Leibniz-Gemeinschaft

Römisch-Germanisches Zentralmuseum
Leibniz-Forschungsinstitut für Archäologie

R G Z M

Taming Ambiguity –

Dealing with doubts in archaeological datasets using LOD

Florian Thiery M.Sc.
Dr. Allard Mees FSA

Römisch-Germanisches Zentralmuseum Mainz
Department of Scientific Information Technology

Archaeological research and Linked Open Data deals with doubts and ambiguities which have to be tamed.

Language causes ambiguities. As an example: „*limes*“. Do we mean the *fruit* or the roman *defence wall*?

There are two different types of doubts. *Vagueness* and *uncertainty*. But what is vagueness – what is uncertainty?

RGZM

Leibniz Gemeinschaft

mainzed

Vagueness

The limes are sour.

This statement is imprecisely formulated and allows for individual scaling.

Uncertainty

Lilly eats limes.

The correctness of this statement is not known, but it can only be true or false.

Creating *reproducible and comprehensible data for re-use* AND *guaranteeing data quality in archaeological LOD* also means…

… disclosure of doubts in Linked Open Data!

Example: Paper „*Uncertainty Handling for Ancient Coinage*“ presented at CAA 2014 by Karsten Tolle and David Wigg-Wolf.

Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology
CAA 2014 - 21st Century Archaeology/F. Giligny, F. Djindjian, L. Costa, P. Moscati, S. Robert (eds.)

Uncertainty Handling for Ancient Coinage

Karsten TOLLE,[1] David WIGG-WOLF[2]

[1] Databases and Information Systems (DBIS), Johann Wolfgang Goethe-University of Frankfurt, Germany
[2] Römisch-Germanische Kommission (RGK), Frankfurt, Germany
tolle@dbis.cs.uni-frankfurt.de, David.Wigg-Wolf@dainst.de

Abstract

For archaeologists uncertainty is part of daily business. When data are recorded it is important to concentrate on facts, yet in many cases the recording of uncertain data are mandatory in order to answer research questions. As long as we stay within one system and the number of users is limited, the handling of such uncertain data is up to the individual users.
But when the data are to be exchanged, the question arises as to how to model uncertainty without risking that others fail to notice this uncertainty and build hypothesis based on uncertain data without knowing it.
In this paper we show how we store our uncertain data internally and discuss different approaches to modelling uncertainty based on ontologies with different drawbacks and benefits.

Keywords: Uncertainty, Ontology, Modelling

Tolle/Wigg-Wolf, 2014

We often use *authoritative repositories* like *'controlled'* vocabularies and *'controlled'* resources in Linked Open Data…

… to create a fixed 'undoubted' anchor in the Linked Data Cloud
… to enable the usage of this resource as a central node.

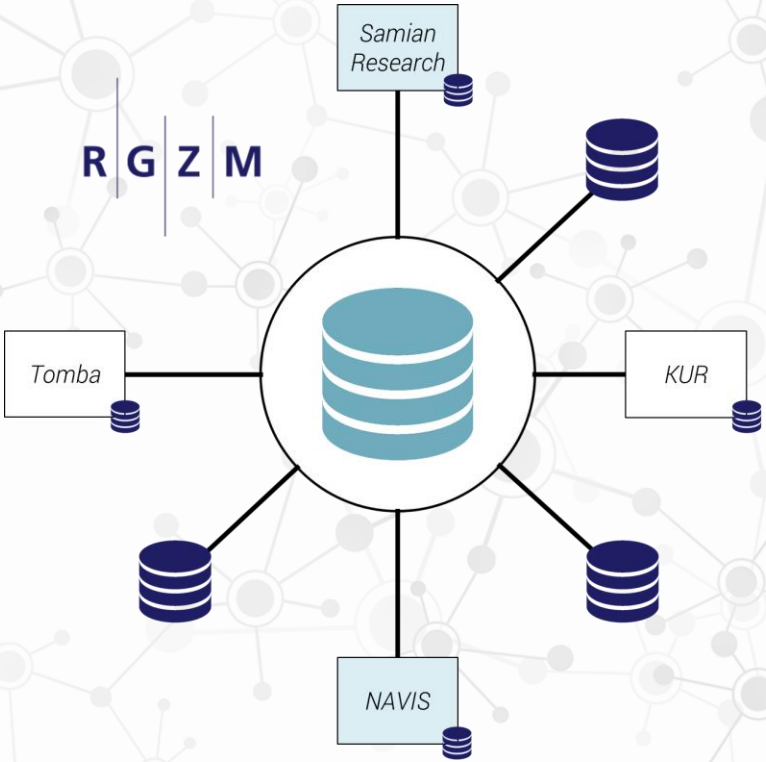
But: The Linked Data Cloud is full of so called *'controlled'* resources, which in fact rapidly run out of control…

Each resource collection, such as a thesaurus, is cooking its own soup related to its research context…

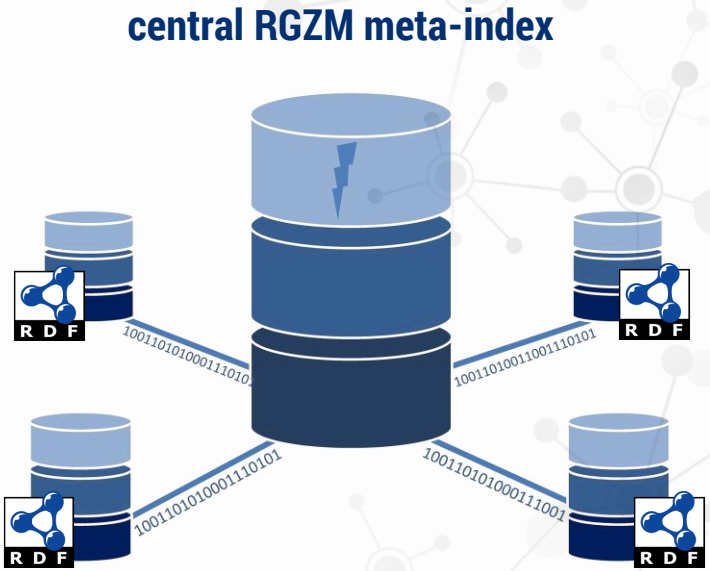
Archaeological items are somehow related to generic instances in the LOD cloud, based on their object orientated nature.

These relations are described by modelled archaeological assumptions, regularly causing ambiguities which have to be tamed to guarantee data quality for reuse.

For aligning our various specialised distributed databases, we are setting up a ‘central index’ containing metadata.



Linking ‘central index terms’ into the LOD Cloud is subject to uncertainties within modelling of relations.



As a result, this process includes dealing with ambiguity challenges.

Use cases: trying to model doubtful statements about Roman objects in RDMBS and graph databases using Linked Open Data.

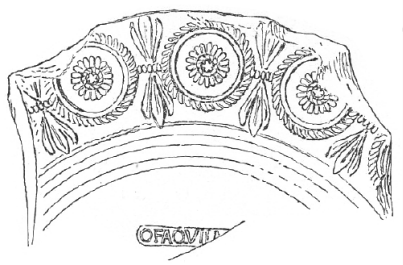


Photo: Samian Research

Photo: NAVIS II, A

Photo: NAVIS III, A

Trying to map pot fragments of Gaulish Terra Sigillata to historically defined concepts of ‘types’ and ‘service families’…

Photo: Mees

…or aligning these typologies, ends up in modelling doubtful assumptions.

Can metrics be used to define ‘type’ attributions?

How can an already existing online research community help?

In Samian studies, the attribution to vessel types is restricted by ‘services’ already established in antiquity.

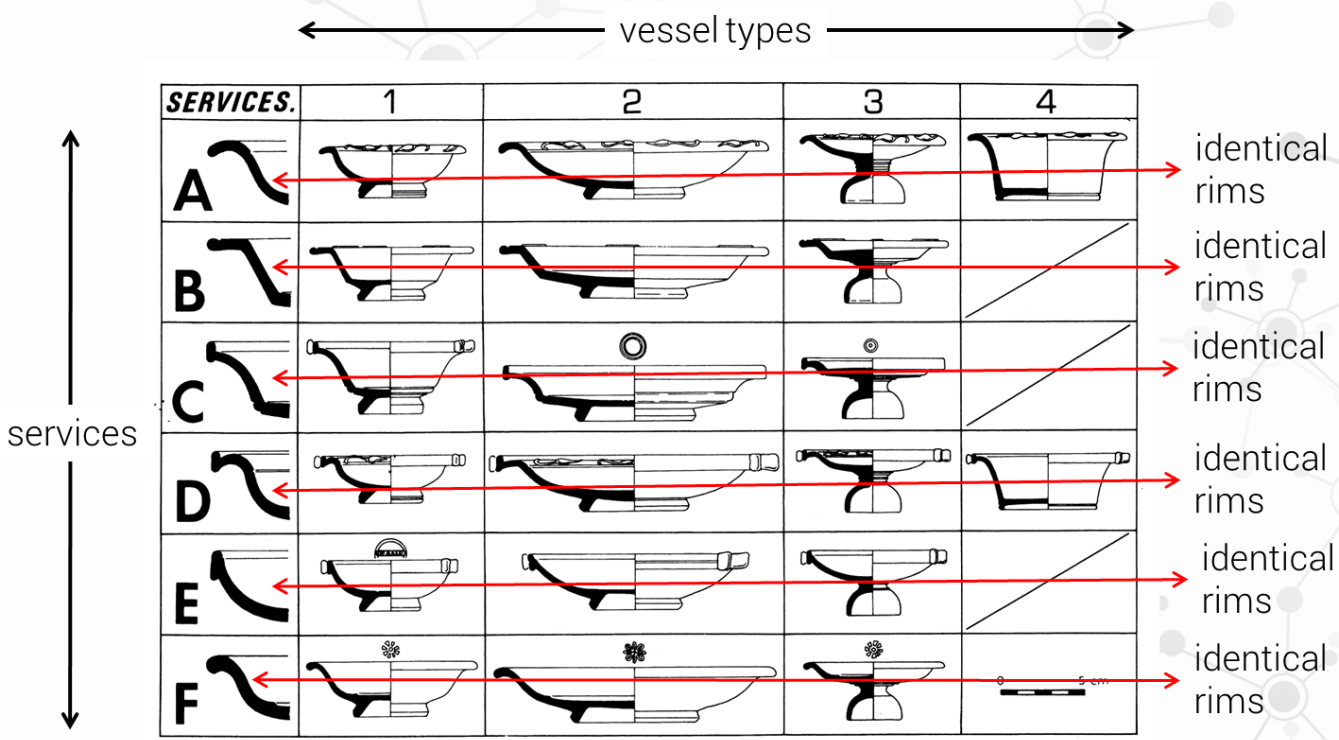
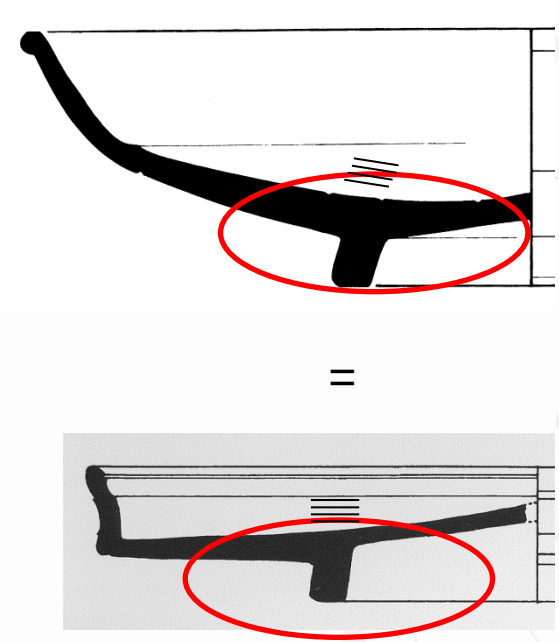
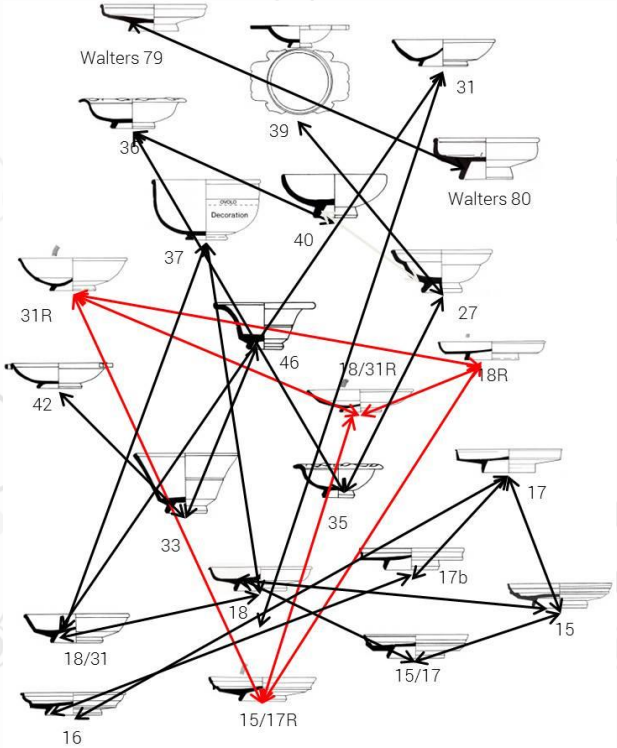


figure based on Polak 2000, Vechten, 19

The uncertain attribution of individual vessels by footrings applies to all Samian services.

Identical footrings on rouletted dishes of different types





figures based on Polak 2000; Webster 1996

The Samian online research community uses abstract „OR“ strings in the RDMBS world to model this uncertainty.

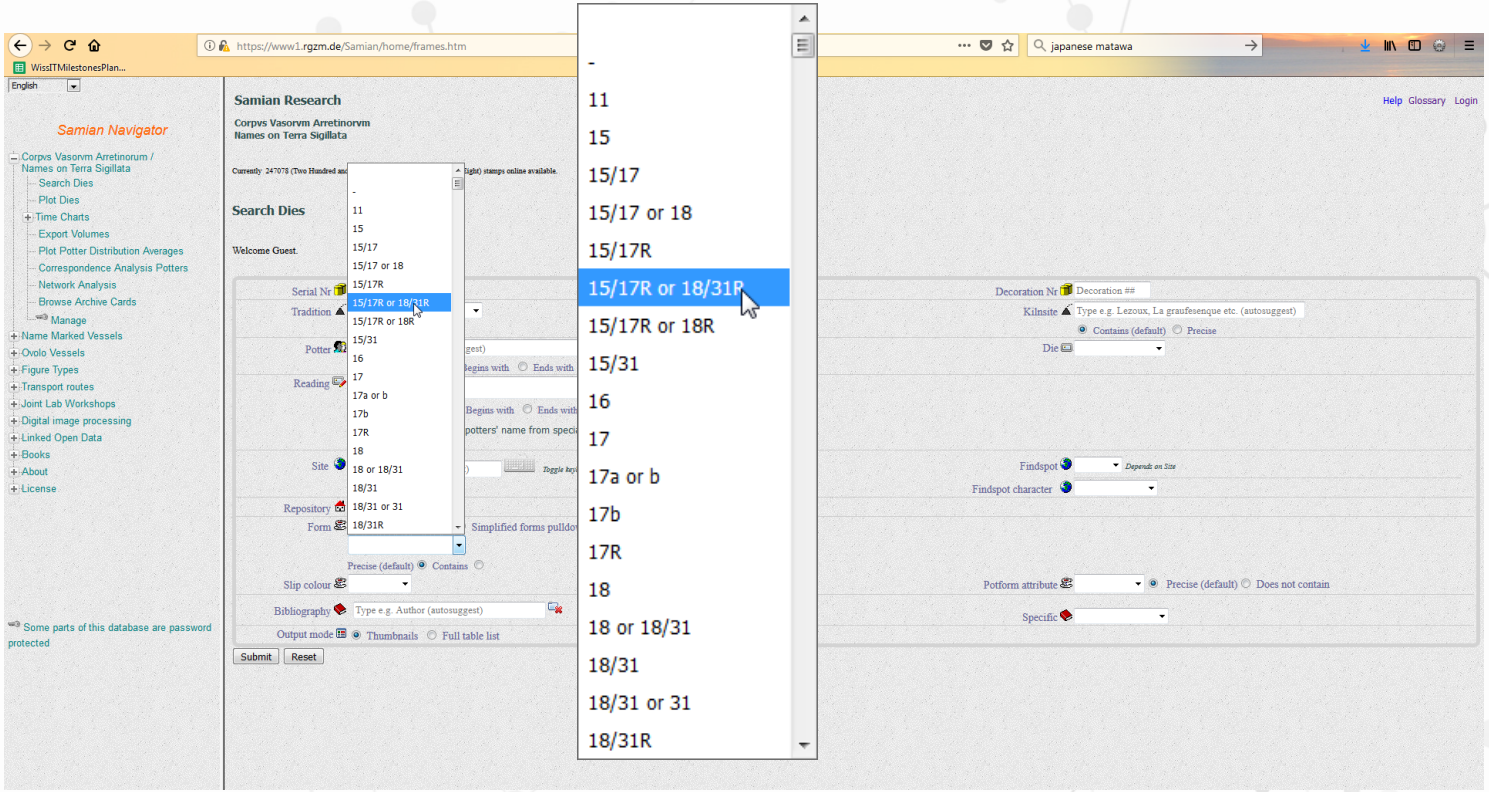


figure based on http://rgzm.de/samian

An effective solution: Creating statistical metrics by specifying the degree of uncertainty and provide this information as LOD.

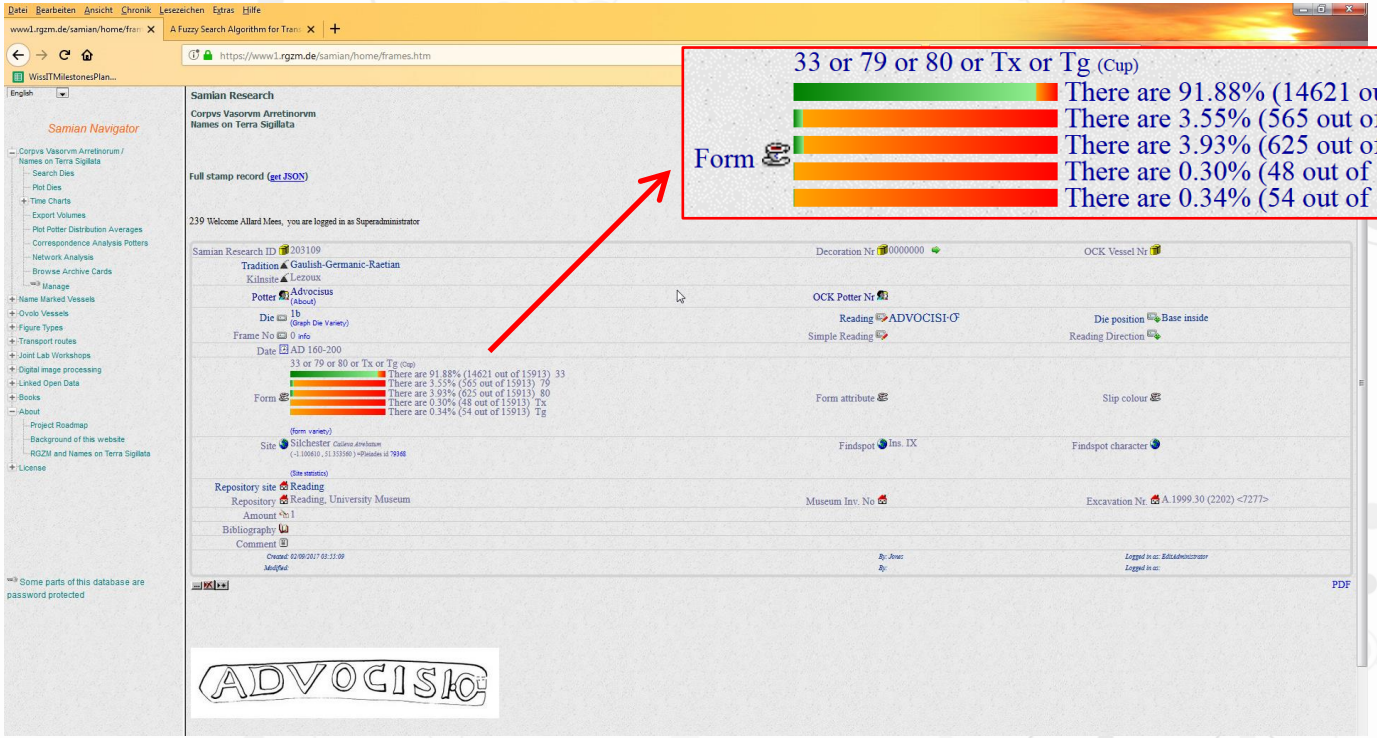
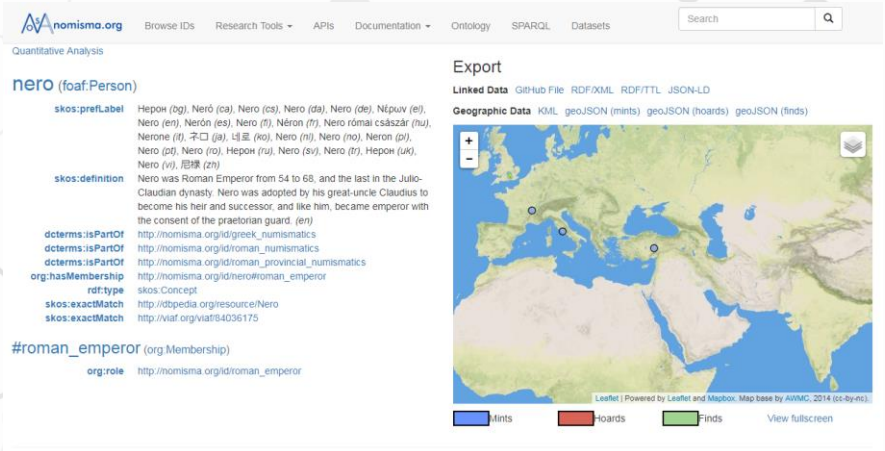


figure based on http://rgzm.de/samian

Example from the NAVIS III database: Describing the emperor on a coin using the LOD cloud and SKOS relations: *"Is it Nero?"*

Photo: NAVIS III, A



http://nomisma.org/id/nero

skos:exactMatch
=100% Nero

Example from NAVIS III: Describing the propulsion of a ship on a coin using the deliberately diffuse SKOS ontology.

Photo: NAVIS III, B

Vagueness: there are depictions with both variants.

available variants: sailed, rowed, paddled, towed, punted
possible variants: sailed, rowed

Photo: NAVIS II, B

skos:relatedMatch
≥50% sailed
≥50% rowed

Example from the NAVIS II database: Describing the sailing gear on a relief using the deliberately diffuse SKOS ontology.

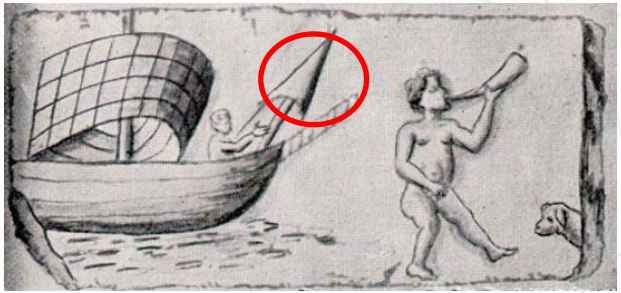


Photo: NAVIS II, C

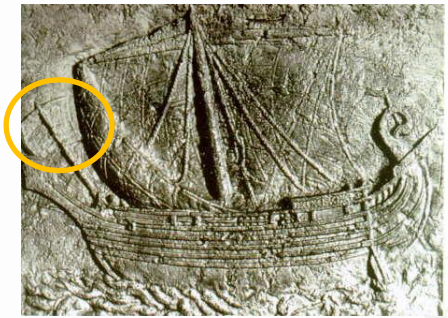


Photo: NAVIS II, D

Vagueness: is it…
- a triangular **lateen sail** used as a 'fore sail'? (=very unlikely)
- actually a squared **fore sail** that is being hoisted? (=likely)

Very important because of the different functions of the sails.

skos:relatedMatch
≥1% lateen sail
≥99% fore sail

Example from NAVIS II. Describing the ship function on a relief using the deliberately diffuse SKOS ontology.

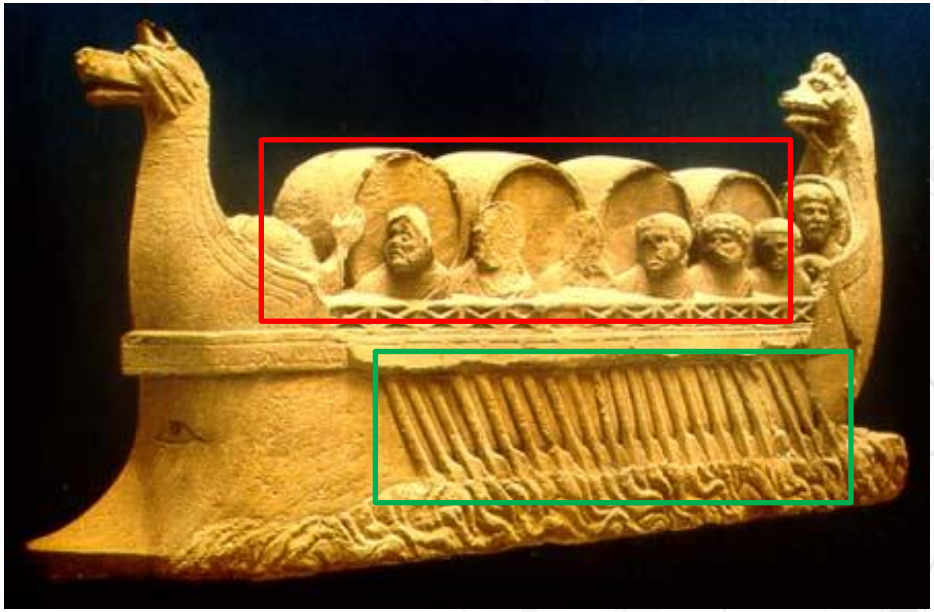


Photo: NAVIS II, E

Vagueness: the ship could be used for warfare *or* trade.

available variants: transport vessel, military vessel, working boat, fishing vessel
possible variants: transport vessel *or* military vessel?

skos:relatedMatch
≥40% transport vessel
≥60% military vessel

Trying to link a triangular ‘lateen sail’ from NAVIS II into the Linked Data Cloud, reveals that each repository has…

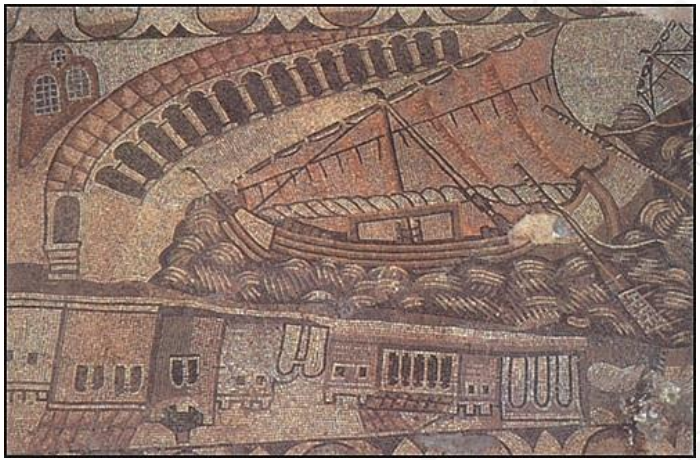


Photo: NAVIS II, A

…completely different “hidden assumptions” in its hierarchies related to their specific scientific domain.
*(Getty ‘**Art and Architecture**’ Thesaurus…)*

Are the usually SKOS based relations able to solve the challenge to model also the degree of doubt? Are different methods required? Which „hidden assumptions“ are implied?

Which hidden hierarchical assumptions are implied by linking a ‘lateen sail’ to Getty AAT or EH Maritime Craft Types?

Photo: NAVIS II, A

Getty Art and Architecture Thesaurus

Objects Facet
Furnishings and Equipment (hierarchy name)
Tools and Equipment (hierarchy name)
equipment
<equipment by process>
power producing equipment
sails (equipment)

English Heritage Maritime Craft Types

MARITIME CRAFT
WARSHIP
ESCORT VESSEL
CORVETTE SAIL

skos:relatedMatch
≥p% sails (equipment)
≥q% CORVETTE SAIL

Academic Meta Tool (AMT) provides web based functions for modelling doubts as Linked Open Data including reasoning.

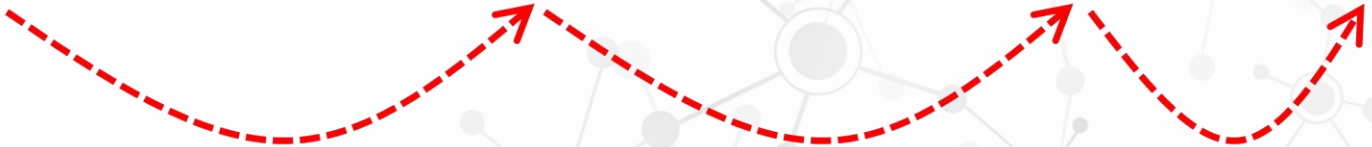
Academic Meta Tool

http://academic-meta-tool.xyz

created by
i3mainz and RGZM

with ideas from
Martin Unold M.Sc. & Florian Thiery M.Sc.

The idea behind it: map depictions to defined concepts and aligning them to authoritative thesauri to obtain additional information.



has depiction
≥p%

matches with
≥q%

is broader than
≥r%

https://github.com/RGZM/amt-caa2018

Using the AMT ontology: concepts and roles are defined and provide the fundament for an LOD export comprising doubts.

RGZM

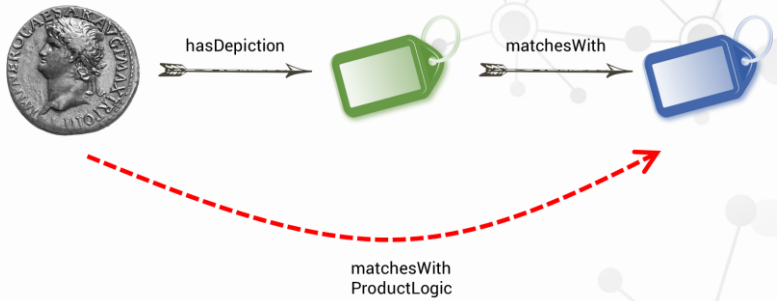
Leibniz-Gemeinschaft

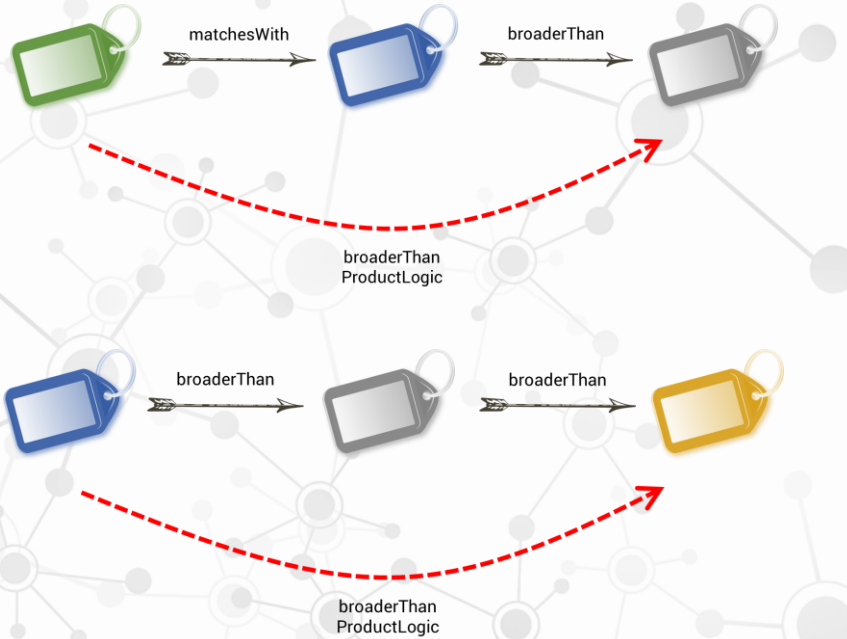
mainzed

concepts	roles	weight
 **Object** (NAVIS I/II)	 *has depiction* 	≥p%
 **Keyword** (Meta-Index)	 *matches with* 	≥q%
 **Concept** (English Heritage)	 *is broader than* 	≥1% percentage increases in top level direction

https://github.com/RGZM/amt-caa2018

Reasoning rules with various defined axioms and logics generates deeper insights based on the knowledge graph.





https://github.com/RGZM/amt-caa2018

Such rules can be formulated as RDF using specific ontologies, based on the prototypical '*Academic Meta Tool Ontology*'.

AMT ontology

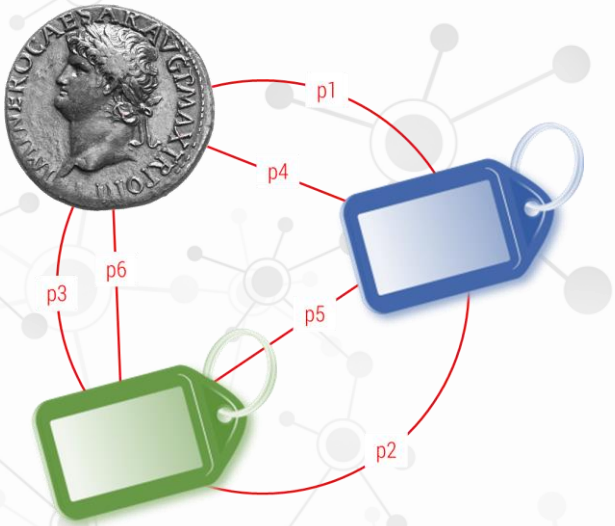
Model

```
@prefix amt: <http://academic-meta-tool.xyz/vocab#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
amt:RoleChainAxiom rdfs:subClassOf amt:InferenceAxiom .
amt:RoleChainAxiom amt:antecedent1 amt:Role .
amt:RoleChainAxiom amt:antecedent2 amt:Role .
amt:RoleChainAxiom amt:consequent amt:Role .
amt:RoleChainAxiom amt:logic amt:Logic .
```

Example

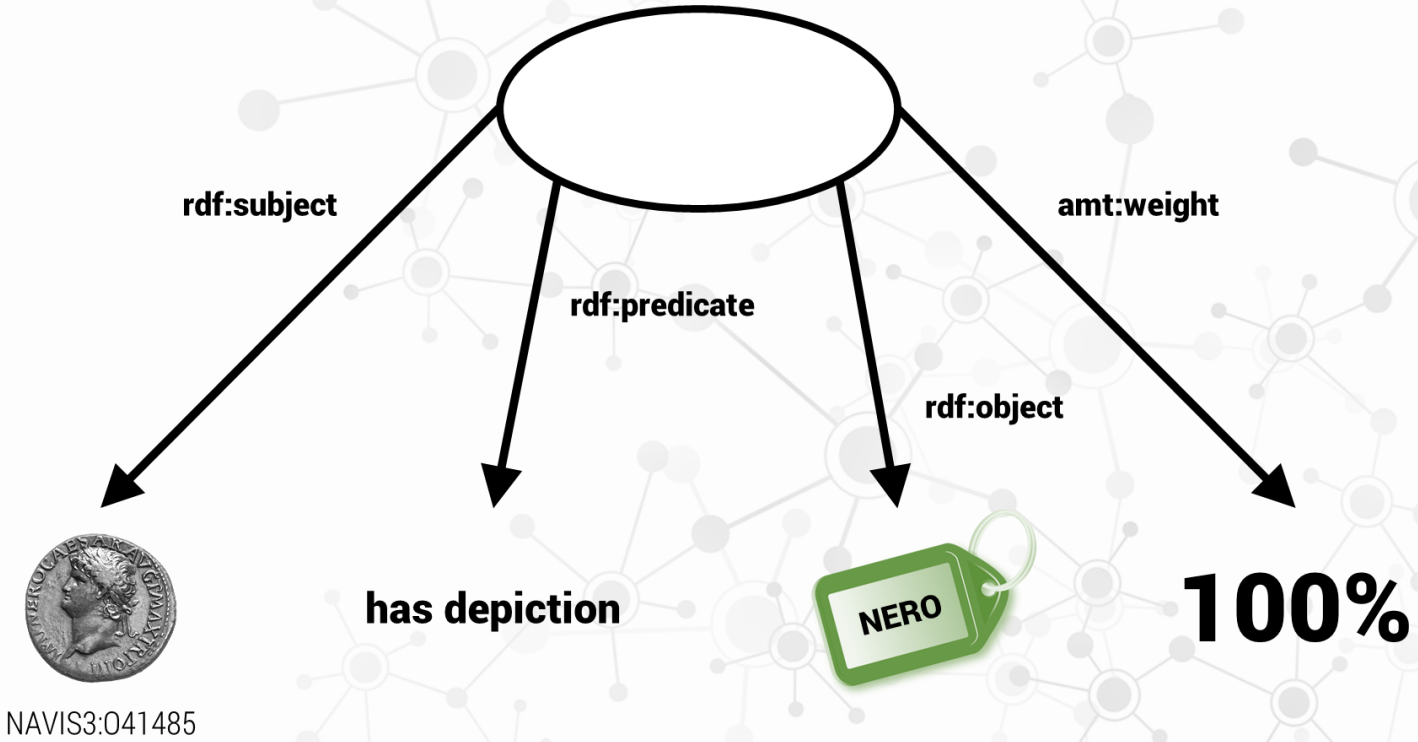
```
@prefix amt: <http://academic-meta-tool.xyz/vocab#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix example: <http://example.com/> .
example:RCA rdf:type amt:RoleChainAxiom .
example:RCA amt:antecedent1 example:connectedWith .
example:RCA amt:antecedent2 example:connectedWith .
example:RCA amt:consequent example:connectedWith .
example:RCA amt:logic amt:ProductLogic .
```

http://academic-meta-tool.xyz/ontology/

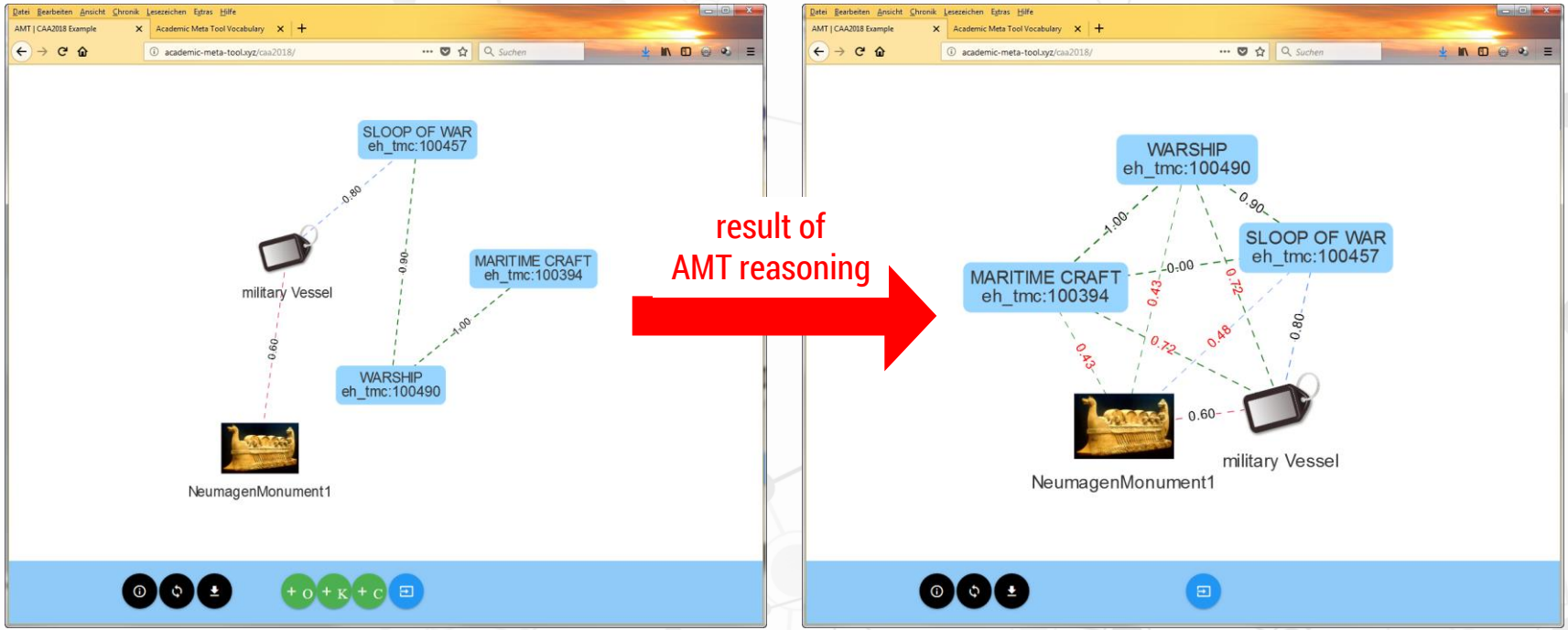


specific ontology

The defined relations between objects, keywords and concepts can be exported as Linked Open Data ‘quatruples’.

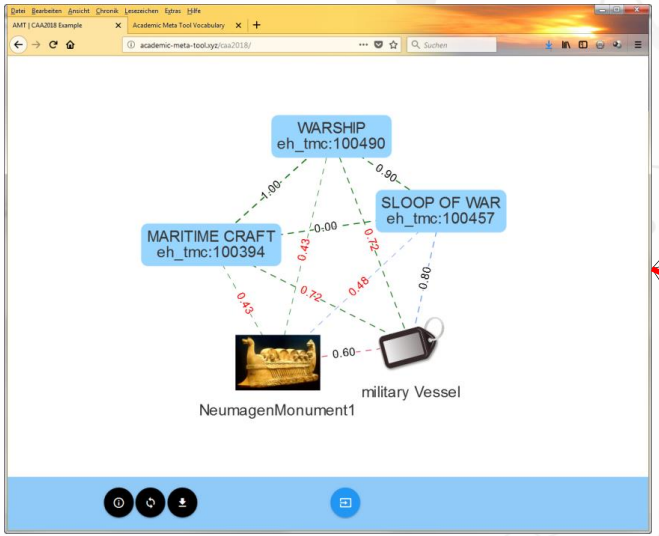


Examples of AMT reasoning, e.g. for the question *"Is it a military vessel or not?"*, can be visualised on the web.



http://academic-meta-tool.xyz/caa2018/

The resulting knowledge graphs are downloadable in different formats (e.g. RDF, JSON, CSV or Cypher) for further usage.

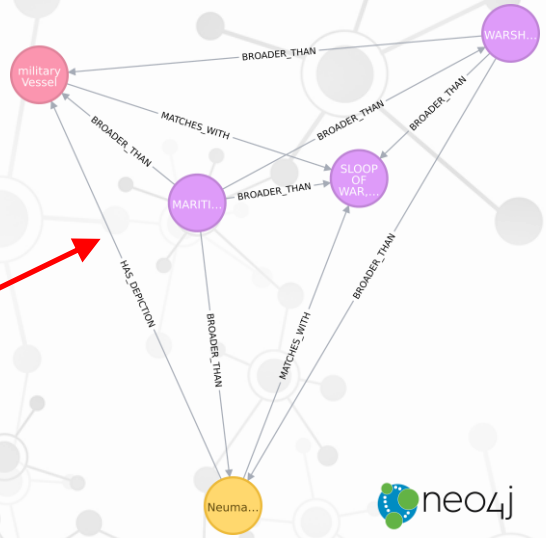


JSON

CSV

CYPHER

RDF



Modelling doubts in archaeological research using an ontology like AMT may help to tame the ambiguities in Linked Open Data.

Leibniz-Gemeinschaft

Römisch-Germanisches Zentralmuseum
Leibniz-Forschungsinstitut für Archäologie
R G Z M

Taming Ambiguity

Dealing with doubts in archaeological datasets using LOD

thiery@rgzm.de & mees@rgzm.de

online references

- Github Repository
  - https://github.com/RGZM/amt-caa2018/
- Live Demo
  - http://academic-meta-tool.xyz/caa2018
- Academic Meta Tool
  - http://www.academic-meta-tool.xyz
  - https://github.com/AcademicMetaTool/amt
  - https://github.com/almende/vis
- Online-Databases
  - http://rgzm.de/navis
  - http://rgzm.de/navis2
  - http://rgzm.de/navis3
  - http://rgzm.de/samian

publications

- Tolle/Wigg-Wolf 2014, Uncertainty Handling for Ancient Coinage
  - Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology CAA 2014 – 21st Century Archaeology/F. Giligny, F. Djindjian, L. Costa, Po. Moscati, S. Robert (eds.)
- Polak 2000, Vechten
  - M. Polak, South Gaulish terra sigillata with potters' stamps from Vechten. RCRF Acta Supplementum 9 (Nijmegen 2000).
- Webster 1996
  - P. V. Webster, Roman samian pottery in Britain. C.B.A. Practical Handbook in Archaeology 13 (York 1996).

online resources

- Samian Research
  - https://www1.rgzm.de/Samian/Queries/Cat29FullOutput.cfm?SerialNumber=0004015&Potter=Aquitanus&DieNo=-
- NAVIS II, A
  - https://www2.rgzm.de/Navis2/Home/SingleObjectOutput.cfm?ObjectName=AydyncikMosaic
- NAVIS II, B
  - https://www2.rgzm.de/Navis2/Home/SingleObjectOutput.cfm?ObjectName=PompeiRegioVIII7Naumachia3
- NAVIS II, C
  - https://www2.rgzm.de/Navis2/Home/SingleObjectOutputDE.cfm?ObjectName=NarbonneRelief7
- NAVIS II, D
  - https://www2.rgzm.de/Navis2/Home/SingleObjectOutputDE.cfm?ObjectName=SidonSarcophag
- NAVIS II, E
  - https://www2.rgzm.de/Navis2/Home/FullDetailImageDE.cfm?ID=100&ShipDepictionCode=DE_00003001&ObjectName=NeumagenMonument1
- NAVIS III, A
  - https://www1.rgzm.de/Navis3/Queries/Fenster1AversPopup.cfm?InvNr=041485
- NAVIS III, B
  - https://www1.rgzm.de/Navis3/Queries/Fenster1ReversPopup.cfm?InvNr=041650