Archaeology 4.0:
Archaeology in the Third Era of Computing

Florian Thiery [a]

[a] *Research Software Engineer, Mainz, Germany*

Abstract

We are in the middle of the digital transformation era, which, as a digital revolution, affects society in economic as well as in scientific terms. This digital revolution as a third part of industrial revolution is based on digital technologies that have already become established in individual stages in various areas of life. Terms such as Industry 4.0, Work 4.0 and Web 4.0 are lived reality in our digital society.

Archeology has also undergone a digital transformation from an analogous science to an Archaeology 4.0. Starting with an *analogue era*, in which research data was kept in books, followed by a *digital era* in which digitization progresses and data are published on the WWW, followed by a *semantic era*, in which semantic modelling and publication of Linked Data is reality, we end up in a *knowledge era*, in which the analysis and the creation of new knowledge through machines will be reality.

Key words: Archaeology 4.0; Linked Data; Semantic Reasoning; Artificial Intelligence.

1 Archaeological Evolution

Today, in 2019 we are living in the centre of the Digital Transformation era. The digital transformation affects all kind of 'humanoid' and 'machine' communities: individuals and society, state, companies and science (research/teaching). In archaeology, researcher develop and use new digital technologies. This results in positive effects on science and archaeology itself. The state promotes (e.g. funding by the Federal Ministry of Education and Research - BMBF) and uses these digital technologies and serves as a regulator. Also the society is using more and more digital technologies. Companies are developing and using new digital technologies as well.

Fig. 1. The Digital and the Ocean of Traditional Things, Florian Thiery [CC BY 4.0]

Thus, science and archaeology should seek collaboration with companies to positively influence the state and society through digital technologies. But does the digital wave overtake us? (Fig. 1) We should surf the wave! On a first wave of digitalisation we have already recorded, saved, transmitted and processed machine-readable data via internet and cloud computing technologies. Furthermore, on the second wave of digitalisation we have to analyse, enhance and use these research data in active ways as machine-interpretable data using artificial intelligence and machine learning.

While surfing the waves, we actively experience the evolution of archaeology from analogue data to Knowledge Graph Computing. That is what I will call Archaeology 4.0 (Fig. 2).

Fig. 2. Archaelogy 4.0, Florian Thiery [CC BY 4.0]

5 April 2019

Archaeologal Sciences fulfilled a digital transformation from an 'analogous science' to modern digital science of humanities. Starting from an *analogue era* (Archaeology 1.0), in which research data was kept in books and monographs, across the *digital era* (Archaeology 2.0) in which digitisation progresses and data are published on the WWW, to a *semantic era* (Archaeology 3.0), where semantic modelling and publication of Linked Data prevail, we end up in a *knowledge era* (Archaeology 3.0), in which the analysis and the creation of new knowledge through a machine results.

1.1 Archaeology 1.0

Fig. 3. Archaelogy 1.0 Symbol, pixabay [Pixabay License]

I will call the era of non digital things ('analogue') the *analogue era*. Research, and its underlying data, was published in printed books and monographs. Data access was only possible via active visiting libraries or complicated interlibrary loan. Research Data was not 'OPEN' at this time, because it was a privilege of an individual.

1.2 Archaeology 2.0

Fig. 4. Archaelogy 2.0 Symbol, pixabay [Pixabay License]

I will call the time using and creating digital things the *digital era*. With the invention of the Internet in ≈1969 by Vinton Gray Cerf ('father of the internet') and others and the invention of the World Wide Web in 1989 by computer scientist Sir Tim Berners-Lee and the so born Web 1.0, digital methods in archaeology come into play. (Archaeological) Research Data is from now on not exclusive any more, it is accessible for everyone via searchable online databases or PDF documents. The Web 2.0 including its so called 'Social Media' and the collaborative potential like Wikipedia and Google Docs enlarge the odds of shared archaeological research around the globe. The new possibilities of digitalisation methods of photographs and 3D documentation, using image analysis and 3D techniques, expand the range even more. Moreover, Open Data initiatives and common data format standards like the OGC web services, the Text Encoding Initiative (TEI) help to make research data interoperable, transparent and reproducible. This enables researcher using computer applications and quantitive

methods to generate there own analyses via using digitised data.
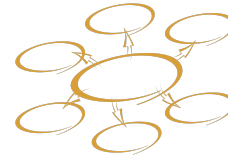
1.3 Archaeology 3.0

Fig. 5. Archaelogy 3.0 Symbol, pixabay [Pixabay License]

I will call the period of modelling data in graph and triple structures the *semantic era*. In this era not only the structure matters - like in relational databases - the semantic meaning of things and their relationships are the backbone of research and scientific data. The idea of a Web 3.0, the so called Semantic Web comes into play. By the invention of a simple Resource Description Framework - RDF - everything stared. The RDF model is a data model based on a directed graph, were data are statements about resources. This statements are modelled as triples. A triple is modelled like a short sentence: Subject - Predicate - Object. Subject and Object are modelled as nodes, the predicate is the name of the directed edge between Subject and Object. Subjects and Predicates are usually URIs, Objects can be also a Literal, which are not unique. The base of the Semantic Web are ontologies. This big conceptual models including several axioms and rules to enable reasoning are usually modelled in the Web Ontology Language - OWL. The most common ontology in archaeology is the CIDOC CRM and its extensions.
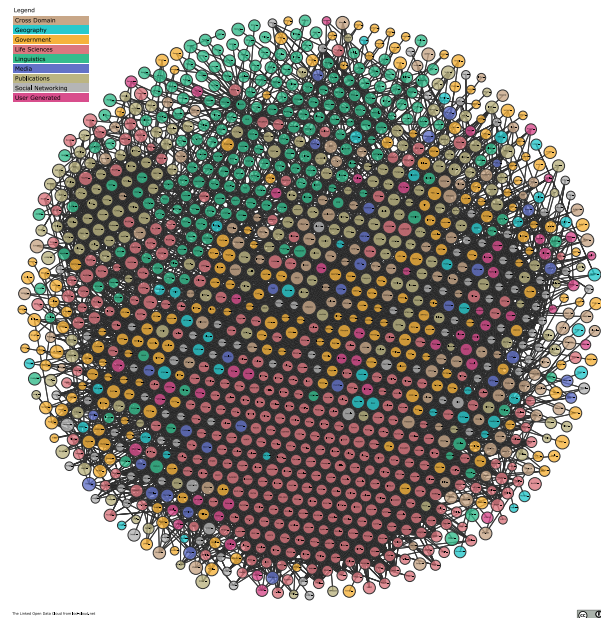

Fig. 6. Linked Open Data Cloud 2019-03-29, lod-cloud.net [CC BY 4.0]

Publishing data using the RDF data model ends up in using Linked Data. Tim Berners-Lee created four Linked Data rules [1] for creating a Linked Data Cloud (Fig. 6): (1) use URIs as names for things, (2) use HTTP URIs so that people can look up those names, (3) when someone looks up a URI, provide useful information, using the standards like RDF and SPARQL, (4) include links to other URIs, so that they can discover more things.
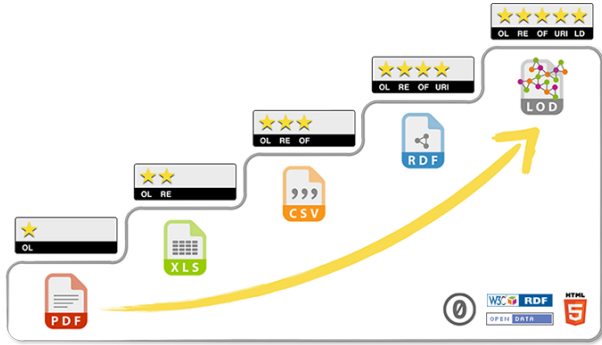

Fig. 7. 5 Star Open Data, 5stardata.info [CC0]

Mentioned in Archaeology 2.0, data is only usable if data will be shared and is open. Berners-Lee also proposed the Five Star Data Model (Fig. 7) for open data where Linked Data will be the best option, the so called Linked Open Data [2].

To make Linked Data and Linked Open Data usable for computer scientists and researchers the term of LOUD - Linked Open Usable Data [6] by Rob Sanderson was established into the Web 3.0 community. LOUD can also be devided into five design principles / stars, the ABCDE model: (1) the right *Abstraction* for the audience, (2) few *Barriers* to entry, (3) *comprehensible* by introspection, (4) *documentation* with working examples and (5) few *Exceptions*, many consistent patterns.

The Web offers a lot of Linked Open Data in different domains (space, time, persons, keywords, ...) to use them for several purposes: LOD as Infrastructure, LOD as Research Data, LOD as Research Tool and LOD as Cloud. Resources can be found in several projects in the WWW like GeoNames, Pleiades, Pelagios, ChronOntology, Getty, Nomisma, ... Examples in archaeology are described in a famous PhD by Leif Isaksen "Archaeology and the Semantic Web" [4] and in my own master thesis [7].

There are also some nice little projects build up in Mainz, Germany, for LOD as Infrastructure purposes (Labeling System, RGZM Meta-Index), for LOD as Research Data purposes (NAVIS 2.0, ARS3D), for LOD as Reseach Tool (Alligator, Academic Meta Tool) to establish a small piece of LOD as Cloud.

As the big aim of the archaeological Linked Open Data community, creating a Archaeological Linked Data

Fig. 8. Archaeological Linked Data Cloud, Florian Thiery [CC BY 4.0]

Cloud comes into play (Fig. 8). This cloud, including all archaeological object data, research data, meta data, para data with all of its hidden assumptions as well as interpretations in an objective, comprehensible, transparent and interoperable machine readable semantic way will open up new possibilities.

1.4 Archaeology 4.0

Fig. 9. Archaelogy 4.0 Symbol, pixabay [Pixabay License]

I will call the period of analysing data a create new knowledge using machines the *knowledge era*. In this era the players are Artificial Intelligence, Data Mining, Machine Learning and Knowledge Graphs.

But for using this methods in archaeology I propose a problem hypothesis: The techniques of archaeology 4.0 often need big amounts of training data, (*BigData*) which are not (yet) available in archaeology.

Using the words of my master thesis supervisor Prof. Kai-Christian Bruhn: As CAPTCHAs (completely automated public Turing test to tell computers and humans apart) turned from training machines learning text for digitalisation projects (Google Books, reCAPTCHA) to training machines learning understanding images (like recognising street signs) the archaeological community missed the boat to place images of archaeological artefacts into that huge survey.

One possible solution could be using techniques of Archaeology 3.0 to create an interlinked semantically correctly described archaeological data cloud as a general base for training data.
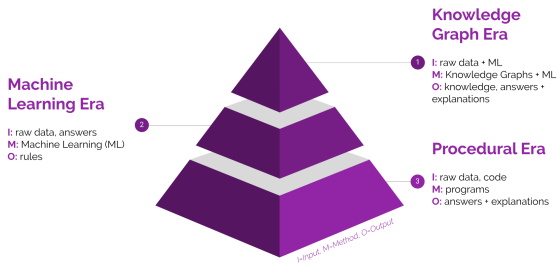
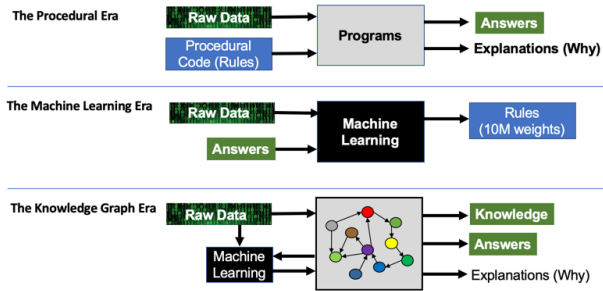Fig. 10. Era of Computing, Florian Thiery [CC BY 4.0]

Fig. 11. Era of Computing, Dan McCreary [5]

If this process is finished, archaeology can start into the third era of computing [3] using the Archaeological Linked Data Cloud as underlying Knowledge graph. Right now we stuck in the *procedural era*

> where we write step-by-step procedures to find answers in our raw data[5]

. In the *machine learning era era*

> where data and answers are fed in and the outcome is a black box model [...] but without explanation of why decisions were made[5]

archaeological research can not be done, because we do not know why the machine has created the knowledge. In the *knowledge graph era*

> where machine learning continuously reads raw data, combines this with existing knowledge and produces new knowledge, answers and explanations[5]

Archaeology 4.0 will be reality. This story have to be continued!

Acknowledgements

References

[1] Tim Berners-Lee. Linked Data - Design Issues, July 2006.

[2] Michael Hausenblas and James G. Boram Kim. 5-star Open Data, August 2015.

[3] Anthony J. G. Hey and Gyuri Ppay. *The computing universe: a journey through a revolution.* Cambridge University Press, New York, NY, USA, 2015.

[4] Leif Isaksen. *Archaeology and the Semantic Web.* phd, University of Southampton, December 2011.

[5] Dan McCreary. Knowledge Graphs: The Third Era of Computing, March 2019.

[6] Rob Sanderson. LOUD: Linked Open Usable Data, May 2019.

[7] Florian Thiery. *Semantic Web und Linked Data: Generierung von Interoperabilitaet in archaeologischen Fachdaten am Beispiel roemischer Toepferstempel.* master thesis, Fachhochschule Mainz, December 2013.