

Call for Papers

Computer Applications & Quantitative Methods in Archaeology Meeting 2020
14th – 17th April 2020, Oxford, United Kingdom
<https://2020.caaconference.org/>



We would like to announce and invite your contribution to our session

S21: Hic sunt dracones – Improving knowledge exchange in the Semantic Web with Linked Open and FAIR data

Standard Session, max. 20 minutes / talk

In historical maps, the phrase ‘Hic sunt dracones’ (‘here be dragons’) is used to describe areas which were unknown to the map creator [UW19]. Today the WWW gives researchers the possibility of sharing their research (data) and enables the community to participate in the scientific discourse to create previously unknown knowledge. But much of these shared data are not findable or accessible, thus resulting in modern ‘unknown data dragons’. Often these ‘data dragons’ lack connections to other datasets, i.e. they are not interoperable and in some cases even lack usefulness or usability. To overcome these shortcomings, a set of techniques, standards and recommendations can be used: Semantic Web and Linked Open Data, the FAIR principles and LOUD data.

Tim Berners-Lee introduced the concept of ‘Semantic Web’, wherein he suggested using the ideas of Open Data, semantically described resources and links, as well as usable (machine readable) interfaces and applications for creating a ‘Giant Global Graph’. In 2016 the FAIR principles were introduced [MW16]: research data and their metadata have to be Findable, Accessible, Interoperable, and Reusable. Linked Data is an essential part of the FAIR principles: “The Semantic Web isn’t just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. [TBL06].” Publishing research data as HTTP URIs with RDF content containing links to other resources makes data FAIR.

On top of that, these data should be ‘open’ for access, re-use and universal participation [ODH19]. A five star rating system of openness [MH12] was introduced to rate Linked Data, i.e. “Linked Open Data (LOD) is Linked Data which is released under an open licence. [TBL06].” Furthermore, LOD have to be ‘usable’ for scientists and programmers to take advantage of all the LOD power. Following the LOUD principles [RS18] will make LOD even more FAIR.

Merging all these principles to create FAIR and LOUD research data results in the ‘Sphere 7 Data Model’ [FT19], which enables a wide array of digital humanities and archaeological (web-)applications using LOUD and FAIR data.

The Linked Data Cloud already offers research data repositories for certain archaeological and humanities domains. Popular examples of FAIR LOUD providers are: Nomisma.org [EG18], Kerameikos.org [GS15], Pelagios [ISBS14], OpenContext [EC07], Portable Antiquities Scheme [EH18], ARIADNE [AN17] and there are more to come, e.g. NAVIS [TM18b], ARS3D [TKR19] and ARIADNEplus [AP19].

The development of more and more repositories poses challenges in handling the complex facets of data quality and completeness. This is especially valid for archaeological data, which are based on a complicated network of concepts from different knowledge domains. Moreover, it is necessary to include means of conveying knowledge about uncertainty in the data models to produce and publish transparent FAIR and LOUD data that can also describe specific stratigraphies or the (archaeological) context of objects. In order to be able to connect different data resources, exchange standards also have to be developed, published, and applied.

To enable non-experts to engage with FAIR, and especially LOUD data, small tools – ‘minions’ – were created for different purposes, such as modelling a relative chronology (Alligator [DS18]), modelling and reasoning on vague edges in graph data (Academic Meta Tool [TM18]), creating annotated texts and images (Recogito [SBIS17]), and creating controlled vocabularies (Labeling System [TE16]). Furthermore, Wikidata [EGKMV14] not only offers community-driven data, but also provides a vast set of tools for using and interacting with it.

The goal of our session is to bring together experts on LOD and FAIR data, as well as anybody interested in learning about FAIR and LOUD data-driven publishing, applications and research projects based on this kind of data. We would like to discuss ideas for FAIR and LOUD data models as a basis for reproducible research and exchange in the Semantic Web.

This session is intended as a starting point for a CAA SIG on ‘Semantics and LOUD in Archaeology’. The core aim of this SIG is to use the CAA’s SIG format to raise awareness for Linked Data in archaeology by creating a friendly and open platform to discuss the role of LOUD and FAIR Data in archaeology, and to enable the CAA community to learn about LOD basics. If you wish to join the SIG, feel free to contact us to be an active part of the discussion [SIG19] and help us to navigate archaeology away from the data dragons.

The success of the sessions on data quality in Linked Data at CAA 2017 and 2018 has raised awareness of the many challenges related to FAIR and LOUD data, and encourages pursuing the debate.

For this session we invite contributions that address part or all of the following issues:

- application of semantic web technologies, such as ontologies or RDF, to the archaeological domain
- modelling archaeological artefacts as FAIR and LOUD data
- modelling archaeological context, including the specificity of stratigraphy, uncertainty, and vagueness as FAIR and LOUD data
- proposals for FAIR and LOUD data exchange standards
- development of research tools producing or using FAIR and LOUD data
- identifying sources and dangers of incorrect or ambiguous LOD
- identifying duplicates across different LOD sources
- keeping track of the provenance of data as a means of solving errors and identifying their source
- setting up methodologies and tools in order to label or assess datasets based on their quality
- dealing with ambiguities resulting from multiple links in the LOD cloud
- computer vision or machine learning applications built upon controlled, semantic data

We encourage presenters to derive the problems from real-world datasets and to formulate proposals for solutions, preferably demonstrating (prototypes of) realised data-driven web applications. Since we target a broad and diverse audience because of the thematic relevance, the challenges described should also be integrated into their archaeological context (excavation, museum, archive, etc.).

Florian Thiery¹, Martina Trognitz², Ethan Gruber³

¹ *Römisch- Germanisches Zentralmuseum*

² *Austrian Centre for Digital Humanities at Austrian Academy of Sciences*

³ *American Numismatic Society*

Deadline for the submission is October the 31st 2019.

For any questions regarding the session please contact the session chairs:
Florian Thiery <thiery@rgzm.de>, Martina Trognitz <Martina.Trognitz@oeaw.ac.at> OR Ethan Gruber <ewg4xuva@gmail.com>.