

# Lebende Systeme in den Digital Humanities - das Projekt SustainLife

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

<sup>1</sup>Data Center for the Humanities (DCH), Universität zu Köln

<sup>2</sup> Institut für Architektur von Anwendungssystemen (IAAS), Universität Stuttgart

## Abstract

Der Beitrag beschreibt einen Lösungsansatz für den Erhalt spezialisierter, webbasierter Forschungsanwendungen in den Geisteswissenschaften. Die Modellierung auf Grundlage des TOSCA-Standards erlaubt eine portable Beschreibung der Systeme unabhängig von konkreten Anbietern, um deren langfristige Verfügbarkeit zu ermöglichen. Anhand konkreter Usecases aus dem Bereich der Digital Humanities (DH) werden im SustainLife-Projekt Schlüsselkomponenten identifiziert und Anwendungsvorlagen erstellt, die über die OpenTosca-Umgebung für die Modellierung von DH-Anwendungen zur Verfügung gestellt werden. Die im Projekt modellierten Usecases werden zudem als Praxisbeispiele zur Verfügung gestellt.

## 1 Ausgangslage

Das exponentielle Wachstum sowie der zunehmende Gebrauch von digitalen Forschungsdaten beeinflussen den Forschungsprozess in den Geisteswissenschaften signifikant. Bisher bleibt jedoch weitgehend unberücksichtigt, dass ein Großteil der digitalen Produkte in den Geisteswissenschaften nicht nur in Forschungsprimärdaten besteht, sondern vor allem auch in Form von Forschungsanwendungen vorliegt. Präsentationssysteme, interaktive Visualisierungen, Recherche-Datenbanken, digitale Editionen und digitale Arbeitsumgebungen - um nur einige Formen zu nennen - stellen gerade im Kontext der Digital Humanities (DH) einen wesentlichen Bestandteil digitaler Ergebnissicherung dar. Die dauerhafte Erhaltung, Betreuung und Bereitstellung solcher „lebender Systeme“ [SAH13] stellt jedoch eine große technische, organisatorische und letztlich vor allem auch finanzielle Herausforderung dar.

Hier lassen sich im wesentlichen vier Defizite ausmachen: (i) Bislang fehlt ein Überblick über die Gesamtlandschaft, was notwendige Voraussetzung für die Bearbeitung des Problems ist. (ii) Bereits erste Stichproben von Usecases aus dem Bereich der DH zeigen eine große Bandbreite sowohl in Technologie als auch Methodik. Anders als im Kontext großer, betrieblicher Informationssysteme liegt in den DH demnach eine große Anzahl kleinerer, stark heterogener Softwarelösungen vor, die alle gleichermaßen dem Pro-

blem des “Software-Aging” [PAR94] unterworfen sind. (iii) Die Exposition dieser Technologien ins Internet ist systembedingt mit einem laufenden Entwicklungsaufwand verbunden, bspw. um die Systeme durch Sicherheitsupdates dauerhaft verfügbar zu halten. (iv) Mittel für eine langfristige Wartung sind jedoch meist nur in geringem Umfang vorgesehen.

## 2 SustainLife

Im Projekt *SustainLife* (IAAS Stuttgart und DCH Köln) arbeiten wir an einer Lösung der oben genannten Probleme auf Grundlage des OASIS-Standards TOSCA (*Topology and Orchestration Specification for Cloud Applications*) [OAS13]. TOSCA ermöglicht eine portable Beschreibung von IT-Systemen in Form von Modellen, um deren Provisionierung und Management zu automatisieren. Dadurch können anpassungsfähige und zukunftssichere Software-Architekturen geschaffen werden. Im Zuge des Projekts werden ausgewählte Usecases aus dem Bereich der DH in Form von TOSCA-Modellen beschrieben, um diese jederzeit automatisiert bereitstellen zu können. Ziel ist es, die darin eingesetzten Systemkomponenten in einer Komponentenbibliothek sowie in Form von TOSCA-konformen Anwendungsvorlagen für die Modellierung anderer DH-Projekte zur Verfügung zu stellen.

## 3 OpenTOSCA

Die Universität Stuttgart hat in mehreren BMWi-geförderten Projekten ein open-source Ökosystem für den TOSCA-Standard entwickelt, das im Projekt auf den spezifischen Bedarf der DH angepasst wird. Das OpenTOSCA Ökosystem beinhaltet (i) das Modellierungswerkzeug *Winery*, das die Erstellung TOSCA-basierter Anwendungsmodelle ermöglicht [KBBL13], (ii) die Laufzeitumgebung *OpenTOSCA* für die automatisierte Provisionierung und das automatisierte Management der beschriebenen Anwendungen [BBH+13], sowie (iii) das Self-Service-Portal *Vinothek* [BBKL14b], das alle im OpenTOSCA-Container installierten Anwendungen auflistet und die grafische Schnittstelle zum Benutzer bildet.

In einer ersten Publikation wurde bereits gezeigt, dass sich der TOSCA-Standard generell zur Sicherung der digitalen Nachhaltigkeit von Forschungsergebnissen eignet [BBFL17]: Der TOSCA-Standard

definiert zu Anwendungsmodellen auch ein Paketierungsformat, mithilfe dessen alle zur Bereitstellung nötigen Softwareartefakte, Modelle, Skripte, etc. in einer ZIP-Datei verpackt werden können, welche anschließend von einer TOSCA-Laufzeitumgebung vollautomatisiert entpackt und die beschriebene Anwendung bereitgestellt werden kann („Cloud Service Archive“, abgekürzt „CSAR“). Wird eine Forschungsanwendung in einem solchen CSAR pakettiert, kann dieses auch Jahre später noch von einer TOSCA-Laufzeitumgebung ausgeführt werden. Dabei stellt TOSCA diverse Erweiterungsmechanismen zur Verfügung, um beliebige Komponententypen modellieren zu können. Zudem können existierende Technologien, wie beispielsweise Container-Systeme wie Docker, nahtlos integriert werden. TOSCA stellt somit keinen Konkurrenzansatz zu existierenden Technologien dar, sondern eine Möglichkeit, diese miteinander zu kombinieren [LBWW17].

Die existierenden Arbeiten und Technologien im TOSCA-Universum leisten einen wichtigen Beitrag, um Anwendungen so zu beschreiben, dass diese prinzipiell und technisch auch Jahre später noch bereitgestellt werden können. Einige der Anforderungen bleiben jedoch offen, für die TOSCA und die Winery noch erweitert werden müssen.

Der typische Usecase ist, dass die Anwendung nur zu bestimmten Zeiten benötigt wird oder nach einer Terminierung zu einem späteren Zeitpunkt wieder lauffähig gemacht werden soll. Ein Beispiel hierfür ist die „Digitale Rätoromanische Chrestomathie“ (DRC, siehe <http://www.crestomazia.ch> sowie [NEU11]), die im SustainLife-Projekt als DH-Usecase bearbeitet wird. Teil der DRC ist ein webbasierter Korrektureditor auf Basis des Eclipse-RAP-Framework, der nach Abschluss der Korrekturarbeiten an der DRC vor allem zu Dokumentationszwecken vorgehalten werden soll. Optimalerweise wird in einem solchen Fall der identische Anwendungszustand vor der Terminierung wieder hergestellt, sodass ein An- und Abschalten der Anwendung ermöglicht wird und Ressourcen nur dann gebunden werden, wenn diese auch tatsächlich von der Anwendung genutzt werden. Eine dementsprechende Funktionalität wird gegenwärtig weder vom *OpenTOSCA*-Ökosystem, noch von anderen Cloud-Managementumgebungen für TOSCA unterstützt. Im SustainLife-Projekt werden wir daher ein Konzept zur Umsetzung dieser Funktionen entwickeln.

Lebende Systeme sind einem ständigen Wandel unterworfen. Neben einfachen Wartungsaufgaben, wie dem Einspielen eines Patches, müssen teilweise komplette Komponenten oder, bei einer Migration, die komplette Umgebung ausgetauscht werden. Dafür ist es hilfreich, mehrere Versionen von TOSCA-Modellen verwalten zu können. Dies bedeutet auch, dass Managementfunktionen nicht notwendigerweise von Anfang an bekannt sind.

Aktuell unterstützt die *Winery* sowohl Manage-

mentoperationen, die an Komponententypen gebunden sind, als auch eigenständige Managementpläne, die automatisiert ausführbare Prozessmodelle darstellen [BBK+14]. Diese in ein laufendes System einzuspielen ist bisher nicht vorgesehen und soll im Rahmen des Projekts ermöglicht werden.

## Literatur

- [SAH13] P. Sahle, S. Kronenwett. „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, 2013.
- [PAR94] D. L. Parnas. „Software Aging“. In: Proceedings of the 16th International Conference on Software Engineering (ICSE 1994). IEEE, Mai 1994, S. 279–287, 1994.
- [BBK+14] U. Breitenbücher, T. Binz, K. Képes, O. Kopp, F. Leymann, J. Wettinger. „Combining Declarative and Imperative Cloud Application Provisioning based on TOSCA“. IC2E, S. 87–96, 2014.
- [BBFL17] U. Breitenbücher, J. Barzen, M. Falkenthal, F. Leymann. „Digitale Nachhaltigkeit in den Geisteswissenschaften durch TOSCA: Nutzung eines standardbasierten Open-Source Ökosystems“. Konferenzabstracts DHd 2017: Digitale Nachhaltigkeit, S. 235–237, 2017.
- [LBWW17] F. Leymann, U. Breitenbücher, S. Wagner, J. Wettinger. „Native Cloud Applications: Why Monolithic Virtualization Is Not Their Foundation“. Cloud Computing and Services Science, Springer, S. 16–40, 2017.
- [OAS13] OASIS: Topology and Orchestration Specification for Cloud Applications Version 1.0, 2013.
- [KBBL13] O. Kopp, T. Binz, U. Breitenbücher und F. Leymann. „Winery – A Modeling Tool for TOSCA-based Cloud Applications“. In: ICSOC, 2013, S. 700–704, 2013.
- [BBH+13] T. Binz, U. Breitenbücher, F. Haupt, O. Kopp, F. Leymann, A. Nowak und S. Wagner. „OpenTOSCA - A Runtime for TOSCA-based Cloud Applications“. In: ICSOC, 2013, S. 692–695, 2013.
- [BBKL14b] U. Breitenbücher, T. Binz, O. Kopp und F. Leymann. „Vinothek - A Self-Service Portal for TOSCA“. In: ZEUS 2014, S. 69–72, 2014.
- [NEU11] C. Neufeind, J. Rolshoven, F. Steeg. „Werkzeuge und Verfahren für die Korpuserstellung durch kollaborative Volltexterschließung“. In: Conference of the German Society for Computational Linguistics and Language Technology (GSCL), Hamburg 2011, S. 163–168, 2011.