Workshop on Modelling and Maintaining Research Applications in TOSCA

Philip Schildkamp[1], Lukas Harzenetter[2], Brigitte Mathiak[1], Claes Neuefeind[1],
Johanna Barzen[2], Uwe Breitenbücher[2], Frank Leymann[2]

[1] University of Cologne, 50923 Cologne, Germany
[2] University of Stuttgart, 70569 Stuttgart, Germany

{c.neuefeind,philip.schildkamp,bmathiak}@uni-koeln.de
{lukas.harzenetter,uwe.breitenbuecher,johanna.barzen,frank.leymann}@iaas.uni-stuttgart.de

BibTeX:

```
@inproceedings{Schildkamp2020_DHWorkshop,
    Author     = {Philip Schildkamp and Lukas Harzenetter and Brigitte Mathiak
                   and Neuefeind and Johanna Barzen Uwe Breitenb{\"u}cher
                   and Frank Leymann},
    Title      = {{Workshop on Modelling and Maintaining Research Applications
                   in TOSCA}},
    Booktitle = {DH2020},
    Publisher = {ADHO},
    Month     = jul,
    Year      = 2020,
}
```

Universität Stuttgart
Germany

Workshop on Modelling and Maintaining Research Applications in TOSCA

Philip Schildkamp[1], Lukas Harzenetter[2], Uwe Breitenbücher[2], Frank Leymann[2], Brigitte Mathiak[1], and Claes Neuefeind[1]

[1] Data Center for the Humanities
University of Cologne
{philip.schildkamp,bmathiak,c.neuefeind}
@uni-koeln.de

[2] Institute of Architecture of Application Systems
University of Stuttgart
{harzenetter,breitenbuecher,leymann}
@iaas.uni-stuttgart.de

Abstract

The project "SustainLife – Sustaining Living Digital Systems in the Humanities" that is currently running at the Institute of Architecture of Application Systems (IAAS, University of Stuttgart) and the Data Center for the Humanities (DCH, University of Cologne) deals with the conservation of research applications in the field of Digital Humanities (DH). By employing the TOSCA standard (Topology and Orchestration Specification for Cloud Applications) to fully automate the deployment of DH applications and to keep them available in the long term, we try to tackle the problem of software obsolescence in the field of DH. To interactively demonstrate our approach to the international DH community, we would like to give a workshop on the topic "Modelling and Maintaining Research Applications in TOSCA" in the run-up to the DH 2020 conference. Therein we will show how to model (DH) software systems with TOSCA and share experiences and best practices on how to work with the OpenTOSCA ecosystem, an open-source implementation of the TOSCA standard.

The Problem

The establishment of the DH as an independent scientific research area as well as the increasing usage of digital methods in the research process require adjustments to common result assurance practices. For example, the long-term archiving (LTA) of primary research data uses well-established practices such as employing standardized data formats and forwarding data to permanent repositories. However, the fact that digital artifacts generated in DH-oriented research do not only consist of primary data but also contain research software is mostly disregarded (Sahle and Kronenwett, 2013). Moreover, the variety of DH research outcomes includes so-called "living systems" in which the software to present, access or analyze the data represents an essential part of the actual research output (Bingert et al., 2016). In contrast to classical research results such as monographs or encyclopedias, living systems cannot be served long-term without maintenance as their instantiation, supervision, and permanent provisioning represent major technical, organizational, and financial challenges. Furthermore, the heterogeneity of the research software generated in the DH requires a highly flexible preservation strategy, i.e., a suitable technology that ensures standardization, reusability, and archiving of as many digital artifacts as possible (Barzen et al., 2018). In addition to the aforementioned challenges, i.e., heterogeneity, underfunding, and obsolescence of digital artifacts, scientific practice requires long-term interoperability and traceability of all research outcomes. With regard to digital systems, these requirements are (1) constant accessibility, (2) the possibility of error-free operation, and (3) the ability to reconstruct any stage of development of a research application at any time without major structural difficulties.

Our Approach

The TOSCA standard (OASIS, 2013 and 2019) allows software systems to be modelled, provisioned, and deployed in a standardized and provider-independent manner. Thus, it is suitable for long-term archiving and operation of research applications produced within the field of DH (Neuefeind et al., 2018 and 2019). Following the TOSCA standard, applications are modelled in "Topology Templates" by describing their components and their relations amongst each other: Components are represented

as "Node Templates", while relations are modeled as "Relationship Templates". Moreover, the semantics of a Node Template or Relationship Template are dictated by reusable types, i.e., "Node Types" and "Relationship Types" respectively. For example, a Python web application can be modelled as a Node Template that is an instance of the "Python Application" Node Type. To express that the Python Application accesses a MySQL database, a second Node Template that is of type "MySQL Database" can be added to the Topology Template. Then, the connection between both components can be described by a Relationship Template that is an instance of the Relationship Type "connectsTo". Additionally, to specify that both components are running on an Ubuntu virtual machine (VM), a Node Template of type "Ubuntu VM" can be added, while Relationship Templates of type "hostedOn" between the Python Application Node Template and the VM Node Template, as well as between the MySQL Database and the VM describe their respective hosting relations.

Hereby, TOSCA's type system enables the modelling of reusable component types, e.g., the "Python Application" Node Type, which can be reused in multiple Topology Templates describing different applications. Therefore, synergic effects emerge as existing Node Types can be reused in other Topology Templates, easing the modelling of new applications. In addition, the open-source TOSCA implementation OpenTOSCA (Breitenbücher et al., 2016) offers the possibility to graphically model applications using the TOSCA editor "Winery" (Kopp et al., 2013) which further simplifies the creation of new applications by providing drag-and-drop modeling capabilities.

Workshop Curriculum

During our four hour workshop, we will (1) give an overview to different solutions for long-term preservation of living systems and (2) describe the modeling language TOSCA. Based on these theoretical units, practical tasks will introduce (3) the modelling of an existing application using TOSCA and (4) how applications can be deployed using the OpenTOSCA ecosystem. Thus, by combining the theoretical foundations and the practical application of TOSCA, the participants will be able to model (research) software systems according to the standard and provision and deploy applications using the OpenTOSCA ecosystem.

The practical tasks are structured as follows: (1) Identify the components of an application and (2) describe them and their relations among each other in an TOSCA-based application topology, i.e., in a Topology Template. By fragmenting an application into its components and mapping them to TOSCA Node Types, the Topology Template describing the application can then be modelled using the OpenTOSCA ecosystem. Afterwards (3), the modelled TOSCA application will be deployed by the OpenTOSCA runtime. Moreover, by sharing our experiences and best practices in using OpenTOSCA with the community, we will introduce concepts such as "software stacks" in a practical way.

Target Group

The workshop is primarily designed for data center employees, libraries and other institutions focusing on infrastructures for long-term archiving and operation of heterogeneous software systems. Previous experience in dealing with Linux and writing shell scripts as well as with software stacks and service orchestration are helpful but not necessary for a successful participation. To provide a productive context for communicating the described content and to enable individual consultation and support, we designed the workshop for about 20 participants but limit it to a maximum of 30 participants.

Technical Prerequisites

For a successful participation in the workshop, it is necessary that each participant brings his/her own laptop. Although a shared instance of the OpenTOSCA ecosystem will be provided, it is desirable that all participants set up an OpenTOSCA instance on their work equipment prior to the workshop in order to perform modelling and deployment tasks on their own devices. Therefore, registered participants will be provided with all necessary information about system requirements and how to setup OpenTOSCA prior to the workshop. Furthermore, relevant documentation, publications, and manuals

will be provided both in advance and in the context of the workshop. In addition, a stable internet connection as well as a sufficient number of power outlets for all electronic devices are indispensable.

About the Instructors

Philip Schildkamp has been researching since 2015 and teaching since 2017 at the University of Cologne. He studied sociology, psychology, and Digital Humanities information processing. The main topics of his employment are technical infrastructure measures in the field of (Digital) Humanities and the orchestration of distributed software systems. Since March 2018, Philip is part of the DFG-LIS project "SustainLife" at the DCH.

Lukas Harzenetter is a research associate at the Institute of Architecture of Application Systems (IAAS) at the University of Stuttgart, Germany. He received his Master of Science degree from the University of Stuttgart in Software Engineering in 2018. His research interests are in the field of cloud deployment and management models focusing on the development and change of such models over time. Lukas is part of the SustainLife project which is working on sustainable application deployments in the domain of digital humanities.

Frank Leymann is a full professor of computer science and director of the Institute of Architecture of Application Systems (IAAS) at the University of Stuttgart, Germany. His research interests include service-oriented architectures and associated middleware, workflow- and business process management, cloud computing and associated systems management aspects, and patterns. Frank is co-author of more than 400 peer-reviewed papers, about 70 patents, and several industry standards. He is elected member of the Academy of Europe.

Brigitte Mathiak is chairwoman of the Data Center for the Humanities and is particularly interested in data management and text mining. The idea for the "SustainLife" project arose after she had experienced again and again how living systems have to be abandoned or neglected. She is Junior Professor for Digital Humanities at the University of Cologne and Senior Scientist at the Leibniz Institute for the Social Sciences (GESIS).

Claes Neuefeind is a postdoc at the Cologne Center for eHumanities (CCeH) at the University of Cologne. He worked with Philip Schildkamp and Lukas Harzenetter on the DFG-LIS project "SustainLife" until October 2019 and changed for a position that is responsible for coordinating the Digital Humanities of the North Rhine-Westphalian Academy of Sciences and the Arts office.

Acknowledgements

References

- **Barzen, J. and Blumtritt, J. and Breitenbücher, U. and Kronenwett, S. and Leymann, F. and Mathiak, B. and Neuefeind, C.** (2018). SustainLife – Erhalt lebender, digitaler Systeme für die Geisteswissenschaften. In: Book of Abstracts of the 5th annual Conference of the Digital Humanities im deutschsprachigen Raum (DHd 2018), pp. 471-474.

- **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, pp. 112-130.

- **Kopp, O. and Binz, T. and Breitenbücher, U. and Leymann, F.** (2013). Winery – A Modeling Tool for TOSCA-based Cloud Applications. In: Proceedings of the 11th International Conference on Service-Oriented Computing (ICSOC 2013), pp. 700-704.

- **Neuefeind, C. and Harzenetter, L. and Schildkamp, P. and Breitenbücher, U. and Mathiak, B. and Barzen, J. and Leymann, F.** (2018). The SustainLife Project – Living Systems in Digital Humanities. In: Proceedings of the 12th Advanced Summer School on Service-Oriented Computing (SummerSoC 2018) (IBM Research Report RC25681), pp. 101-112.

- **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.