

Modellierung und Verwaltung von DH- Anwendungen in TOSCA

Schildkamp, Philip

philip.schildkamp@uni-koeln.de
Universität zu Köln, Data Center for the Humanities
(DCH)

Neuefeind, Claes

c.neuefeind@uni-koeln.de
Universität zu Köln, Data Center for the Humanities
(DCH)

Mathiak, Brigitte

bmathiak@uni-koeln.de
Universität zu Köln, Data Center for the Humanities
(DCH)

Harzenetter, Lukas

lukas.harzenetter@iaas.uni-stuttgart.de
Universität Stuttgart, Institut für Architektur von
Anwendungssystemen (IAAS)

Breitenbücher, Uwe

uwe.breitenbuecher@iaas.uni-stuttgart.de
Universität Stuttgart, Institut für Architektur von
Anwendungssystemen (IAAS)

Leymann, Frank

frank.leymann@iaas.uni-stuttgart.de
Universität Stuttgart, Institut für Architektur von
Anwendungssystemen (IAAS)

Abstract

Das aktuell vom Institut für Architektur von Anwendungssystemen (IAAS) der Universität Stuttgart und vom Data Center for the Humanities (DCH) der Universität zu Köln bearbeitete Projekt *SustainLife – Erhalt lebender, digitaler Systeme für die Geisteswissenschaften* befasst sich mit der Konservierung von Forschungssoftware im Bereich der Digital Humanities (DH). Dabei wird der Topology Orchestration Specification for Cloud Applications (TOSCA) Standard verwendet, um das Deployment von DH-Anwendungen vollständig zu automatisieren und diese langfristig verfügbar zu halten. Um der DH Community unseren Ansatz interaktiv zu demonstrieren, möchten wir im Vorfeld der DHd 2020 einen Workshop zur *Modellierung*

und Verwaltung von DH-Anwendungen in TOSCA durchführen. Dabei sollen Kernkompetenzen bezüglich der Modellierung von Softwaresystemen mit TOSCA sowie Erfahrungen und Best Practices im Umgang mit OpenTOSCA, einer open-source Implementierung des TOSCA Standards, vermittelt werden.

Problemstellung

Die zunehmende Etablierung der DH als ein eigenes Forschungsfeld sowie der damit einhergehend vermehrte Einsatz von digitalen Methoden im Forschungsprozess erfordern daran angepasste Mittel der Ergebnissicherung. Zur Langzeitarchivierung von Forschungsprimärdaten gibt es bereits etablierte Strategien, bspw. die Nutzung standardisierter Datenformate und die Übermittlung relevanter Daten an einschlägige Repositorien. Weitestgehend unberücksichtigt bleibt dabei, dass viele der in DH-orientierten Forschungsprozessen erzeugten digitalen Artefakte nicht in Form von Primärdaten, sondern in Form von Forschungssoftware vorliegen. Die Vielfalt der in den DH erzeugten Software beinhaltet auch sog. *lebende Systeme* (Sahle, Patrick / Kronenwett, Sabine: 2013), deren Laufzeitumgebung unerlässliche Daten enthält und die somit nicht statisch abbildbar sind. Da solche lebenden Systeme im Gegensatz zu klassischen Erkenntnisträgern wie bspw. Monographien oder Lexika nicht ohne kontinuierliche Wartung auskommen, stellen Erhalt, Betreuung und dauerhafte Bereitstellung große technische, organisatorische und finanzielle Hürden dar. Weiterhin erfordert die Heterogenität der in den DH erzeugten Forschungssoftware eine höchst flexible Methodologie bzw. Technologie, die Standardisierung, Nachnutzbarkeit und Archivierung von möglichst vielen digitalen Artefakten gewährleisten kann (Barzen, Johanna et al.: 2018). Neben den genannten Herausforderungen (Heterogenität, Unterfinanzierung und Überalterung digitaler Artefakte) fordert die wissenschaftliche Praxis dauerhafte Interoperabilität und Nachvollziehbarkeit aller Erkenntnisträger. Bezogen auf digitale Systeme sind diese Forderungen (1) konstante Zugänglichkeit, (2) die Möglichkeit eines fehlerfreien Betriebs und (3) die Möglichkeit jeden Entwicklungsstand einer Forschungsanwendung zu jedem Zeitpunkt und ohne große strukturelle Hürden nachzuvollziehen bzw. wiederherzustellen zu können.

Lösungsansatz

Da der TOSCA-Standard (OASIS: 2013, 2019) es erlaubt, Anwendungen standardisiert und anbieterunabhängig zu modellieren, zu provisionieren und zu deployen, eignet er sich auch zum langfristigen Archivieren und Betreiben von in den DH erzeugter Forschungssoftware (vgl. Neuefeind et al.: 2018). Hierbei werden Anwendungen mithilfe von wiederverwendbaren Komponententypen,

sog. *Node Types*, modelliert. Um Abhängigkeiten zwischen diesen unterschiedlichen Komponenten einer Anwendung darzustellen, werden verschiedene Beziehungstypen, sog. *Relationship Types*, verwendet. So kann bspw. eine einfache PHP Webanwendung, die auf eine Datenbank zugreift, als eine Instanz des Node Types *PHP Anwendung* modelliert werden, welche sich zu einer Instanz des *MySQL Datenbank* Node Types verbindet. Die Verbindung der beiden Komponenten zueinander wird durch den Relationship Type *connectsTo* dargestellt. Zusätzlich kann bspw. angegeben werden, dass beide Komponenten auf einer Ubuntu virtuellen Maschine (VM) installiert werden müssen, welche wiederum eine Instanz des Node Types *Ubuntu VM* ist (vgl. Neufeind et al.: 2019).

Solch eine Beschreibung der Anwendungskomponenten und deren Beziehungen untereinander wird *Anwendungstopologie* genannt. Weiterhin ermöglicht TOSCA durch sein Typensystem die Modellierung von wiederverwendbaren Komponententypen, sodass bspw. der *PHP Anwendung* Node Type in mehreren unterschiedlichen Anwendungen verwendet werden kann. Dadurch kommen Synergieeffekte zum Tragen, da bereits existierende Node Types von anderen Modellen wiederverwendet werden können, womit die Modellierung neuer Anwendungen deutlich schneller und einfacher wird. Darüber hinaus bietet die open-source TOSCA Implementierung OpenTOSCA die Möglichkeit Anwendungen grafisch per drag-and-drop zu modellieren, wodurch die Modellierung nochmals vereinfacht wird.

Inhalte des Workshops

Neben einem Einblick in verschiedene Lösungsansätze wird den Teilnehmenden zunächst der konzeptuelle Rahmen des TOSCA-Standards vermittelt. Auf Basis dieser theoretischen Vorarbeit sollen praxisorientierte Arbeitseinheiten in den Umgang mit OpenTOSCA einführen. Durch die Vermittlung sowohl der theoretischen Grundlagen als auch der praktischen Anwendung des TOSCA-Standards werden die Teilnehmenden in die Lage versetzt, (Forschungs-) Software standardkonform zu modellieren und mit Hilfe von OpenTOSCA bereitzustellen.

Die praxisorientierten Arbeitseinheiten werden wie folgt strukturiert: Ausgehend von der Identifikation aller Komponenten eines Softwaresystems soll dieses im Hinblick auf den TOSCA-Standard als Anwendungstopologie erfasst und abgebildet werden. Dabei werden auch theoretische Konzepte wie sog. *Software-Stacks* praxisnah eingebunden. Daraufhin soll die erarbeitete Anwendungstopologie mittels OpenTOSCA und dem darin enthaltenen graphischen Editor *Winery* (vgl. Kopp et al.: 2013) modelliert werden, um die modellierte Anwendung letztendlich mit OpenTOSCA automatisiert bereitzustellen. Des Weiteren werden unsere Erfahrungen und Best-Practices im Umgang mit TOSCA und der

Modellierung von Anwendungen in OpenTOSCA an die Community weitergegeben.

Zielgruppe des Workshops

Der Workshop richtet sich in erster Linie an Mitarbeiter von Datenzentren, Bibliotheken und sonstigen Institutionen mit Ausrichtung auf Infrastrukturen für Langzeitarchivierung und -betrieb heterogener lebender Systeme. Vorerfahrungen im Umgang mit Linux und mit den Themen *Shell-Scripting*, *Software-Stacks* und *Service-Orchestrierung* sind hilfreich, aber nicht notwendig zur erfolgreichen Teilnahme. Um einen produktiven Kontext zur Vermittlung der aufgezeigten Inhalte zu schaffen und individuelle Beratung und Betreuung zu ermöglichen, streben wir ein Ideal von 20 bis maximal 30 Teilnehmenden an.

Technische Vorbedingungen

Zur erfolgreichen Teilnahme am Workshop ist es notwendig, dass jeder Teilnehmer ein eigenes Arbeitsgerät mitbringt. Weiterhin ist es wünschenswert, dass alle Teilnehmer im Vorfeld des Workshops eine OpenTOSCA-Instanz auf ihren Geräten aufsetzen, um eigene Modellierungen durchzuführen und zu sichern. Zwar wird eine zentral erreichbare Instanz bereitgestellt, jedoch kann keine Garantie für den langfristigen Erhalt dieser Instanz und damit auch der dort hinterlegten Ergebnisse übernommen werden (es ist jedoch problemlos möglich diese Ergebnisse zum Abschluss des Workshops aus der zentralen Instanz zu exportieren und somit weiterhin nutzen zu können). Darüber hinaus sind eine stabile Internetverbindung sowie eine umfassende Versorgung der Teilnehmer mit Netzstrom unabdingbar.

Für den erfolgreichen Ablauf des Workshops werden alle angemeldeten Teilnehmer im Vorfeld des Workshops mit allen notwendigen Informationen zur Inbetriebnahme von OpenTOSCA ausgestattet. Weiterhin werden einschlägige Dokumentationen, Publikationen und Anleitungen sowohl vorab als auch im Kontext des Workshops bereitgestellt.

Forschungsgebiete der Referenten

Brigitte Mathiak

Brigitte Mathiak ist Vorsitzende Sprecherin des Data Center for the Humanities und insbesondere an den Themen Datenmanagement und Text Mining interessiert. Die Idee zum *SustainLife* LIS-Projekt entstand, nachdem sie immer wieder erlebt hat wie lebende Systeme aufgegeben oder vernachlässigt werden müssen. Sie ist Juniorprofessorin für Digital Humanities an der Universität zu Köln und

darüber hinaus Senior Scientist am Leibniz-Institut für die Sozialwissenschaften (GESIS).

Claes Neuefeind

Claes Neuefeind ist Postdoc am Cologne Center for eHumanities (CCeH) der Universität zu Köln. Bis Oktober 2019 bearbeitete er gemeinsam mit Philip Schildkamp das DFG-LIS-Projekt *SustainLife* für das DCH und ist seither am CCeH verantwortlich für die Geschäftsführung der Koordinierungsstelle Digital Humanities der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste.

Frank Leymann

Frank Leymann ist Professor für Informatik und Direktor des Institute of Architecture of Application Systems (IAAS) an der Universität Stuttgart. Seine Forschungsinteressen umfassen serviceorientierte Architekturen und zugehörige Middleware, Workflow- und Geschäftsprozessmanagement, Cloud Computing und damit verbundene Aspekte des Systemmanagements sowie Design Patterns. Frank ist Mitautor von mehr als 400 peer-reviewed Papers, etwa 70 Patenten und mehreren Industriestandards. Er ist ein gewähltes Mitglied der Europäischen Akademie.

Lukas Harzenetter

Lukas Harzenetter ist wissenschaftlicher Mitarbeiter am Institut für Architektur von Anwendungssystemen (IAAS) an der Universität Stuttgart. Seinen Master of Science Abschluss erhielt er von der Universität Stuttgart im Studiengang Software Engineering im Jahr 2018. Seine Forschungsinteressen liegen im Bereich Cloud-Deployment und Management. Er beschäftigt sich vor allem damit, wie sich Deploymentmodelle entwickeln. Lukas ist Teil des DFG-LIS-Projekts *SustainLife* und arbeitet an nachhaltigen Anwendungsimplementierungen im Bereich der digitalen Geisteswissenschaften.

Philip Schildkamp

Philip Schildkamp forscht seit 2015 und lehrt seit 2017 an der Universität zu Köln. Er studierte Soziologie, Psychologie und Informationsverarbeitung. Schwerpunktthemen seiner Beschäftigung sind technische Infrastrukturmaßnahmen im Bereich der (digitalen) Geisteswissenschaften und die Orchestrierung von verteilten Softwaresystemen. Seit März 2018 bearbeitet er am DCH das DFG-LIS-Projekt *SustainLife*.

Uwe Breitenbücher

Uwe Breitenbücher ist wissenschaftlicher Mitarbeiter und Postdoc am Institut für Architektur von Anwendungssystemen (IAAS) der Universität Stuttgart. Seine Forschungsvision ist die Verbesserung der Bereitstellung von Cloud-Anwendungen und des Anwendungsmanagements durch die Automatisierung der Anwendung mithilfe von Managementmustern. Uwe war Teil des CloudCycle-Projekts, in dem das OpenTOSCA Ecosystem entwickelt wurde. Seine aktuellen Forschungsinteressen umfassen cyber-physikalische Systeme, Blockchains und Microservices.

Acknowledgments

Dieser Workshop wird teilweise durch das DFG-LIS Projekt *SustainLife* (379522012) finanziert.

Bibliographie

Barzen, Johanna / Blumtritt, Jonathan / Breitenbücher, Uwe / Kronenwett, Sabine / Leymann, Frank / Mathiak, Brigitte / Neuefeind, Claes (2018): *SustainLife – Erhalt lebender, digitaler Systeme für die Geisteswissenschaften*. In: Book of Abstracts der 5. Jahrestagung der Digital Humanities im deutschsprachigen Raum (DHD 2018), S. 471-474.

Kopp, Oliver / Binz, Tobias / Breitenbücher, Uwe / Leymann, Frank (2013): *Winery – A Modeling Tool for TOSCA-based Cloud Applications*. In: Proceedings of the 11th International Conference on Service-Oriented Computing (ICSOC 2013), S. 700-704.

Neuefeind, Claes / Harzenetter, Lukas / Schildkamp, Philip / Breitenbücher, Uwe / Mathiak, Brigitte / Barzen, Johanna / Leymann, Frank (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), S. 101-112.

Neuefeind, Claes / Schildkamp, Philip / Mathiak, Brigitte / Mar#i#, Aleksander / Hentschel, Frank / Harzenetter, Lukas / Breitenbücher, Uwe / Barzen, Johanna / Leymann, Frank (2019): *Sustaining the Musical Competitions Database. A TOSCA-based Approach to Application Preservation in the Digital Humanities*. In: Book of Abstracts der 29. Digital Humanities Conference (DH 2019), <https://dev.clariah.nl/files/dh2019/boa/0574.html> (Stand: 10.09.2019).

OASIS (2013): *Topology and Orchestration Specification for Cloud Applications Version 1.0*, <http://docs.oasis-open.org/tosca/TOSCA-v1.0.html> (Stand: 10.09.2019).

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> (Stand: 10.09.2019).

Sahle, Patrick / Kronenwett, Sabine (2013): Jenseits der Daten. Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner Data Center for the Humanities. In: LIBREAS. Library Ideas 23, S. 76-96.