

# Hic sunt dracones - Improving knowledge exchange in the Semantic Web with Linked Open and FAIR data

Florian Thiery<sup>1</sup>, Martina Trognitz<sup>2</sup>, Ethan Gruber<sup>3</sup>

<sup>1</sup> Römisch-Germanisches Zentralmuseum, <sup>2</sup> Austrian Centre for Digital Humanities  
at Austrian Academy of Sciences, <sup>3</sup> American Numismatic Society

<sup>1</sup> thiery@rgzm.de, <sup>2</sup> Martina.Trognitz@oeaw.ac.at, <sup>3</sup> gruber@numismatics.org

In historical maps, the phrase ‘Hic sunt dracones’ (engl. here be dragons) is used to describe areas which were unknown to the map creator [1]. 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 this 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’, where 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 [2]: Research data and its 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. [3]. 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 [4]. A five star rating system of openness [5] was introduced to rate Linked Data, i. e. Linked Open Data (LOD) is Linked Data which is released under an open licence. [3]. Furthermore, LOD have to be ‘usable’ for scientists and programmers to take advantage of all the LOD power. Following the LOUD principles[6] will make LOD even more FAIR.

Merging all these principles to create FAIR and LOUD research data results in the ‘Sphere 7 Data Model’ [7], 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 [8], Kerameikos.org [9], Pelagios [10], OpenContext [11], Portable Antiquities Scheme [12], ARIADNE [13] and there are more to come, e.g. NAVIS [14], ARS3D [15] and ARIADNEplus [16].

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 in engaging with FAIR, and especially LOUD data, small tools - ‘minions’ - were created for different purposes, such as modelling a relative chronology (Alligator [17]), modelling and reasoning on vague edges in graph data (Academic Meta Tool [18]), creating annotated texts and images (Recogito [19]), and creating controlled vocabularies (Labeling System [20]). Furthermore, Wikidata [21] 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 CAAs 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 [22] 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 the archaeological context (excavation, museum, archive, etc.).

## References

- [1] U. Wuttke, “Here be dragons: Open Access to Research Data in the Humanities,” Apr. 2019. [Online]. Available: <https://ulrikewuttke.wordpress.com/2019/04/09/open-data-humanities/>
- [2] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. t Hoen, R. Hoof, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, “The FAIR Guiding Principles for scientific data management and stewardship,” *Scientific Data*, vol. 3, p. 160018, Mar. 2016. [Online]. Available: <https://doi.org/10.1038/sdata.2016.18>
- [3] T. Berners-Lee, “Linked Data - Design Issues,” Jul. 2006. [Online]. Available: <https://www.w3.org/DesignIssues/LinkedData.html>
- [4] O. D. Handbook, “What is Open Data?” 2019. [Online]. Available: <http://opendatahandbook.org/guide/en/what-is-open-data/>
- [5] M. Hausenblas and J. G. Boram Kim, “5-star Open Data,” Aug. 2012. [Online]. Available: <http://5stardata.info/en/>
- [6] R. Sanderson, “LOUD: Linked Open Usable Data,” May 2019. [Online]. Available: <https://linked.art/loud/>
- [7] F. Thiery, “Sphere 7 Data: LOUD and FAIR Data for the Research Community,” Apr. 2019. [Online]. Available: <https://zenodo.org/record/2643469>
- [8] E. Gruber, “Linked Open Data for Numismatic Library, Archive and Museum Integration,” in *CAA2016. Oceans of Data. Proceedings of the 44th Conference on Computer Applications and Quantitative Methods in Archaeology*, M. Matsumoto and E. Uleberg, Eds. Oxford: Archaeopress, 2018, pp. 55–62, oCLC: 907629230.
- [9] E. Gruber and T. J. Smith, “Linked Open Greek Pottery,” in *CAA2014. 21st Century Archaeology. Concepts, methods and tools. Proceedings of the*

*42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology.*, F. Giligny, L. Costa, P. Moscati, and S. Robert, Eds. Oxford: Archaeopress, 2015, pp. 205–214, oCLC: 907629230.

- [10] L. Isaksen, R. Simon, E. T. Barker, and P. de Soto Caamates, “Pelagios and the Emerging Graph of Ancient World Data,” in *Proceedings of the 2014 ACM Conference on Web Science*, ser. WebSci ’14. New York, NY, USA: ACM, 2014, pp. 197–201, event-place: Bloomington, Indiana, USA. [Online]. Available: <http://doi.acm.org/10.1145/2615569.2615693>
- [11] E. Kansa, “Publishing Primary Data on the World Wide Web: Opencontext.org and an Open Future for the Past.” *Technical Briefs in Historical Archaeology*, vol. 2, pp. 1–11, 2007. [Online]. Available: <http://www.sha.org/publications/technical-briefs/volume02/article.01.pdf>
- [12] E. Harper, “Toys and the Portable Antiquities Scheme: A Source for Exploring Later Medieval Childhood in England and Wales,” *Childhood in the Past*, vol. 11, no. 2, pp. 85–99, Jul. 2018. [Online]. Available: <https://www.tandfonline.com/doi/full/10.1080/17585716.2017.1348647>
- [13] Consiglio Nazionale Delle Ricerche ISTI, N. Aloia, C. Binding, S. Cuy, M. Doerr, B. Fanini, A. Felicetti, J. Fihn, D. Gavrilis, G. Geser, H. Hollander, C. Meghini, F. Niccolucci, F. Nurra, C. Papatheodorou, J. Richards, P. Ronzino, R. Scopigno, M. Theodoridou, D. Tudhope, A. Vlachidis, and H. Wright, “Enabling European Archaeological Research: The ARIADNE E-Infrastructure,” *Internet Archaeology*, no. 43, 2017. [Online]. Available: <http://intarch.ac.uk/journal/issue43/11/>
- [14] F. Thiery and A. Mees, “Taming Ambiguity - Dealing With Doubts In Archaeological Datasets Using Lod,” Mar. 2018. [Online]. Available: <https://zenodo.org/record/1200111>
- [15] F. Thiery, A. Karmacharya, and L. Rokohl, “ARS3d - Documenting facts and interpretations of African Red Slip Ware,” Apr. 2019. [Online]. Available: <https://zenodo.org/record/2648210>
- [16] ARIADNEplus, “ARIADNEplus: Advanced Research Infrastructure for Archaeological Data Networking in Europe,” 2019. [Online]. Available: [https://ariadne-infrastructure.eu/wp-content/uploads/2019/02/ariadne-plus\\_A4.researcher\\_web.pdf](https://ariadne-infrastructure.eu/wp-content/uploads/2019/02/ariadne-plus_A4.researcher_web.pdf)
- [17] D. Seidensticker, F. Thiery, A. Mees, and C. Schmid, “Rdf Based Modeling Of Relative And Absolute Chronological Data: Examples From The Central African Rainforest And Roman Periodisation,” Sep. 2018. [Online]. Available: <https://zenodo.org/record/1410516>
- [18] F. Thiery and A. Mees, “Taming the chronology of South Gaulish Samian found at Hadrian’s Wall and the German Limes using Linked Open Data,” Oct. 2018. [Online]. Available: <https://zenodo.org/record/1469298>
- [19] R. Simon, E. Barker, L. Isaksen, and P. De Soto Caamates, “Linked Data Annotation Without the Pointy Brackets: Introducing Recogito 2,” *Journal of Map & Geography Libraries*, vol. 13, no. 1, pp. 111–132, Jan.

2017. [Online]. Available: <https://www.tandfonline.com/doi/full/10.1080/15420353.2017.1307303>

- [20] F. Thiery and T. Engel, “The Labeling System: The Labelling System: A Bottom-up Approach for Enriched Vocabularies in the Humanities,” in *CAA2015. Keep the Revolution Going. Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology.*, S. Campana, R. Scopigno, G. Carpentiero, and M. Cirillo, Eds. Oxford: Archaeopress, 2016, pp. 259–268, oCLC: 907629230.
- [21] F. Erxleben, M. Gnther, M. Krtzsch, J. Mendez, and D. Vrandei, “Introducing Wikidata to the Linked Data Web,” in *The Semantic Web ISWC 2014*, P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandei, P. Groth, N. Noy, K. Janowicz, and C. Goble, Eds. Cham: Springer International Publishing, 2014, vol. 8796, pp. 50–65. [Online]. Available: [http://link.springer.com/10.1007/978-3-319-11964-9\\_4](http://link.springer.com/10.1007/978-3-319-11964-9_4)
- [22] F. Thiery, M. Trogitz, E. Gruber, K. Tolle, and D. Wigg-Wolf, “CAA SIG on ”Semantics and LOUD in Archaeology”,” Jul. 2019. [Online]. Available: <https://zenodo.org/record/3337979>