CAA
2018

Florian Thiery & Allard Mees
https://dx.doi.org/10.15496/publikation-87762

Taming Ambiguity - Dealing with Doubts in Archaeological Datasets using LOD

Florian Thiery
Department of Scientific IT
Leibniz-Zentrum für
Archäologie (LEIZA)
florian.thiery@leiza.de

Allard Mees
Department of Scientific IT
Leibniz-Zentrum für
Archäologie (LEIZA)
allard.mees@leiza.de

Abstract

The Linked Data Cloud is full of controlled resources, which quality is in fact difficult to handle. Firstly, each resource collection, e.g. a thesaurus, is cooking its own soup related to its research context. Secondly, conceptualisation of Linked Open Data (LOD) assumes standardised data, but in reality, either generic concepts or real instances exist. Thirdly, archaeological items are usually related to generic instances in the LOD cloud, based on their object oriented nature. Describing these relations by modelling archaeological assumptions causes ambiguities which have to be tamed in order to guarantee data quality which can be reused. In this paper we will demonstrate this in three examples from the databases of the Römisch-Germanisches Zentralmuseum Mainz and are proposing ways to solve the handling of ambiguities with a software framework called Academic Meta Tool (AMT).

Keywords: Linked Open Data; graph modelling; ontology; vagueness; uncertainty

Introduction

Archaeological research implicitly deals with doubts and ambiguities in data modelling, aiming to overcome them. Creating reproducible and comprehensible data for the purpose of re-use, whilst also guaranteeing data quality in archaeological data, involves disclosing any doubts and ambiguities. This could be done in any data modelling strategy, e.g. relational or graph based modelling as Linked Open Data (LOD) defined by Berners-Lee (2006).

The Römisch Germanisches Zentralmuseum in Mainz (RGZM) is increasingly engaged with the topic of LOD and aims to provide transparent data to achieve interoperability. In order to achieve this, the RGZM is setting up a 'central metadata index' for aligning its various specialised distributed databases. In this context, it is trying to control the doubts and ambiguities and model them semantically. Data and implicit knowledge from these databases will be modelled as LOD.

A considerable amount of archaeological objects are bearing figural representations (e.g. humans, animals, plants). In the case of identifying these iconographical items, ambiguities may appear. These ambiguities are often combined with doubts. With regards to 'doubts', and trying to model them, we have to consider that two different types of doubts exist: 'uncertainty' and 'vagueness' (Unold et al. 2017).

Moreover, in archaeological research deploying LOD, normally authoritative repositories and controlled vocabularies are used, suggesting that we create a fixed 'undoubted anchor' in the LOD Cloud in order to enable the usage of this resource as a central node. Each resource collection is biased to its own research context, e.g. the Getty Art and Architecture Thesaurus (Getty AAT) (J. Paul Getty Trust 2017) or the Heritage Data (2013) vocabularies like the 'FISH Archaeological Sciences Thesaurus' (FISH 2018a). Because the LOD Cloud is full of isolated resource collections which are build according to research community specific criteria, the LOD anchoring runs rapidly out of control.

Archaeological items are usually related to generic instances in the LOD Cloud based on their object oriented nature. These relations are described by modelled archaeological assumptions, regularly causing ambiguities which have to be tamed to guarantee data quality and ensure the data can be reused.

In this paper we propose a solution for taming the doubts and ambiguities in LOD using a software framework called Academic Meta Tool (AMT) by Thiery and Unold (2018a). We continue the more theoretical work already done (Tolle and Wigg-Wolf 2015; Bruhn et al. 2015) by providing a low-threshold generic web-based software tool, which enables researchers to define their own research specific ontology, describing vague relations in graphs. On top of this, AMT is able to do reasoning according user defined ontology rules.

After a general introduction to the handling of modelling doubts in the Digital Humanities (DH), LOD, and graphs (cf. section 2), the theoretical concepts of uncertainty and vagueness in graphs are presented (cf. section 3). This is followed by some use cases for doubtful statements in relational database management systems (RDBMS) and LOD (cf. section 4). The actual software tool AMT involved is being discussed (cf. section 5) and some examples of its usage in the context of databases at the RGZM are demonstrated (cf. section 6).

Modelling Doubts

This section gives a short overview on data modelling in the DH using relational RDBMS, NoSQL, including graph structures. Within archaeology, semantic LOD modelling has gained increasingly acceptance. But the problem of modelling doubts has not been solved yet. A possible way to tackle this is discussed below.

Data Modelling in the DH

Data modelling in the DH has changed considerably throughout the last few decades. In the first phase of web-based databases in the mid-1990s, common rigid RDBMS in software such as Microsoft Access, Oracle, Informix, MySQL or FoxPro were considerably popular (Beagrie 1993). Some databases at the RGZM, funded by the European Union back in the 1990s are based on the standards of those days, which nowadays are still accessible and remain important research tools, e.g. NAVIS I-III (RGZM 2002; RGZM 2003). However, what was at that time considered as an extremely innovative project is nowadays experienced as a heavy burden to maintain. In the last few years, the 'classic SQL world' has been enhanced by newly developed open source RDBMS such as PostgreSQL and more specifically, with the geometry extension PostGIS.

Large industrial software and web application companies like Facebook and Google work nowadays with NoSQL storage solutions. The simple data model without the need of defining relations or structures allows for quick and efficient data management in distributed systems (Weber 2019; Meier and Kaufmann 2019). As a side effect, this storage technology is increasingly applied in several (archaeological) DH projects, which are based on NoSQL databases such as MongoDB or CouchDB (Lambers 2017).

Furthermore, a specific NoSQL data modelling in graphs is also becoming popular in archaeology. However, the hardest part is certainly the migration from existing authoritative large data repositories to make them available in formats that enable access to them from various analytical packages (Graham 2014: 43). Data is stored and used in classical graph databases like Neo4j. Also triplestores like RDF4J, Apache Jena Fuseki, Ontotext, Parliament and Virtuoso are now found in archaeology, such as the numismatic Nomisma project (Gruber 2013). Triplestores are based on the concept of the Semantic Web (W3C 2015) and Linked Data modelled using the Resource Description Framework (RDF) by RDF Working Group (2014) as well as the Web Ontology Language (OWL) by the OWL Working Group (2012) in a triple structure, following the rules of defining data according to subject - predicate - object.

Sharing and providing interoperable data as LOD is increasingly envisaged by archaeological institutions. Not only the RGZM, but also large research institutes like the Deutsches Archäologisches Institut, English Heritage and the Getty Research Institute are building web resources based on this kind of software architecture. However, it remains to be seen how popular these techniques will really become in archaeology or whether this

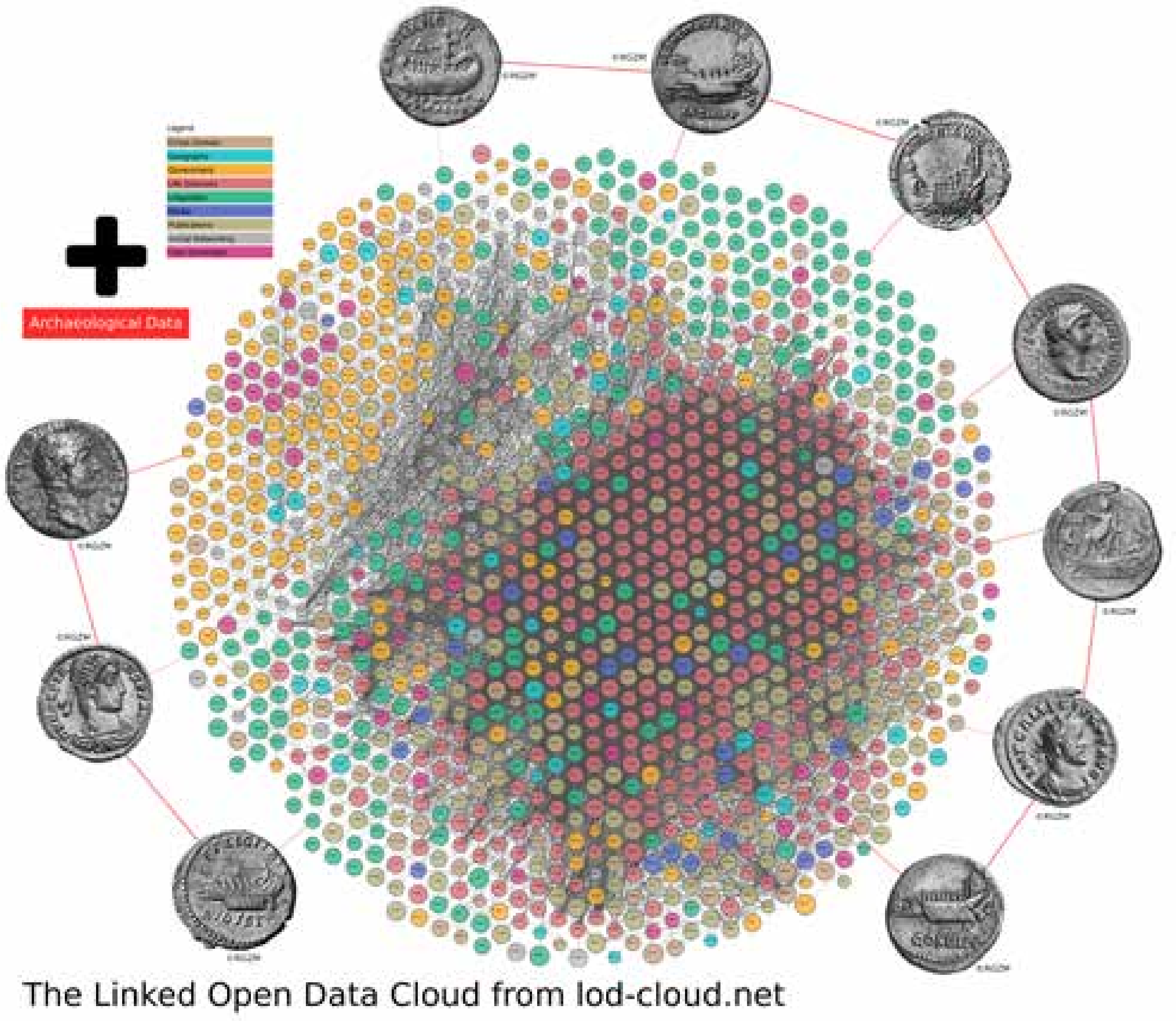

Figure 1. Archaeological LOD Cloud based on the LOD Cloud from lod-cloud.net (CC BY 4.0).

remains limited to the few who can afford the necessary resources. Stepping up these kinds of technologies also implies investing in considerably more advanced human IT resources, since archaeologists themselves are typically not able to handle these technologies themselves anymore. Therefore, it is important to balance the judgement as to whether these techniques should be applied or not. The grapes for graph databases are, by nature, hanging especially low in historical and archaeological branches where social relationships can be modelled with network analysis tools (Deicke 2017; Graham 2014). However, this also implies that there remain large research areas where these techniques do not really make sense - and generating knowledge here can be treated with rigid classical RDBMS methods, cf. chapter 4.1.

LOD in Archaeology

In 2011, Leif Isaksen described the application of Semantic Web technologies to the discipline of archaeology (Isaksen 2011) whilst constructing a foundation for further research in the field of ancient studies, including popular projects like 'Pelagios Commons'. On top of this, some well-known projects (e.g. Nomisma (Gruber and Lockyear 2015), Pelagios Commons (Simon et al. 2016) and PeriodO (Golden and Shaw 2016)) were established within the last few years, and an increasing number of researchers are getting involved. Following from that, the scientific LOD community is continuously growing. Researchers are linking datasets and resources from various sources to create an »Archaeological LOD Cloud« extension (cf. figure 1) as part of the 'Giant Global

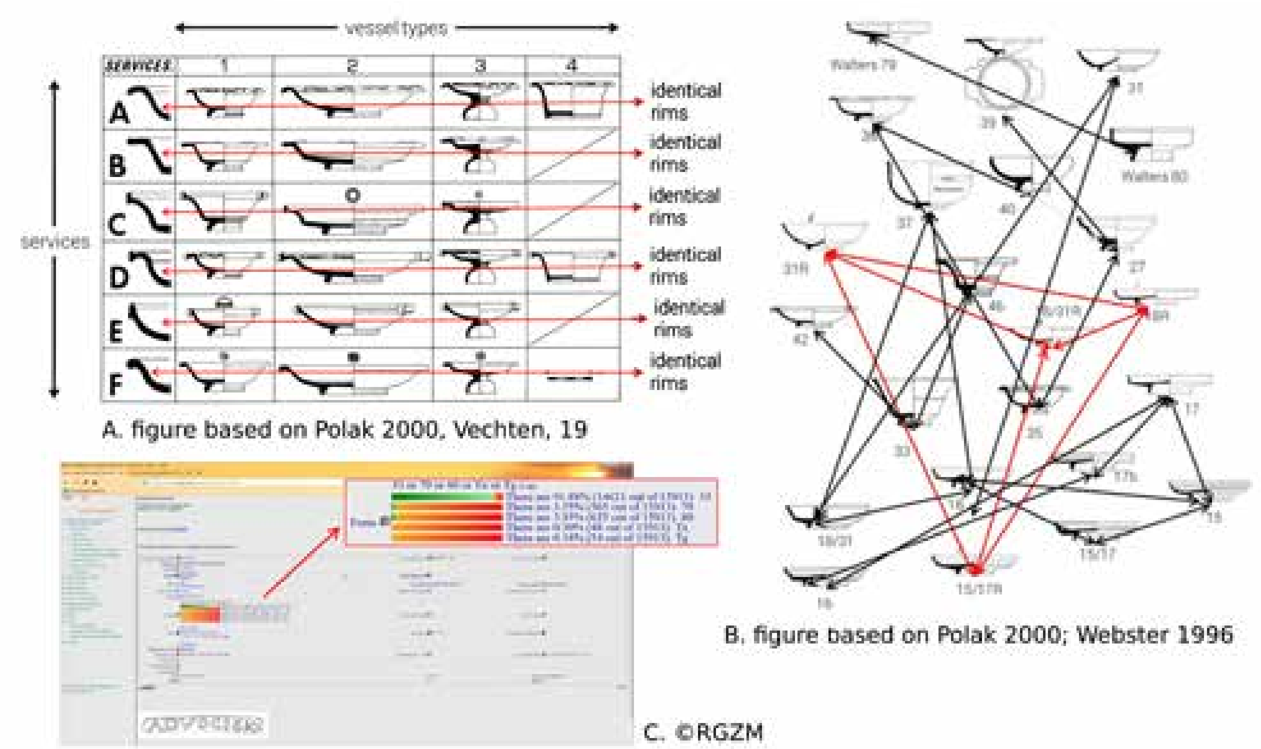

Figure 2. Historical 'service families' (A), archaeological form type attributions (B) and visualisation of degrees of vagueness in Samian Research (C).

Graph' (Berners-Lee 2007), or 'LOD Cloud' (McCrae 2018).

Modelling Doubts in Graphs

Theoretical work on uncertainty and imprecision and their relation to knowledge bases, has been carried out in the 1990s (Parsons 1996; Parsons 1998).

Following from that, Karsten Tolle and David Wigg-Wolf discussed a proposal for semantic modelling as LOD to describe uncertainties in the determination of coin representations (Tolle and Wigg-Wolf 2015). In particular, here the W3C Uncertainty Ontology (W3C-UN) has been used.

W3C-UN is based on the fact that a 'sentence' is subject to 'uncertainty', which has different characteristics: type, nature, derivation and the mathematical model (Laskey et al. 2008).

Moreover, statements or annotations without an exact degree of relation can be solved in the Semantic Web with the 'Open Annotation Ontology'. In this case, two web resources are linked together via a 'body' and a 'target' attribute (Sanderson, Ciccarese & Young 2017). The Pelagios Commons Initiative uses e.g. this ontology for linking data sets and resources of the gazetteer Pleiades (Muccigrosso 2018).

Efforts to enable working with the popular CIDOC CRM ontology (Niccolucci and Felicetti 2018) as a graph resulted in an extension of the CIDOC CRM enabling dealing with attributes of uncertainty (Bruhn et al. 2015).

Modelling Doubts in Thesauri: SKOS

The Simple Knowledge Organization System (Miles and Bechhofer 2009), short SKOS, is a formal language for encoding keywords in thesauri, using RDF and RDFS schema (Brickley and Guha 2014). SKOS offers semantic relations and mapping properties to express vague relationships between skos:Concepts. However, this raises the problem of transitivity and the general problem of 'fuzzy statements' about relations that cannot be quantitatively measured and evaluated:

> *"Note that skos:related is not a transitive property."* (Miles and Bechhofer 2009: #L2344)

> *"A skos:closeMatch link indicates that two concepts are sufficiently similar that they can be used*

Figure 3. Historical 'service families' (A), archaeological form type attributions (B) and visualisation of degrees of vagueness in Samian Research (C).

> *interchangeably in some information retrieval applications. A skos:exactMatch link indicates a high degree of confidence that two concepts can be used interchangeably across a wide range of information retrieval applications.*" (Miles and Bechhofer 2009: #L4858)

Uncertainty and Vagueness in Graphs

As demonstrated in section 2, modelling of doubts by using graphs is a challenge in archaeology. Since the perspective of interoperable and transparent research data for the scientific research community is so promising, it is worth the efforts. In this section, basic concepts and ideas for treating vagueness in graphs are introduced.

Uncertainty vs. Vagueness

There are two different types of doubts: 'uncertainty' and 'vagueness'. Vagueness can be seen as a statement which is not clearly formulated and allows room for individuals to draw their own conclusions. Uncertainty, however, is applicable when the correctness of a statement is not known, but can only be true or false.

Vagueness is a measure of the precision of a statement. A vague statement only applies to a certain extent. For example, if the weather report says »Tomorrow there will be rainfall« it could be a light drizzle, a moderate rain, or a heavy thunderstorm. A remedy could be, for example, the indication of the rainfall.

But a vague statement is not to be confused with an uncertain statement. Within uncertainty, it is completely unknown whether the statement is true at all. For example, if a weather report says, »Tomorrow there is a 75% chance of rain«, this is an uncertain statement. It indicates that in 3 out of 4 cases the message is true, so it rains tomorrow, and in 1 out of 4 cases it is wrong, so it does not rain tomorrow. In Dubois and Prade (2001) a more detailed clarification of the differences between vagueness and uncertainty is described.

In this paper, we only concentrate on vague statements and we assume that all vague statements can be expressed with values somewhere between 0 and 1. For example, a light rain with the value 0.25 (25%) could be used to say »Tomorrow there will be rainfall«. A heavy rain with the value 1.00 (100%) could

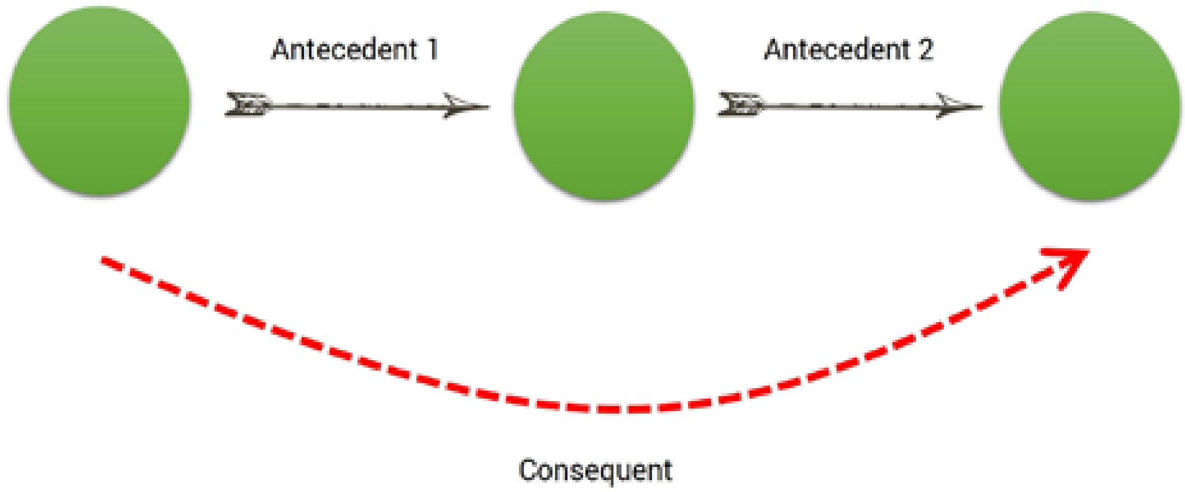

Figure 6. Schematic representation: Role-Chain-Axiom, green dots: concept instances, black arrows: role properties (antecedents), red arrow: inference (consequent).

be used to say »Tomorrow there will definitely be rainfall«

Vagueness in Graph-Based Data

Vagueness can theoretically occur in different places in a graph, but the range of vagueness in graph databases can be complex. The most common case is the assignment of a weight to a vague edge. This expresses to what degree or with what intensity the connection between the two nodes connected by the edge exists.

Analogously, it is also possible to provide additional information in a graph with a vague value. However, in this case we limit ourselves in our use of vagueness to the edge weight. In fact, in this paper, only values between 0 and 1 can be assigned a weighting. Such edge weights can be stored relatively easily in a decimal format in graph databases. More interesting is the actual processing of the edge weights by making rule-based conclusions, which means creating new edges that also carry vagueness.

Processing of the Edge Weights by Rule-based Conclusions

To process the edge weight by rule-based conclusions, techniques commonly utilised for description logics are used and applied to graph-based data as RDF. This has the advantage that the resulting graphs are directly connectable to other ontologies and LOD. A transformation of a vague description logic, interpreted in propositional logic, is realised through the use of multi-valued logic (Lukasiewicz and Straccia 2008). The disadvantage of multi-valued logic is that within them not all rules of classical propositional logic can apply, such as the 'law of De Morgan' or the 'double negation'. Although this disadvantage is persistent, various multi-valued logics can be used in combination.. This makes it possible to assign an individual interpretation to each rule.

Use Cases: Doubtful Statements in RDBMS and LOD

In this section, we describe ambiguous statements about Roman objects. All examples stem from work we did on this modelling issue during the last few years, available in online databases.

Use Case: Modelling Doubts in Samian Research

The Samian Research database comprises more than 245,000 stamped vessels on Roman Terra Sigillata (also called Samian). In antiquity, this pottery was highly standardised and in several cases even conceived as 'service families'. A 'service family' can e.g. consist of a cup, a dish and a bowl, having the same kind of rim or footring (Polak 2000). In modern times, archaeologists define different sub types

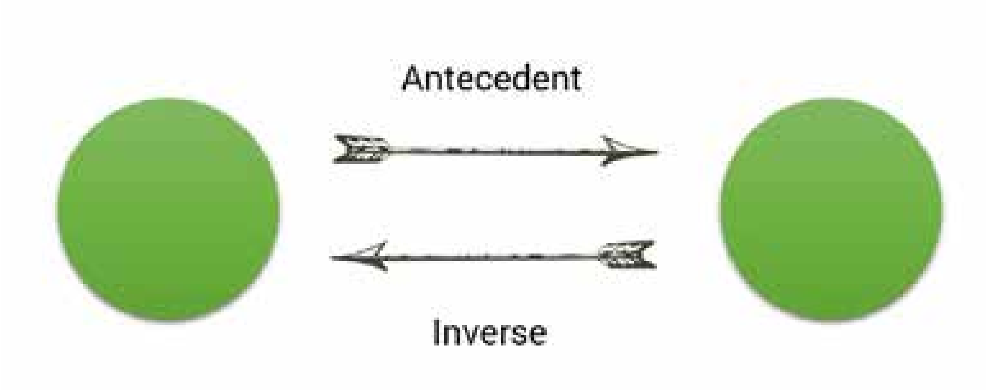

Figure 5. Schematic representation: Inverse-Axiom, green dots: concept instances, black arrows: role properties (antecedent / inverse)

(cf. figure 2B). Because the potters created e.g. identical rims across different pot forms, for modern archaeologists it is easy to identify a 'service family' if only the rim of a pot is preserved. But with only a rim or a footring in ones hand, the specific sub type can frequently not be identified (cf. red lines in figure 2A). Since archaeologists usually find broken parts of vessels (only a footring or only a rim), it is frequently not possible to attribute a pot fragment to one specific form type and usually there remains a (limited) range of form type possibilities (cf. red lines in figure 2B), resulting in vagueness within the typological attribution. Thus, trying to map pot fragments of Terra Sigillata to historically defined concepts of 'types' or 'service families' or even aligning these with typologies, frequently ends up in modelling doubtful assumptions. Typically, identical footrings occur on rouletted dishes of different pot form types (cf. red lines in figure 2B). The Samian online research community uses abstract 'OR' strings in the RDBMS world to model this vagueness. Such an 'OR' statement on its own is not particularly meaningful, but a function within the SQL database querying routine can provide statistical metrics and to specify the degrees of vagueness of the possible form type attributions involved. A simple visualisation of the vagueness degree in coloured bars, indicating the statistical likelihood of the possible pot form attributions, is easily interpretable for the scientific community (cf. figure 2C) (RGZM 2018).

Use Case: Modelling Doubts in the NAVIS Ship Databases

The NAVIS ship databases I, II and III are comprised of thousands of images that show Roman ancient shipwrecks, ship depictions on ancient monuments and coins to be described using SKOS.

Use Case: is it Nero?

A Roman coin with the head of Nero within the NAVIS III database can be described by using a LOD resource (e.g. Nomisma.org (2017)). The question »Is it Nero?« can easily be answered by using skos:exactMatch =100% Nero (cf. figure 3A).

Use Case: Sailed or Rowed?

Another example from NAVIS III can be used for describing the propulsion of a ship on a coin. In this ship depiction (cf. figure 3B), there is vagueness involved. It is not clear whether this kind of ship was sailed or rowed, since there are also fresco depictions where this ship type is displayed in a sailing mode. Therefore, we could describe the situation by using skos:relatedMatch ≥50% sailed ≥50% rowed (cf. figure 3B-C).

Use Case: Lateen Sail or Foresail?

A more difficult example (cf. figure 3D) from the NAVIS II database deals with the 'sailing gear' on a relief. The triangular sail in front of the scene can be described as a triangular lateen sail being used as a 'foresail', a very unlikely scenario. Hence the idea that it could actually be a squared foresail that is being hoisted, this being the more likely scenario because of the adjacent person. It is important to distinguish between the sail types, because there are completely different functions attached to them. From the contextual evidence, this can be solved as follows: skos:relatedMatch ≥1% 'lateen sail' ≥99%' fore sail'.

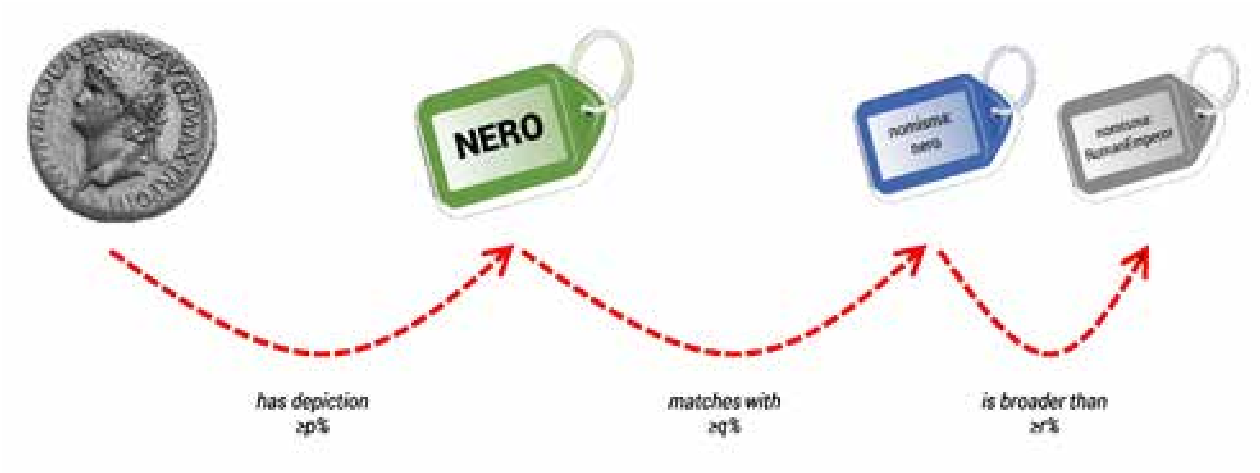

Figure 6. Schematic representation: NAVIS ontology connections.

Use Case: Transport Vessel or Military Vessel?

A further example from NAVIS II (cf. figure 3E) describes the ship function on the so called 'Neumagen relief'. This is an example of vagueness: the ship could have been either used for warfare (the ship type is known on coin depictions with military contexts only) or trade (looking at the loaded wine barrels). Therefore, the available options are: transport vessel or military vessel, in which case we use skos:relatedMatch ≥40% 'transport vessel' ≥60% 'military vessel'.

Use Case: Linking a Triangular Lateen Sail to the LOD Cloud

When trying to link a triangular 'lateen sail' from the NAVIS II database (cf. figure 3F) into the LOD Cloud, it is revealing that each external repository has completely different 'hidden assumptions' in its hierarchies that are related to its specific scientific domain. The internal organisation of the Getty AAT thesaurus or the FISH Maritime Craft Types Thesaurus resources, follow entirely different principles and a correct entry level may be missing. SKOS based relations cannot solve this challenge to model the degree of doubt involved. In such cases, a different approach is required to cope with the hierarchical 'hidden assumptions' implied in these thesauri. As a generic rule, in such cases we can only link a 'lateen sail' to Getty AAT or FISH Maritime Craft Types Thesaurus by skos:relatedMatch ≥p% sails (equipment) ≥q% 'CORVETTE SAIL'.

Academic Meta Tool

Since there is no tool available to our needs in order to model vagueness and ambiguity in combination with reasoning using LOD techniques, we implemented it ourselves. AMT provides web based functions for modelling doubts as LOD including reasoning. It is developed by Martin Unold and Florian Thiery from the Mainz Center for Digitality in the Humanities and Cultural Studies (mainzed), the Institute for Spatial Information and Surveying Technology (i3mainz) and the RGZM (Thiery and Unold 2018a).

AMT Meta Ontology

When using AMT, it is necessary to develop an ontology (Thiery and Unold 2018b; Thiery and Unold 2018c). This ontology describes the schema and axioms for a particular application scenario. In the ontology four types of statements are available.

For demonstrating the individual types of statements we are using an example ontology to model locations that face each other in different directions. To compare the AMT ontology with OWL, we align it with corresponding expressions in the Web Ontology Language (Hitzler et al. 2012).

First of all, it is possible to specify categories for nodes. We also call such categories 'Concepts'. This corresponds to the predicate owl:Class. Each concept can be assigned a name and a short description. For example, the concept of a 'Place' as a point or an area on Earth's surface can be assigned to a place instance

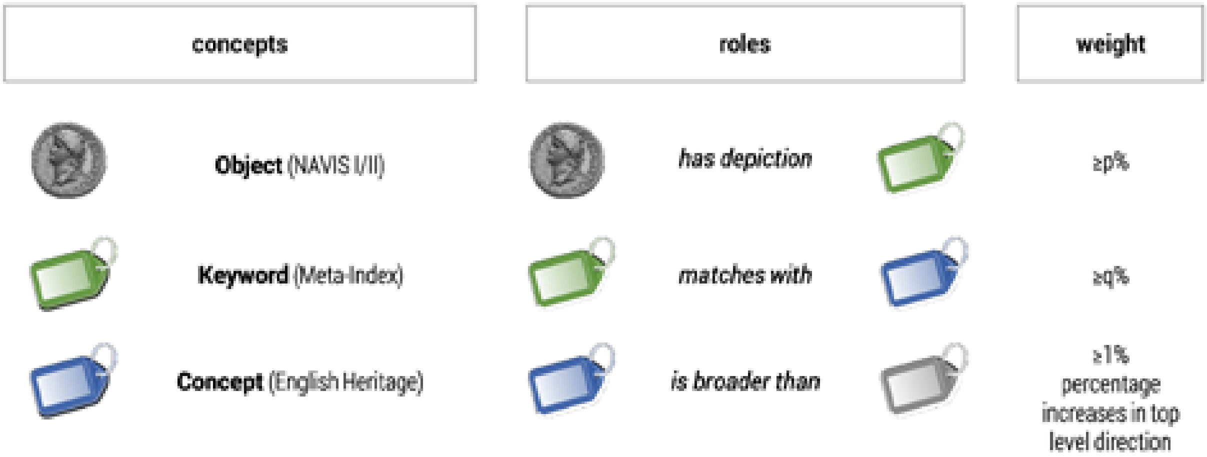

Figure 7. Schematic representation: NAVIS ontology concepts and roles.

'Budapest' in its location at 120 AD, which has a different location compared to the medieval or modern 'Budapest'.

Analogously, categories for edges can also be specified. We call these Roles. This corresponds to the predicate owl:ObjectProperty. Each role can be assigned a name and a concept for source nodes (=rdfs:domain) and destination nodes (=rdfs:range). Example: the roles 'northOf', 'eastOf', 'southOf' and 'westOf' which have a 'Place' as a source and a destination node.

In addition, two types of axioms can be formulated. One type is the role-chain-rule (cf. figure 4). It roughly corresponds to owl:ObjectPropertyChain. However, in addition to specify the roles in the chain, AMT must directly declare the resulting role. Moreover, it must also be determined according to which multivalued logic (e.g. Goedel) the reasoning should take place.

The other type of axiom is the inverse (cf. figure 5). It corresponds to owl:inverseOf. Here you have to specify the role and its inverse. Example: 'northOf' is the inverse of 'southOf'.

AMT JavaScript Library

For web implementation of use cases, a JavaScript library can be used (Unold and Thiery 2018). The amt.js library provides data management functionality, communication with a database server (here: RDF4J triplestore) and a reasoning program. However, each example ontology requires the implementation of an individual web viewer to display and edit the data.

AMT Example of the NAVIS Ship Databases

The aim of this paper is to demonstrate a proposal related to defining keywords for subject indexing depictions of several items within the NAVIS ship databases in the central-index of the RGZM, aligning them to authoritative thesauri to obtain additional information like the hierarchical information (cf. figure 6).

One attempt to find a solution for semantic modelling of uncertainty using Linked Data was done by Tolle and Wigg-Wolf (2015) when dealing with ancient coinage: On a Roman coin, a portrait of a person is shown. Important for further usage is the clear identification of the person. In a survey of experts a 100% certainty could not be ensured: »I am 80% sure that the portrayed person is Titus, or the likelihood is 60% Titus and 40% Nero« (Tolle and Wigg-Wolf 2015: 173). This result was achieved by letting the scholars identify the person according to their own standards without permitting them to indicate likelihood. Only in the post processing the statistical distribution of the identifications was used to indicate the likelihood of the identification.

A similar problem arises in the NAVIS ship databases of the RGZM. In NAVIS II, depictions of ships on mosaics, monuments, etc. are made available on the Web (RGZM 2002); in NAVIS III, ship representations on coins are made available to the scientific community (RGZM 2003). In both databases, analogous to the case of Tolle and Wigg-Wolf, the representations are assigned to an attribute, e.g.

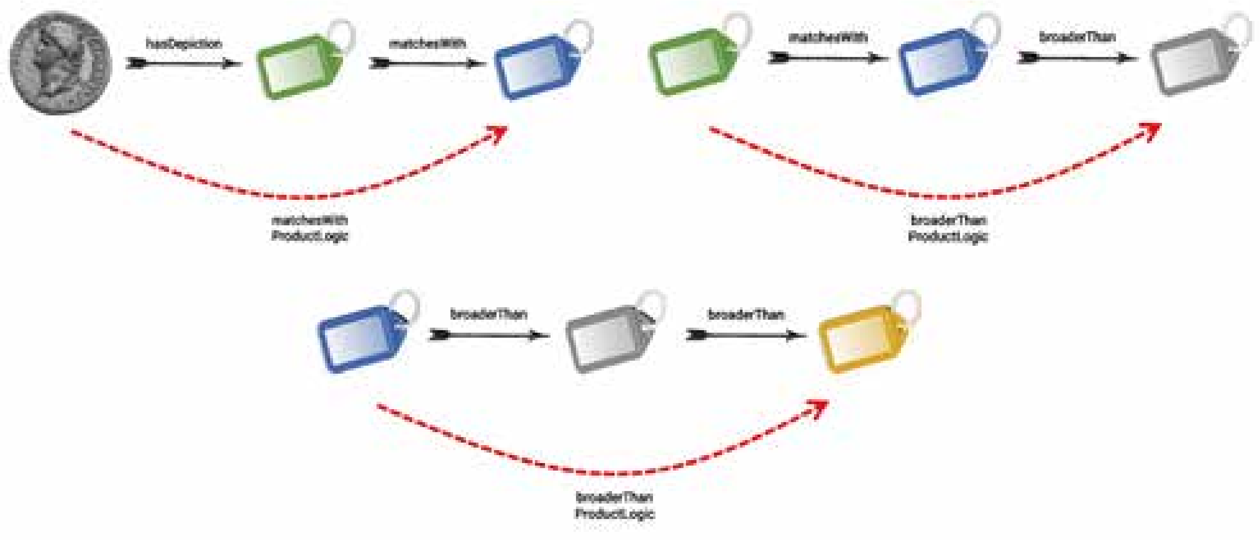

Figure 8. Schematic representation: NAVIS ontology axioms.

Titus and Nero, but also trade and war or paddled and rowed. So far, these links are modelled 1:1 with a 100% possible security in the data model. In order to give objectivity to this very subjective perception, a vague connection that exists only to a certain degree would be transparent and comprehensible. In addition, keywords of object representations are aligned to LOD thesauri concepts. In these thesauri, however, there are again dependencies to a certain degree, which cannot be mapped exactly by means of the used SKOS ontology. However, this is necessary for the content development. Using AMT is a way to semantically model the process from looking at the object for keyword tagging to linking to a thesaurus concept.

AMT NAVIS Ontology

For this use case, a small AMT NAVIS ontology must be implemented. It consists of three concepts, six roles and 18 axioms (Thiery and Mees 2018b).

The NAVIS ontology contains the concepts Object (O), Keyword (K) and Concept (C). The roles (O)-[hasDepiction]->(K) and (K)-[isDepictionOf]->(O) are used to link the object to the keyword. For the connection between keyword and thesaurus concept there are the roles (K)-[matchesWith]->(C) and (C)-[matchedBy]->(K), as well as for the hierarchical order in the thesaurus the roles (C)-[broaderThan->(C) and (C)-[narrowerThan]->(C). Here we assume that the degree of linkage increases the further it goes in the direction of the top-level concept, the degree of the other links must be determined by the scientist himself. Figure 7 shows the concepts and roles.

Three role-chain-rules (including the respective inverse) with suitable logics are introduced (see figure 8): These are:

* Axiom01: (A)-[hasDepiction]->(B)-[matchesWith]->(C) => (A)-[matchesWith;ProductLogic]->(C)
* Axiom02: (A)-[matchesWith]->(B)-[broaderThan]->(C) => (A)-[broaderThan ;ProductLogic]->(C)
* Axiom03: (A)-[broaderThan]->(B)-[broaderThan]->(C) => (A)-[broaderThan ;ProductLogic]->(C)

In addition, six inverse axioms and six disjoint axioms are added.

The role-chain-rules in the NAVIS ontology lead to the following conclusions: If an object is tagged with a keyword and linked to a concept in a thesaurus, there is also a certain degree of connection between the object and the thesaurus concept, cf. Axiom01. If this thesaurus concept is a broader concept in the thesaurus, the keyword is also linked to it with a certain degree, cf. Axiom02. In addition, all hierarchically organized keywords in the thesauri have relationships to some degree, cf. Axiom03.

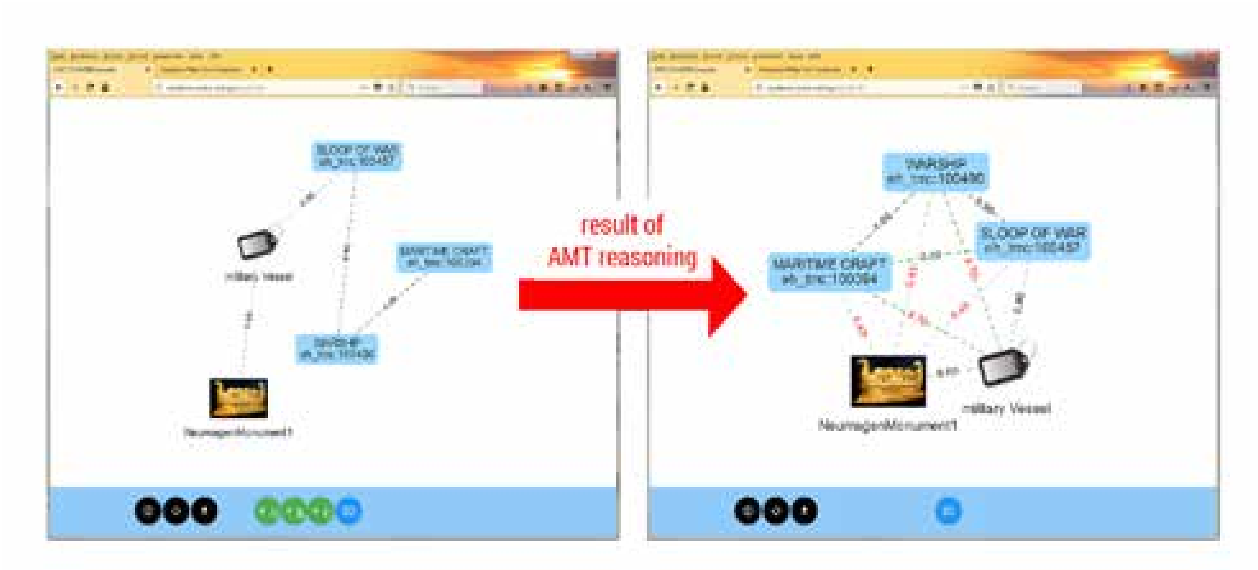

Figure 9. NAVIS ontology reasoning example in the web viewer.

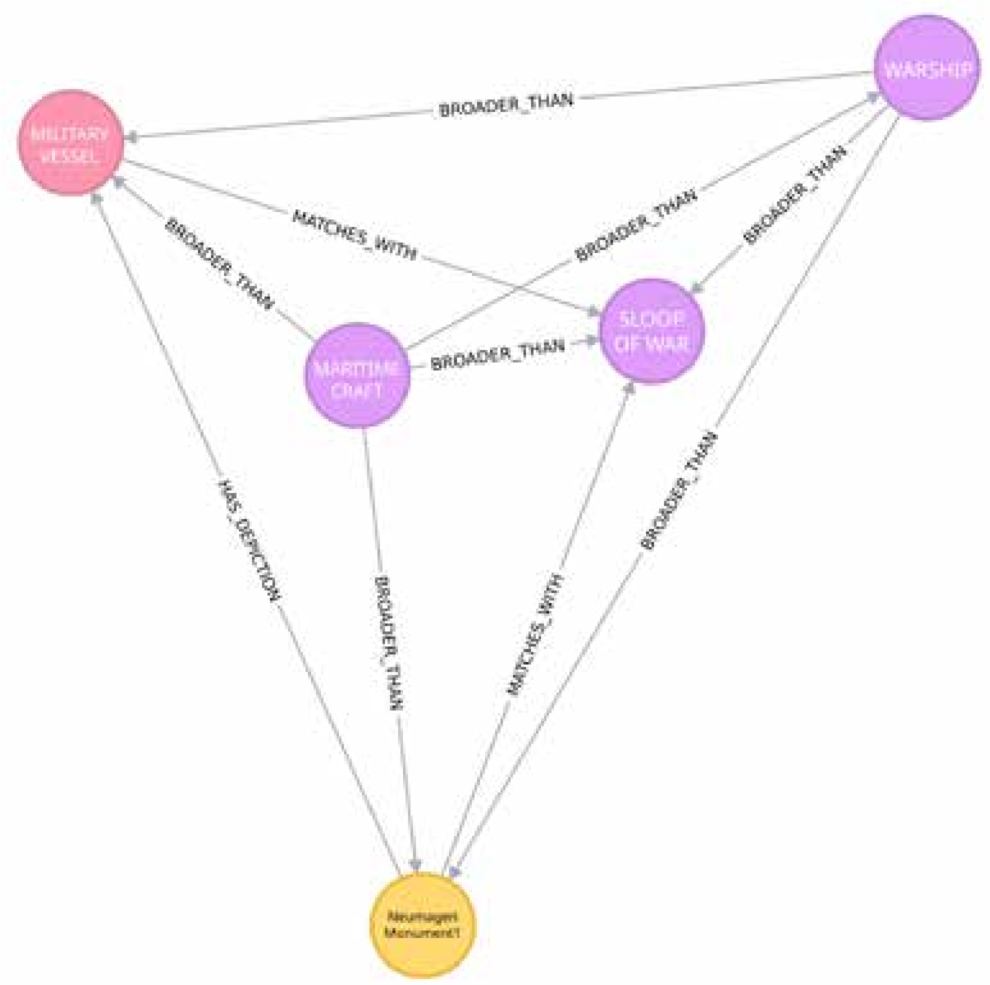

Figure 10: Schematic representation: Inverse-Axiom, green dots: concept instances, black arrows: role properties (antecedent / inverse)

We can demonstrate this in two examples from the NAVIS II ship databases: a depiction shows a ship: Is it a rudder or a sailing ship? Here the scientist can decide for ≥50% sailing ship or ≥50% rowing ship. Another illustration shows a relief. The ship depicted on it could represent a transport or a military ship, since both wine barrels and soldiers can be seen. Again, the scientist can now decide, probably ≥40% transport ship or ≥60% military vessel, cf. figure 3, image B and D.

Example Reasoning in the Web Viewer

AMT reasoning can be visualised on the web (Thiery, Mees & Unold 2018). The visualisation is based on the existing vis.js framework (visjs 2019). This package allows for web based, low-threshold usability. The reasoning implementation is available in a specific amt.js library (cf. section 5.2).

Figure 9 illustrates an example of the question »Is it a military vessel or not?«. On the left, the input graph is visualized. In our example the 'Neumagen Monument 1' Object is connected with a degree of 60% to the 'military vessel' keyword. This keyword matches with 80% to the 'SLOOP OF WAR' item (eh_tmc:100457) in the 'FISH Maritime Craft Types Thesaurus', short eh_tmc (English Heritage 2013). In this thesaurus a hierarchical structure is modelled (FISH 2018b): the thesaurus concept of 'sloop of war' has a broader concept 'WARSHIP' (eh_tmc:100490) and this concept is attached to the top level item 'MARITIME CRAFT' (eh_tmc:100394). As described in chapter 6.1 the degree of linkage increases in the direction of the top-level concept. So here we modelled (eh_tmc:100457) -[90%]-> (eh_tmc:100490) -[100%]-> (eh_tmc:100394). After AMT reasoning, new conclusions can be drawn, cf. figure 9, right side (red numbers): The monument is connected with a degree of 48% to 'SLOOP OF WAR', 43% to 'WARSHIP' and 43% to 'maritime craft'. Our keyword 'military vessel' is connected with a degree of 72% to 'WASHIP' and 72% to 'MARITIME CRAFT'.

The resulting knowledge graphs can be downloaded in different formats (e.g. RDF, JSON, CSV or cypher) for further usage, cf. figure 10.

Outlook

In the last sections we demonstrated that modelling doubts in archaeological research by using an ontology from AMT can help to tame the ambiguities in LOD. However, there still remain challenges and work that will have to be done in the future:

1. Rules in AMT are currently limited, cf. chapter 5. There are ideas of the AMT developers to add more rules in newer versions of AMT. However, AMT rules cannot be extended to the expressive power of OWL because of limitations in the handling of contradictions.

2. The role-chain-rule, even without considering the vagueness, is supported by only a few reasoners, e.g. Straccia (2015), Bobillo and Straccia (2008), Bobillo and Straccia (2011), Tsatsou et al. (2014) but not yet for LOD purposes.

3. Using web standards such as RDF and OWL makes it easy to connect the AMT reasoned data directly to other LOD. Thus, information created with AMT can be linked to other resources and contribute to the enrichment of the Giant Global Graph. Unfortunately, the modelling of vagueness in the Semantic Web is not yet standardized by the World Wide Web Consortium (W3C). Therefore, for the moment there is no way around using an in-house development such as AMT to model vagueness in LOD.

4. As discussed in chapter 3, the AMT ontology only supports vagueness and not uncertainty. The software is therefore suitable for modelling humanities research questions, in which a lot of knowledge is considered as secured, but not all. A classical modelling (without vagueness) fails because of decent categorisation, examples discussed in chapter 6. By using AMT, data modelling is not based on binary decisions (yes or no), but based on decisions that are only valid to a certain degree.

5. The current JavaScript library will be enhanced by a full server based Java library using Apache Jena which will be made freely available to the scientific community.

NAVIS database Update

After finishing the manuscript, the NAVIS I-III databases have been updated and merged into one unified platform using a single CIDOC CRM based data model, called NAVIS (LEIZA 2023: Thiery and Mees 2023a). The thesauri, which were dispersed between NAVIS I, II and III, have been unified and are now available as SKOS based resources 'Maritime Thesaurus' (Thiery and Mees 2023b; Thiery and Mees 2023c).

Acknowledgements

We would like to thank the colleagues of mainzed for their support, above all Martin Unold who developed the AMT core concept (Unold and Cruz 2017; Unold et al. 2017) in the context of his PhD at the Université de Bourgogne, Dijon. We are grateful to Toni Marie Goldsmith and Jonathan Meadows for making this article readable in English.

The example data is published under a CC BY 4.0 licence in the research data repository Zenodo (Thiery and Mees 2018a). The source code of the web viewer prototype is published under a MIT licence (Thiery 2018).

References

Beagrie, N 1993 The Computerisation of the National Archaeological Record, *Revue Informatique et Statistique dans les Sciences Humaines*, 29(1): 9–16.

Berners-Lee, T 2006 Linked Data. 27 July 2006. Available at https://www.w3.org/DesignIssues/LinkedData.html [Last accessed 14 August 2018].

Berners-Lee, T 2007 Giant Global Graph. 21 November 2007. Available at https://web.archive.org/web/20170629153546/http://dig.csail.mit.edu:80/breadcrumbs/node/215 [Last accessed 14 August 2018].

Bobillo, F and Straccia, U 2008 fuzzyDL: An expressive fuzzy description logic reasoner. *2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence)*. June 2008 Hong Kong, China: IEEE, pp. 923–930. DOI: https://doi.org/10.1109/FUZZY.2008.4630480.

Bobillo, F and Straccia, U 2011 Fuzzy ontology representation using OWL 2, *International Journal of Approximate Reasoning*, 52(7): 1073–1094. DOI: https://doi.org/10.1016/j.ijar.2011.05.003.

Brickley, D and Guha, R V 2014 *RDF Schema 1.1.* 25 February 2014. Available at https://www.w3.org/TR/rdf-schema/ [Last accessed 14 August 2018].

Bruhn, K-C, Engel, T, Kohr, T and Gronenborn, D 2015 Integrating Complex Archaeological Datasets from the Neolithic in a Web-Based GIS. In: Giligny, F., Djindjian, F., Costa, L., Moscati, P., and Robert, S. (eds.) *CAA2014. 21st Century Archaeology. Concepts, methods and tools. Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology*. Oxford: Archaeopress, pp. 341–348.

Deicke, A 2017 Networks of Conflict: Analyzing the 'Culture of Controversy' of Polemical Pamphlets of Intra-Protestant Disputes (1548-1580), *Journal of Historical Network Research*, 71-105 Pages. DOI: https://doi.org/10.25517/jhnr.v1i1.8.

Dubois, D and Prade, H 2001 Possibility theory, probability theory and multiple-valued logics: A clarification. *Annals of mathematics and Artificial Intelligence*, 32(1-4): 35–66.

English Heritage 2013 *MARITIME CRAFT TYPE (EH).* 17 July 2013. Available at http://purl.org/heritagedata/schemes/eh_tmc [Last accessed 14 August 2018].

Forum on Information Standards in Heritage 2018a *FISH TERMINOLOGIES. Archaeological Science Thesaurus.* January 2018. Available at http://www.heritage-standards.org.uk/wp-content/uploads/2018/02/ArchSci_alpha.pdf [Last accessed 14 August 2018].

Forum on Information Standards in Heritage 2018b *FISH TERMINOLOGIES. Maritime Craft Type Thesaurus.* January 2018. Available at http://www.heritage-standards.org.uk/wp-content/uploads/2018/02/MarCraft_alpha.pdf [Last accessed 14 August 2018].

Golden, P and Shaw, R 2016 Nanopublication beyond the sciences: the PeriodO period gazetteer, *PeerJ Computer Science*, 2: e44. DOI: https://doi.org/10.7717/peerj-cs.44.

Graham, S 2014 On Connecting Stamps - Network Analysis and Epigraphy. *Les Nouvelles de l'archéologie*, 135(3): 39–44.

Gruber, E 2013 *Nomisma: Using XForms to Manage and Publish Linked Open Data.* 18 July 2013. Available at http://numishare.blogspot.com/2013/07/nomisma-using-xforms-to-manage-and.html [Last accessed 14 August 2018].

Gruber, E and Lockyear, K 2015 From dBase III+ to the Semantic Web: Twenty Five Years of the Coin Hoards of the Roman Republic Database. In: Traviglia, A. (ed.) *Across Space and Time. Papers from the 41st Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA)*. 2015 Amsterdam: Amsterdam University Press, pp. 336–346.

Heritage Data 2013 *Heritage Data Vocabularies.* Available at http://www.heritagedata.org/blog/vocabularies-provided/ [Last accessed 14 August 2018].

Hitzler, P, Krötzsch, M, Parsia, B, Patel-Schneider, P F and Rudolph, S 2012 *OWL 2 Web Ontology Language Primer* (Second Edition). 11 December 2012. Available at http://www.w3.org/TR/2012/REC-owl2-primer-20121211/ [Last accessed 14 August 2018].

Isaksen, L 2011 *Archaeology and the Semantic Web*. Thesis (Doctoral). University of Southampton.

J. Paul Getty Trust 2017 *Art & Architecture Thesaurus® Online*. 7 March 2017. Available at http://www.getty.edu/research/tools/vocabularies/aat/ [Last accessed 14 August 2018].

Lambers, R 2017 *NoSQL databases in Archaeology, a Funerary Case Study.* Thesis (MSc) Digital Archaeology, Universiteit Leiden.

Laskey, K J, Laskey, K B, Costa, P C G, Kokar, M, Martin, T and Lukasiewicz, T 2008 *Uncertainty Reasoning for the World Wide Web.* 31 March 2008. Available at https://www.w3.org/2005/Incubator/urw3/XGR-urw3-20080331/ [Last accessed 14 August 2018].

Leiza 2023 *NAVISone Ship Database.* Available at https://www.leiza.de/navis [Last accessed 20 March 2023]

Lukasiewicz, T and Straccia, U 2008 Managing uncertainty and vagueness in description logics for the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4): 291–308. DOI: https://doi.org/10.1016/j.websem.2008.04.001.

McCrae, J P 2018 *The Linked Open Data Cloud.* Available at https://lod-cloud.net/ [Last accessed 14 August 2018].

Meier, A and Kaufmann, M 2019 NoSQL Databases. In: *SQL & NoSQL Databases*. Wiesbaden: Springer Fachmedien Wiesbaden. pp. 201–218. DOI: https://doi.org/10.1007/978-3-658-24549-8_7.

Miles, A and Bechhofer, S 2009 *SKOS Simple Knowledge Organization System Reference.* 18 August 2009. Available at https://www.w3.org/TR/2009/REC-skos-reference-20090818/ [Last accessed 14 August 2018].

Niccolucci, F and Felicetti, A 2018 A CIDOC CRM-based Model for the Documentation of Heritage Sciences. *2018 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018)*. October 2018 San Francisco, CA, USA: IEEE. pp. 1–6. DOI: https://doi.org/10.1109/DigitalHeritage.2018.8810109.

Muccigrosso, J 2018 *Joining Pelagios.* 22 June 2018. Available at https://github.com/pelagios/pelagios-cookbook/wiki/Joining-Pelagios [Last accessed 14 August 2018].

Nomisma.org 2017 nero (foaf:Person). 28 April 2017. Available at http://nomisma.org/id/nero [Last accessed 14 August 2018].

OWL Working Group 2012 *Web Ontology Language (OWL).* 11 December 2012. Available at https://www.w3.org/OWL/ [Last accessed 14 August 2018].

Parsons, S 1996 Current approaches to handling imperfect information in data and knowledge bases, *IEEE Transactions on Knowledge and Data Engineering*, 8(3): 353–372. DOI: https://doi.org/10.1109/69.506705.

Parsons, S 1998 Addendum to 'Current approaches to handling imperfect information in data and knowledge bases', *IEEE Transactions on Knowledge and Data Engineering*, 10(5): 862.

Polak, M 2000 *South Gaulish Terra Sigillata with Potters' Stamps from Vechten*. Rei Cretariae Romanae Fautorum Acta, Supplementum 9. Nijmegen: Katholieke Universiteit Nijmegen.

RDF Working Group 2014 *Resource Description Framework (RDF).* 25 February 2014. Available at https://www.w3.org/RDF/ [Last accessed 14 August 2018].

RGZM 2002 *The NAVIS II project.* Available at http://www.rgzm.de/navis2 [Last accessed 14 August 2018].

RGZM 2003 NAVIS III. *Römische Schiffsmünzen im RGZM.* Available at http://www.rgzm.de/navis3 [Last accessed 14 August 2018].

RGZM 2018 *Samian research.* Available at http://rgzm.de/samian [Last accessed 14 August 2018].

Sanderson, R, Ciccarese, P and Young, B 2017 *Web Annotation Data Model.* 23 February 2017. Available at https://www.w3.org/TR/2017/REC-annotation-model-20170223/ [Last accessed 14 August 2018].

Simon, R, Isaksen, L, Barker, E and de Soto Cañamares, P 2016 The Pleiades Gazetteer and the Pelagios Project. In: Berman, M L, Mostern, R, and Southall, H (eds.) *Placing names: enriching and integrating gaz-*

etteers. The spatial humanities. Bloomington: Indiana University Press. pp. 97–109.

Straccia, U 2015 The fuzzyDL System. Available at http://www.umbertostraccia.it/cs/software/fuzzyDL/fuzzyDL.html [Last accessed 14 August 2018].

Thiery, F 2018 *Academic Meta Tool - NAVIS example web app (CAA2018)*. DOI: https://doi.org/10.5281/zenodo.1344297.

Thiery, F and Mees, A 2018a *Academic Meta Tool - NAVIS example data (CAA2018)*. DOI: https://doi.org/10.5281/zenodo.1344295.

Thiery, F and Mees, A 2018b *Academic Meta Tool - Navis Ontology.* DOI: https://doi.org/10.5281/zenodo.1341111.

Thiery, F, Mees, A and Unold, M 2018 *AMT. Academic Meta Tool. CAA 2018 Example.* Available at http://academic-meta-tool.xyz/caa2018/ [Last accessed 14 August 2018].

Thiery, F and Mees, A 2023a NAVIS.one: Challenges and Opportunities for RSE and RDM by designing a common web application from a 1990s online database, *Squirrel Papers* 2022, (4)3, No. 15. DOI: https://doi.org/10.5281/zenodo.7142170.

Thiery, F and Mees, A 2023b NAVISone Maritime Thesaurus, *Squirrel Papers* 2022, (4)2, No.4. DOI: https//doi.org/10.5281/zenodo.7353771.

Thiery, F and Mees, A 2023c *NAVISone Maritime Thesaurus* [GitHub archaeolink]. Available at https://github.com/archaeolink/navisone-maritime-thesaurus/blob/main/data/navisone_maritime_thesaurus.ttl [Last accessed 20 March 2023]

Thiery, F and Unold, M 2018a *Academic Meta Tool – Ein Web-Tool zur Modellierung des Zweifels.* DOI: https://doi.org/10.5281/zenodo.1155727.

Thiery, F and Unold, M 2018b *Academic Meta Tool – Ontology.* DOI: https://doi.org/10.5281/zenodo.1342536.

Thiery, F and Unold, M 2018c *Academic Meta Tool - Vocabulary.* DOI: https://doi.org/10.5281/zenodo.1342530.

Tolle, K and Wigg-Wolf, D 2015 Uncertainty Handling for Ancient Coinage. In: Giligny, F., Djindjian, F., Costa, L., Moscati, P., and Robert, S. (eds.) *CAA2014. 21st Century Archaeology. Concepts, methods and tools. Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology.* 2015 Oxford: Archaeopress, pp. 171–178.

Tsatsou, D, Dasiopoulou, S, Kompatsiaris, I and Mezaris, V 2014 LiFR: A Lightweight Fuzzy DL Reasoner. In: *The Semantic Web: ESWC 2014 Satellite Events. Lecture Notes in Computer Science.* 25 May 2014 Springer, Cham. pp. 263–267. DOI: https://doi.org/10.1007/978-3-319-11955-7_32.

Unold, M, Cruz, C, Boochs, F and Bruhn, K-C 2017 *Basic types of non-boolean description logics.* Available at http://unold.net/research/p_dls_20170320.pdf [Last accessed 14 August 2018].

Unold, M and Cruz, C 2017 How to enrich description logics with fuzziness. In: *2017 Computing Conference.* July 2017 London: IEEE. pp. 51–57. DOI: https://doi.org/10.1109/SAI.2017.8252080.

Unold, M and Thiery, F 2018 *Academic Meta Tool - amt.js.* DOI: https://doi.org/10.5281/zenodo.1342310.

visjs 2019 Network. Available at https://visjs.github.io/vis-network/docs/network/ [Last accessed 20 September 2019].

W3C 2015 *SEMANTIC WEB.* Available at https://www.w3.org/standards/semanticweb/ [Last accessed 14 August 2018].

Weber, S 2019 *NoSQL Databases.* Available at https://wiki.hsr.ch/Datenbanken/files/Weber_NoSQL_Paper.pdf [Last accessed 10 September 2019].

Webster, P 1996 *Roman samian pottery in Britain*. C.B.A. Practical Handbook in Archaeology 13. New York.