

# Digitale Nachhaltigkeit in den Geisteswissenschaften durch TOSCA: Nutzung eines standardbasierten Open-Source Ökosystems

## Breitenbücher, Uwe

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

## Barzen, Johanna

johanna.barzen@iaas.uni-stuttgart.de  
Institut für Architektur von Anwendungssystemen,  
Universität Stuttgart, Deutschland

## Falkenthal, Michael

michael.falkenthal@iaas.uni-stuttgart.de  
Institut für Architektur von Anwendungssystemen,  
Universität Stuttgart, Deutschland

## Leymann, Frank

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

## Einleitung

Die digitale Nachhaltigkeit von IT-Anwendungen in der Forschung spielt eine immer größer werdende Rolle, da IT-gestützte Forschungsergebnisse auch Jahre nach deren Publikation reproduzierbar sein müssen, um Dritten das Nachvollziehen und Überprüfen der Ergebnisse zu ermöglichen. Wenn das Forschungsergebnis auf der automatisierten Auswertung strukturiert dokumentierter Daten mittels Softwareprogrammen basiert, wird die stetige und zügige Weiterentwicklung von IT-Technologien jedoch zu einem immer größeren Problem: Werden Forschungsergebnisse beispielsweise mittels eines Windows 95-basierten Programms ermittelt, wird dessen Ausführung mit jeder neuen Generation von Betriebssystemen umständlicher, da sich Schnittstellen ändern und Annahmen nicht mehr erfüllt sind.

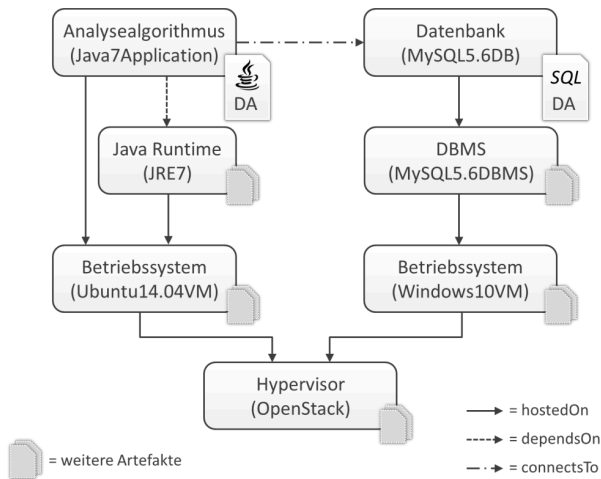
Während diese Probleme für einfache Softwareanwendungen mittels virtueller Maschinen gelöst werden können, sind komplexere Anwendungen mit diesem Ansatz nicht ohne großen manuellen Aufwand reproduzierbar. Basiert ein Forschungsergebnis beispielsweise auf einer umfangreichen softwarebasierten

Simulation, welche unterschiedliche Dienste aufruft, die auf verschiedenen Betriebssystemen ausgeführt werden müssen, erfordert das Aufsetzen der Maschinen und Softwarekomponenten sowie deren Konfiguration detaillierte Expertise und ist mit großem Aufwand verbunden (Breitenbücher et al. 2013).

In diesem Beitrag zeigen wir auf, wie die standardbasierte open-source Technologie *OpenTOSCA* in den Digital Humanities eingesetzt werden kann, um die Reproduzierbarkeit IT-gestützter Forschungsergebnisse unabhängig von Technologieentwicklungen zu ermöglichen. Insbesondere verdeutlichen wir, wie auch komplexe Softwareanwendungen automatisiert bereitgestellt werden können, ohne detaillierte technische Expertise aufweisen zu müssen. Dadurch wird die nachhaltige Entwicklung von Forschungssoftware ermöglicht, indem diese auch Jahre später von Laien ausgeführt werden kann.

## Nutzung des OpenTOSCA Ökosystems zur Sicherung der digitalen Nachhaltigkeit von Forschungsergebnissen

Das OpenTOSCA Ökosystem ist eine Werkzeugsammlung, welche die automatisierte Bereitstellung und Verwaltung von IT-Anwendungen ermöglicht. Die Werkzeuge basieren auf der *Topology and Orchestration Specification for Cloud Applications (TOSCA)* (OASIS 2013), einem OASIS Standard zur portablen Beschreibung von IT-Anwendungen. Der Standard definiert ein Metamodell zur Modellierung von *Anwendungsmodellen*, die alle Komponenten einer Anwendung, beispielsweise Webserver und Datenbanken, sowie deren Beziehungen untereinander beschreiben. TOSCA ist anbieter- und technologieagnostisch, wodurch ein Vendor-Lock-in verhindert wird. Dadurch können beliebige Komponententypen mittels TOSCA beschrieben und in Anwendungsmodellen miteinander kombiniert werden. Zur automatisierten Bereitstellung der modellierten Anwendungen definiert TOSCA die Konzepte der *Deployment Artifacts (DA)* und der *Implementation Artifacts*. Deployment Artifacts stellen die Implementierung einer Komponente dar. Beispielsweise kann die Java-Implementierung eines Analysealgorithmus als Deployment Artifact an das zugehörige Komponentenelement des Modells annotiert werden, siehe Abbildung 1. Managementoperationen, wie beispielsweise ein Installationsskript für einen Webserver, können mittels Implementation Artifacts modelliert werden. Um Anwendungsmodelle inklusive aller Artefakte zu paketieren, definiert TOSCA das selbstbeschreibende Archivformat *Cloud Service Archive (CSAR)*.



*Abbildung 1: Simplifiziert dargestelltes Anwendungsmodell*

Zur automatisierten Bereitstellung TOSCA-basierter Anwendungen werden TOSCA-Laufzeitumgebungen eingesetzt, welche die Anwendungsmodelle interpretieren und alle nötigen Bereitstellungsaktivitäten ausführen, d.h. modellierte virtuelle Maschinen provisionieren, Webserver durch Ausführung von Implementation Artifacts installieren, Komponentenimplementierungen in Form von Deployment Artifacts ausliefern, etc. An der Universität Stuttgart wurde die open-source Laufzeitumgebung *OpenTOSCA* (Binz et al. 2013) sowie das TOSCA-Modellierungswerkzeug *Winery* (Kopp et al. 2013) entwickelt, um TOSCA-basierte Anwendungsmodelle auszuführen und zu erstellen. Das Selbstbedienungsportal *Vinothek* (Breitenbücher et al. 2014) ermöglicht es Nutzern, mittels eines Klicks, die Bereitstellung einer Anwendung zu veranlassen. Abbildung 2 zeigt das Zusammenspiel der Werkzeuge.

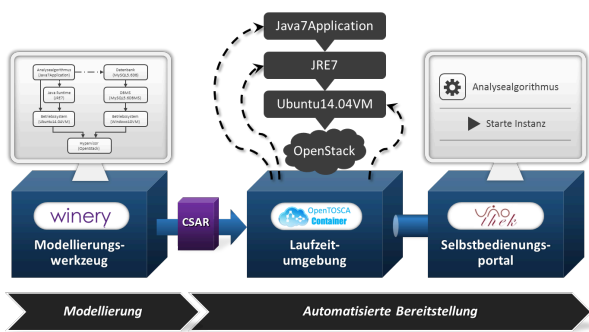


Abbildung 2: Werkzeuge des OpenTOSCA Ökosystems

Dieses OpenTOSCA Ökosystem kann zur Sicherung der digitalen Nachhaltigkeit von Forschungsergebnissen eingesetzt werden, indem Forschungssoftware in Form von CSARs paketiert wird. Durch die Möglichkeit, mit Winery alle erforderlichen Implementierungen in

Form von Deployment und Implementation Artifacts zu spezifizieren, sowie die Struktur der Anwendung inklusive aller Beziehungen zwischen Komponenten zu modellieren, können Anwendungen selbstbeschreibend als CSAR archiviert werden. Diese CSARs können auch Jahre nach deren Entwicklung mittels der OpenTOSCA Laufzeitumgebung provisioniert werden, da alle nötigen Softwareartefakte und Modelle im CSAR enthalten sind und dadurch keine Abhängigkeiten zu externen Dateien existieren. Durch dieses Konzept können beispielsweise „Snapshots“ mehrerer virtueller Maschinen unterschiedlicher Betriebssysteme in Form von Virtual Machine Images in das CSAR gelegt und miteinander assoziiert werden, oder auch spezifische Webserver-Implementierungen, die Jahre später in der genutzten Form nur schwierig auffindbar sind bzw. von Laien nicht gemäß der erforderlichen Konfiguration installiert werden können. Die OpenTOSCA Laufzeitumgebung unterstützt zudem gängige Bereitstellungstechnologien wie Ansible (Hochstein 2014) oder Docker (Mouat 2015), wodurch Artefakte dieser Technologien ohne zusätzlichen Aufwand in das Anwendungsmodell eingebunden werden können. OpenTOSCA ermöglicht dadurch auch die effiziente Orchestrierung mehrerer Bereitstellungstechnologien.

Zur Reproduktion der Forschungsergebnisse muss die Software typischerweise mit auszuwertenden Forschungsdaten gestartet und parametrisiert werden. Häufig ist dies nicht trivial, beispielsweise wenn Data-Mining-Algorithmen auf Basis von Daten über Kostüme in Filmen wiederkehrende Muster finden sollen (Falkenthal et al. 2016). Das Konzept der CSARs ermöglicht auch diese Automatisierung, indem individuelle *Provisionierungspläne* für eine Anwendung modelliert werden können. Ein solcher Plan kann dann automatisiert von OpenTOSCA ausgeführt werden, um die Anwendung zu installieren und wie vorgesehen zu starten.

## Bibliographie

**Binz, Tobias / Breitenbücher, Uwe / Haupt, Florian / Kopp, Oliver / Leymann, Frank / Nowak, Alexander / Wagner, Sebastian** (2013): „OpenTOSCA - A Runtime for TOSCA-based Cloud Applications“, in: *Proceedings of the 11th International Conference on Service-Oriented Computing (ICSOC 2013)*. Springer.

**Breitenbücher, Uwe / Binz, Tobias / Kopp, Oliver / Leymann, Frank / Wettinger, Johannes** (2013): „Integrated Cloud Application Provisioning: Interconnecting Service-Centric and Script-Centric Management Technologies“, in: *On the Move to Meaningful Internet Systems: OTM 2013 Conferences (CoopIS 2013)*. Springer.

**Breitenbücher, Uwe / Binz, Tobias / Kopp, Oliver / Leymann, Frank** (2014): „Vinothek - A Self-Service Portal for TOSCA“, in: *Proceedings of the 6th Central-European Workshop on Services and their Composition (ZEUS 2014)*. CEUR-WS.org.

**Falkenthal, Michael / Barzen, Johanna / Breitenbücher, Uwe / Brüggemann, Sascha / Joos, Daniel / Leymann, Frank / Wurster, Michael** (2016): „Pattern Research in the Digital Humanities - How Data Mining Techniques Support the Identification of Costume Patterns“, in: *Proceedings of the 10th Symposium and Summer School On Service-Oriented Computing (SummerSOC)*. Springer.

**Hochstein, Lorin** (2014): *Ansible: Up and Running*. O'Reilly Media.

**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)*. Springer.

**Mouat, Adrian** (2015): *Using Docker: Developing and Deploying Software with Containers*. O'Reilly Media.

**OASIS** (2013): *Topology and Orchestration Specification for Cloud Applications Version 1.0*.