



## Analysis and Categorisation of Research Software in the Digital Humanities

Lukas Harzenetter<sup>2</sup>, Johanna Barzen<sup>2</sup>, Uwe Breitenbücher<sup>2</sup>, Frank Leymann<sup>2</sup>,  
Brigitte Mathiak<sup>1</sup>, Claes Neufeind<sup>1</sup>, Philip Schildkamp<sup>1</sup>

<sup>1</sup> University of Cologne, 50923 Cologne, Germany

<sup>2</sup> University of Stuttgart, 70569 Stuttgart, Germany

{c.neufeind,philip.schildkamp,bmathiak}@uni-koeln.de

{lukas.harzenetter,uwe.breitenbuecher,johanna.barzen,frank.leymann}@iaas.uni-stuttgart.de

---

### BIBTEX:

```
@inproceedings{Harzenetter2020_SoftwareCategorisationInDH,  
  Author    = {Lukas Harzenetter and Johanna Barzen Uwe Breitenb{"u}cher  
              and Frank Leymann and Brigitte Mathiak and Philip Schildkamp  
              and Claes Neufeind},  
  Title     = {{Analysis and Categorisation of Research Software  
              in the Digital Humanities}},  
  Booktitle = {DH2020},  
  Publisher = {ADHO},  
  Month    = jul,  
  Year     = 2020,  
}
```



# Analysis and Categorisation of Research Software in the Digital Humanities

Lukas Harzenetter<sup>1</sup>, Johanna Barzen<sup>1</sup>, Uwe Breitenbücher<sup>1</sup>, Frank Leymann<sup>1</sup>,  
Brigitte Mathiak<sup>2</sup>, Claes Neufeind<sup>2</sup>, Philip Schildkamp<sup>2</sup>

<sup>1</sup> Institute of Architecture of Application Systems  
University of Stuttgart  
{harzenetter,breitenbuecher,barzen,leymann}  
@iaas.uni-stuttgart.de

<sup>2</sup> Data Center for the Humanities  
University of Cologne  
{bmathiak,c.neufeind,philip.schildkamp}  
@uni-koeln.de

## Abstract

In the Digital Humanities (DH) research software is a major output of the scientific process besides primary research data. However, long-term sustainability and preservation of living systems, in contrast to long-term archiving (LTA) of research data, is still a non-trivial, often institutionally segmented enterprise (Smithies et al., 2019). To counter this, the SustainLife project, running at the Institute of Architecture of Application Systems (IAAS, University of Stuttgart) and the Data Center for the Humanities (DCH, University of Cologne), uses the Topology and Orchestration Specification for Cloud Applications (TOSCA) standard (OASIS, 2013 and 2019), and the open source ecosystem OpenTOSCA (Breitenbücher et al. 2016). To realize our vision of a sustainable long-term DH research software archive, we are building a repository of software components modeled in TOSCA, i.e., Node Types, and use them to create application blueprints, i.e., Topology Templates in TOSCA, describing the structure of an application.

To identify the most used software component types within the DH, we approach the vast field of DH research software from multiple perspectives (Neufeind et al., 2018 and 2019): (1) We investigate multiple use cases in depth and extract the components used in the respective applications, starting from the operating system (OS) level up to the user interface (UI) layer. These components are then modelled as reusable Node Types in TOSCA which can be utilized to describe the respective application in a TOSCA Topology Template. (2) Qualitative case studies aim at extracting all employed components and technologies of selected applications and (3) quantitative surveys, targeted at DH research software experts (i.e. researchers), are performed to collect community practices and demands. The results from both the case studies and the surveys are then used to identify key components and derive application stacks, i.e., Node Types and Topology Templates.

We intend our poster about the “Categorisation of Research Software in the Digital Humanities” to spawn discourse about the perspectives of DH research software in contrast to research data (Sahle and Kronenwett, 2013) and how this vast landscape can be indexed in a way that researchers may find the most appropriate set of component types or a suitable application blueprint, when searching for their own software. In this context, we investigate the use of categorial keywords which in combination with technological identifiers, e.g., programming languages and database types, can be used to provide a means for categorizing research software in the DH. Therefore, our poster will present selected DH research software as a basis for discussion. Hereby, we introduce component types and application stacks from our repository, e.g., an Apache web server or the application stack of a DH research software respectively. Moreover, a set of technological identifiers and categorial keywords, such as digital editions, virtual research environments, and interactive visualisations (Bingert et al., 2016), associated with these projects will be presented.

## Acknowledgements

This Poster is partially funded by the DFG-LIS project “SustainLife” (GEPRIS 379522012).

## References

- **Bingert, S. and Blumtritt, J. and Buddenbohm, S. and Engelhardt, C. and Kronenwett, S. and Kurzawe D.** (2016). Anwendungskonservierung und die Nachhaltigkeit von Forschungsanwendungen. In: Forschungsdaten in den Geisteswissenschaften (FORGE 2016), pp. 14-16.
- **Breitenbücher, U. and Endres, C. and Képes, K. and Kopp, O. and Leymann, F. and Wagner, S. and Wettinger, J. and Zimmermann, M.** (2016). The OpenTOSCA Ecosystem. Concept & Tools. In: European Space project on Smart Systems, Big Data, Future Internet - Towards Serving the Grand Societal Challenges - Volume 1: EPS Rome 2016. SciTePress, pp. 112-130.
- **Neuefeind, C. and Schildkamp, P. and Mathiak, B. and Harzenetter, L. and Barzen, J. and Breitenbücher, U. and Leymann, F.** (2018). Technologienutzung im Kontext Digitaler Editionen. Eine Landschaftsvermessung. In: Book of abstracts of the 6th annual conference of the Digital Humanities im deutschsprachigen Raum (DHd 2019), pp. 219-222.
- **Neuefeind, C. and Schildkamp, P. and Mathiak, B. and Marčić, A. and Hentschel, F. and Harzenetter, L. and Breitenbücher, U. and Barzen, J. and Leymann, F.** (2019). Sustaining the Musical Competitions Database. A TOSCA-based Approach to Application Preservation in the Digital Humanities. In: Book of Abstracts of the 29th Digital Humanities Conference (DH 2019), <https://dev.clariah.nl/files/dh2019/boa/0574.html> (retrieved: 2019-09-10).
- **OASIS** (2013). Topology and Orchestration Specification for Cloud Applications Version 1.0, <http://docs.oasis-open.org/tosca/TOSCA/v1.0/TOSCA-v1.0.html> (retrieved: 2019-09-10).
- **OASIS** (2019). TOSCA Simple Profile in YAML Version 1.2, <http://docs.oasis-open.org/tosca/TOSCA-Simple-Profile-YAML/v1.2/TOSCA-Simple-Profile-YAML-v1.2.html> (retrieved: 2019-09-10).
- **Sahle, P. and Kronenwett, S.** (2013). Jenseits der Daten. Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner Data Center for the Humanities. In: LIBREAS. Library Ideas #23, pp. 76-96.
- **Smithies, J. and Sichani, A. M. and Westling, C. and Mellen, P. and Ciula, A.** (2019). Managing 100 Digital Humanities Projects. Digital Scholarship & Archiving in King's Digital Lab. In: Digital Humanities Quarterly, <http://www.digitalhumanities.org/dhq/vol/13/1/000411/000411.html> (retrieved: 2019-09-10)