

Hic sunt dracones!

The modern unknown Data Dragons

Florian Thiery ^a, Martina Trognitz ^b, Ethan Gruber ^c, David Wigg-Wolf ^d

^a ORCID: 0000-0002-3246-3531

Römisch-Germanisches Zentralmuseum, Mainz, Germany

^b ORCID: 0000-0003-0485-6861

Austrian Centre for Digital Humanities at Austrian Academy of Sciences, Austria

^c ORCID: 0000-0002-4691-9747

American Numismatic Society, USA

^d ORCID: 0000-0002-8604-544X

Römisch-Germanische Kommission des Deutschen Archäologischen Instituts, Germany

Abstract

The world of digital data is full of unknown data, with are not FAIR (Findable, Accessible, Interoperable and Reusable). This modern species of dragons on a (digital) map of the internet as sign for unknown and dangerous areas, the unknown digital data dragons, have to be resolved by digital techniques. We could use the Semantic Web and all of its ideas for this challenge. This paper focuses on data dragons in the humanities domain, especially in archaeology and proposes ideas to overcome the data dragons to achieve FAIR LOUD in archaeology!

Key words: Linked Data; Linked Open Data; Five Star Data; FAIR Data; Semantic Web; Data Dragon.

1 Hic sunt dracones digitales!

In historical maps (cf. Fig. 1), the phrase **Hic sunt dracones** (engl. *here be dragons*) is used to describe areas which were unknown to the map creator [1].



Figure 1. Nature and Science, <https://t1p.de/wmtz>

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** (cf. Fig. 2) 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 Dragon

Figure 2. Data Dragon

2 Shout it out, FAIR and LOUD!

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 meta-data have to be **F**indable, **A**ccessible, **I**nteroperable and **R**eusable. Linked Data (cf. Fig. 3) 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!

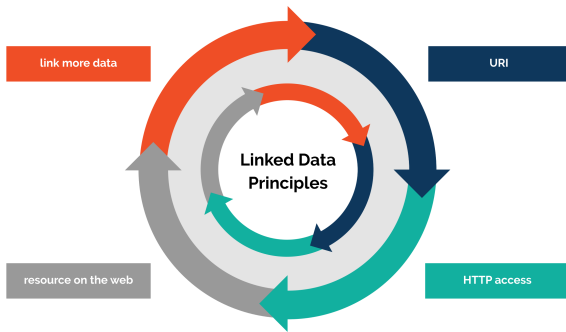


Figure 3. Linked Data, Florian Thiery [CC BY 4.0]

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, cf. Fig. 4) 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 LOD (cf. Fig. 5) principles[6] will make LOD even more FAIR.

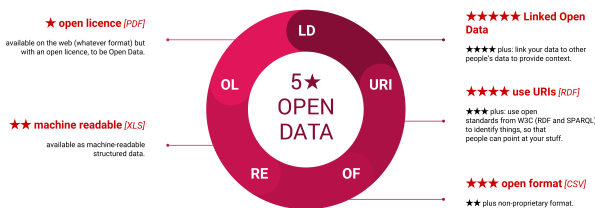


Figure 4. 5 Star Open Data, Florian Thiery [CC BY 4.0]

Merging all these principles to create FAIR and LOD research data results in the *Sphere 7 Data Model* (cf. Fig. 6) [7], which enables a wide array of digital humanities and archaeological (web-)applications using LOD and FAIR data.

The *Linked Data Cloud* already offers research data repositories for certain archaeological and humanities

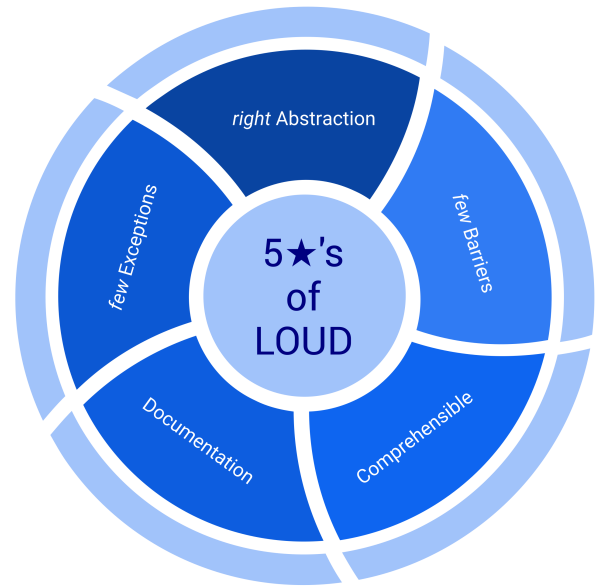


Figure 5. 5 Star LOD, Florian Thiery [CC BY 4.0]

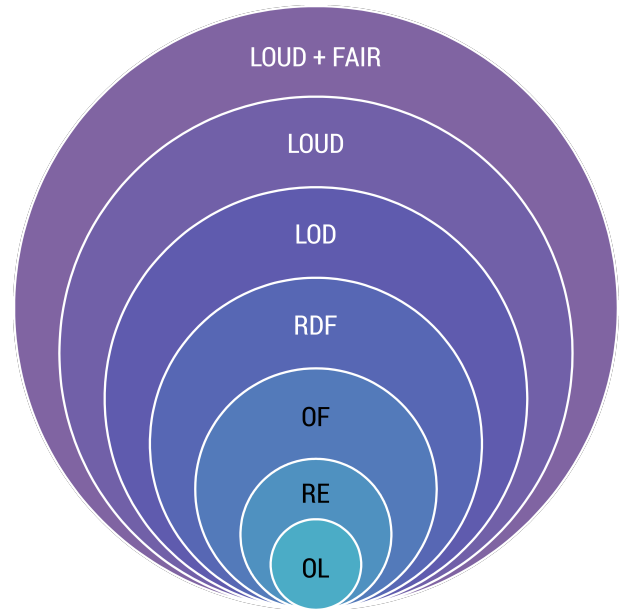


Figure 6. Seven Sphere Data, Florian Thiery [CC BY 4.0]

domains. Popular examples of FAIR LOD 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.

3 Digital Dragons at CAA

The goal of the CAA SIG on **Semantics and LOUD in Archaeology** 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. 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.

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. Hooft, 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: <http://www.nature.com/articles/sdata201618>
- [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 Caameres, "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

- [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 Caameres, “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
- [22] F. Thiery, M. Trognitz, 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>