted aspects of knowledge in a structured dataset remains a formidable challenge.

It is essential for cultural heritage experts and professionals to call for adequate formalisms and tools to capture and preserve these capta. By increasing the quality and quantity of digital collections with non-objective facts, we can achieve an important objective for our future. All the reification methods presented in this work enable us to make statements about statements within RDF. This implies that in addition to the claimed content (e.g., :Wall C2:creator:Severus), supplementary triples can be included to provide context for the information about the claim itself (e.g., with Conjectures, :C2 prov:wasAttributedTo:Bede; prov:wasDerivedFrom:HistEcclesiastica). Consequently, several connections can be established through various claims, such as indicating agreement, or disagreement, specifying if a claim is based on another, referencing both primary and secondary sources, and declaring motivations and evidence on which the claim is based. Through such contextual annotation, the structure of critical discourse in the Humanities can be formalised and, therefore, retrieved.

This proposal provides a more flexible and nuanced way to represent capta proposing Conjectures in their weak and strong form, acknowledging the complexities and uncertainties inherent in these phenomena. Ultimately, it is through a comprehensive approach to scholarly knowledge that we can continue to expand our understanding of the world around us.

Bibliography

Barabucci, G.; Tomasi, F.; Vitali, F. (2021). “Supporting Complexity and Conjectures in Cultural Heritage Descriptions”. *CEUR Workshop Proceedings*, 2810, 104-15. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2736994

Blasch, E.P. et al. (2019). “Uncertainty Ontology for Veracity and Relevance”. *2019 22th International Conference on Information Fusion (FUSION)*. Ottawa: Institute of Electrical and Electronics Engineers (IEEE), 1-8. https://doi.org/10.23919/FUSION43075.2019.9011402

Carroll, J.J. et al. (2005). “Named Graphs”. *Journal of Web Semantics*, 3(4), 247-67. https://doi.org/10.1016/j.websem.2005.09.001

Checkland, P.; Holwell, S.E. (2005). “Data, Capta, Information and Knowledge”. Hinton, M. (ed.), *Introducing Information Management: The Business Approach*. London; New York; Amsterdam: Elsevier, 47-55.

Cyganiak, R.; Wood, D.; Lanthaler, M. (2014). “RDF 1.1 Concepts and Abstract Syntax”. *W3C Recommendation 25 February 2014*. https://www.w3.org/TR/rdf11-concepts/

Daquino, M. et al. (2022). “Expressing Without Asserting in the Arts”. Di Nunzio, G.M. et al. (eds), *Proceedings of the 18th Italian Research Conference on Digital Libraries* (Padua, 24-25 February 2022). CEUR Workshop Proceedings 3160. http://ircdl2022.dei.unipd.it/downloads/papers/IRCDL_2022_paper_31.pdf.

Di Pasquale, A. et al. (2024). “On Assessing Weaker Logical Status Claims in Wikidata Cultural Heritage Records”. *Semantic Web Journal*. https://www.semantic-web-journal.net/content/assessing-weaker-logical-status-claims-wikidata-cultural-heritage-records-1

Doerr, M.; Ore, Ch.-E.; Stead, S. (2007). “The CIDOC Conceptual Reference Model: A New Standard for Knowledge Sharing”. *Proceedings of the Conference: Challenges in Conceptual Modelling. Tutorials, Posters, Panels and Industrial Contributions at the 26th International Conference on Conceptual Modeling – ER 2007* (Auckland, 5-9 November 2007). ACM: Association for Computing Machinery 83, 51-6. https://dl.acm.org/doi/pdf/10.5555/1386957.1386963.

Drucker, J. (2011). “Humanities Approaches to Graphical Display”. *Digital Humanities Quarterly*, 5(1). http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

Erxleben, F. et al. (2014). “Introducing Wikidata to the Linked Data Web”. Mika, P. et al. (eds), *Proceedings of the Conference: The Semantic Web – ISWC 2014 = 13th International Semantic Web Conference* (Riva del Garda, 19-23 October 2014, part 1). Cham: Springer International Publishing, 50-65. Lecture Notes in Computer Science 8796. https://doi.org/10.1007/978-3-319-11964-9_4

Hartig, O. (2017). “Foundations of RDF* and SPARQL* (An Alternative Approach to Statement-Level Metadata in RDF)”. Reutter, J.L.; Srivastava, D. (eds), *Proceedings of the 11th Alberto Mendelzon International Workshop on Foundations of Data Management and the Web 2017* (Montevideo, 7-9 June 2017), article ID 12. CEUR Workshop Proceedings 1912. https://ceur-ws.org/Vol-1912/paper12.pdf.

Hayes, P. (2004). “RDF Semantics”. *W3C Recommendation 10 February 2004*. http://www.W3.Org/TR/Rdf-Mt/

Laskey, K.J. et al. (2008). “Uncertainty Reasoning for the World Wide Web”. *W3C Incubator Group Report 31 March 2008*. http://www.w3.org/2005/Incubator/urw3/XGR-urw3/

Nguyen, V. et al. (2015). “On Reasoning with RDF Statements about Statements Using Singleton Property Triples”. *arXiv*, 15 September. https://doi.org/10.48550/arXiv.1509.04513

Noy, N.; Rector, A. (2006). “Defining N-Ary Relations on the Semantic Web”. *W3C Working Group Note 12 April 2006*.
https://www.w3.org/TR/swbp-n-aryRelations/.

Rolfini, A. (2021). “Semantics of Conjectures”, *arXiv*, 17 October.
https://arxiv.org/abs/2110.08920