

Technologienutzung im Kontext Digitaler Editionen – eine Landschaftsvermessung

Neuefeind, Claes

c.neuefeind@uni-koeln.de
Universität zu Köln, Deutschland

Schildkamp, Philip

philip.schildkamp@uni-koeln.de
Universität zu Köln, Deutschland

Mathiak, Brigitte

bmathiak@uni-koeln.de
Universität zu Köln, Deutschland

Harzenetter, Lukas

lukas.harzenetter@iaas.uni-stuttgart.de
Universität Stuttgart, Deutschland

Barzen, Johanna

johanna.barzen@iaas.uni-stuttgart.de
Universität Stuttgart, Deutschland

Breitenbücher, Uwe

uwe.breitenbuecher@iaas.uni-stuttgart.de
Universität Stuttgart, Deutschland

Leymann, Frank

frank.leymann@iaas.uni-stuttgart.de
Universität Stuttgart, Deutschland

Einleitung

Geisteswissenschaftliche Forschung ist schon seit langem eine auch digitale Praxis. Dies spiegelt sich u. a. in den Methoden der Ergebnissicherung wider: Präsentationssysteme, interaktive Visualisierungen, Recherche-Datenbanken und nicht zuletzt Digitale Editionen haben sich neben der klassischen Publikation längst als digitale Instrumente der Ergebnissicherung etabliert. Während für die Persistenz statischer Daten wie z. B. digitaler Dokumente bereits gut etablierte Strategien und Standards wie TEI-XML (TEI 2018) oder OASIS-DITA (OASIS 2018) existieren, stellt Forschungssoftware im Hinblick auf die langfristige Ergebnissicherung noch immer eine besondere Herausforderung dar.

Dies äußert sich in erster Linie in einem Mangel an konkreten Nachhaltigkeitsstrategien und erwächst u. a. aus einem als „Software-Aging“ (Parnas 1994) bekannten Problem, dem Forschungsanwendungen als „Lebende Systeme“ (Sahle/Kronenwett 2013) grundsätzlich unterworfen sind, da sie, wie jede Software, nicht unabhängig von ihrer Laufzeitumgebung bzw. ihrem digitalen Ökosystem gedacht werden können. Die kontinuierliche Evolution dieser Umgebungen sorgt dafür, dass Softwaresysteme, die nicht stetig an diese veränderten Umweltbedingungen angepasst werden, mit der Zeit veralten und letztendlich unbenutzbar werden.

In dem von der DFG geförderten Kooperationsprojekt „SustainLife – Erhalt lebender, digitaler Systeme für die Geisteswissenschaften“, das in einer Zusammenarbeit zwischen dem Data Center for the Humanities (DCH, siehe <http://dch.phil-fak.uni-koeln.de>) der Universität zu Köln und dem Institut für Architektur von Anwendungssystemen (IAAS, siehe <http://www.iaas.uni-stuttgart.de>) der Universität Stuttgart durchgeführt wird, arbeiten wir an einem Lösungsvorschlag für dieses Problem (vgl. dazu Barzen et al. 2018 sowie Neuefeind et al. 2018). Gegenstand des Projekts ist die Adaption und Weiterentwicklung von Verfahren und Technologien aus dem Cloud-Deployment für die Digital Humanities (DH) mit dem Ziel, Management und Provisionierung von DH-Anwendungen zu optimieren und deren Sicherung und nachhaltigen Betrieb zu realisieren.

Das Projekt setzt hierbei auf den OASIS-Standard TOSCA (*Topology and Orchestration Specification for Cloud Applications*, siehe OASIS 2013, OASIS 2016) sowie dessen Open-Source-Implementierung OpenTOSCA (Binz et al. 2013, Breitenbücher et al. 2017). Mithilfe von TOSCA können (Forschungs-)Anwendungen mitsamt ihrer jeweiligen Laufzeitumgebung als Topologien von zusammenhängenden Software-Artefakten und Schnittstellen in standardisierter Weise modelliert werden. Daraufhin können diese Topologie-Modelle und damit alle benötigten Komponenten und Dateien der modellierten Applikation in sogenannten CSARs (*Cloud Service Archives*) paketiert werden, so dass diese portablen Archive von jeder TOSCA-Runtime interpretiert und automatisiert bereitgestellt werden können.

Methodisches Vorgehen

Eine wesentliche Voraussetzung für eine Lösung, die eine standardisierte, TOSCA-konforme Beschreibung von Softwaresystemen vorsieht, ist ein möglichst genaues Bild der technologischen Landschaft innerhalb der Digital Humanities. Grundlage für den im Projekt verfolgten Ansatz ist daher eine ausführliche Bedarfs- und Anforderungsanalyse, anhand derer die Spezifikation häufig eingesetzter Systemkomponenten sowie die Identifikation von Schlüsselkomponenten zur Erstellung von Anwendungsvorlagen erfolgen kann.

Im Projekt setzen wir hierfür im Wesentlichen auf zwei Instrumente: Zum einen quantitativ angelegte Umfragen in der Community in Form von Fragebögen, zum anderen die qualitativ orientierte Untersuchung von ausgewählten Beispielanwendungen auf Basis von gezielten Codeanalysen. Letzteres zielt auf eine genauere Analyse der quantitativ erhobenen Daten bezüglich der eingesetzten Technologien, um konkrete Systemkomponenten identifizieren zu können, die im weiteren Projektverlauf u. a. für die Modellierung von Usecases auf Basis des TOSCA-Standards eingesetzt werden sollen.

Eine erste Kurzumfrage wurde im Rahmen der DHd2018 in Köln unter den Teilnehmern der Konferenz durchgeführt (Barzen et al. 2018). Primäres Ziel des eingesetzten Fragebogens war es, eine explorative Typisierung von DH-Anwendungen vorzunehmen und das Spektrum der dabei eingesetzten Technologien zu konturieren. Die Umfrage ergab im Wesentlichen, dass sich Entscheidungen bezüglich der Technologienutzung vor allem am jeweiligen Anwendungstyp orientieren. Um eine möglichst zielgenaue Analyse typischer Technologien zu erreichen bietet es sich deshalb an, den Skopus der Befragung entsprechend auf einzelne Anwendungstypen hin einzugrenzen.

Technologienutzung im Kontext Digitaler Editionen

Nachdem sich der Zusammenhang zwischen der Art der Anwendung und den konkreten technologischen Entscheidungen besonders deutlich bei Digitalen Editionen zeigte, und da diese zugleich als besonders repräsentative Form digitaler Ergebnissicherung im Bereich der Digitalen Geisteswissenschaften angesehen werden können, haben wir uns dazu entschieden, hier den ersten Ausgangspunkt für eine gezielte Erhebung von Daten über die Technologienutzung zu setzen. Hierfür richteten wir gemeinsam mit dem Cologne Center for eHumanities (CCEH, siehe <http://cceb.uni-koeln.de>) sowie in Kooperation mit der Landesinitiative NFDI der Digitalen Hochschule NRW (siehe <https://fdm-nrw.de>) und der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste (AWK, siehe <http://www.awk.nrw.de>) einen Workshop zum Thema „Nachhaltigkeit Digitaler Editionen“ aus, der am 17.9.2018 in den Räumen der AWK stattfand (siehe <http://dch.phil-fak.uni-koeln.de/nde-workshop.html>).

Der Workshop hatte das Ziel, vorhandene Lösungsansätze und Aktivitäten, die der nachhaltigen Bereitstellung von Digitalen Editionen gewidmet sind, mit der Fachcommunity zu diskutieren. So wird bereits seit einigen Jahren an verschiedenen Stellen an Konzepten gearbeitet, die dieses Problem adressieren. Bspw. wird in der Schweiz derzeit eine „Nationale Infrastruktur für Editionen“ (NIE-

INE, siehe <https://www.nie-ine.ch>) aufgebaut, die auf die Homogenisierung von Editionsprojekten zielt, und in Österreich wurde mit dem „Kompetenznetzwerk Digitale Editionen“ (KONDE, siehe <http://www.digitale-edition.at>) ein umfangreiches Verbundprojekt eingesetzt, das ebenfalls den Aufbau und die Weiterentwicklung einer nationalen Forschungsinfrastruktur für Digitale Editionen zum Ziel hat. Auch in Deutschland gibt es vergleichbare, wenn auch lokaler ausgerichtete Initiativen, die auf die Schaffung einer gemeinsamen Infrastruktur für Editionsprojekte und damit längerfristig auf eine Standardisierung von Digitalen Editionen zielen. Im Rahmen des Workshops hatten wir die Gelegenheit viele der maßgeblichen Akteure als Sprecher zu gewinnen. Unter den Teilnehmern fanden sich ebenfalls viele Personen, die selbst an Digitalen Editionen arbeiten oder an der Nachhaltigkeitsproblematik anderweitig interessiert waren, z.B. aus Perspektive der Drittmittelgeber.

Begleitende Umfrage

Im Vorfeld des Workshops führten wir unter den insgesamt 80 Teilnehmern (70 Gäste und 10 eingeladene Sprecher) eine Umfrage durch, mit dem Ziel, Informationen über die Technologielandschaft speziell in diesem Bereich der DH zu gewinnen. Die Umfrage enthielt Fragen zu persönlichen Vorerfahrungen in der Arbeit mit Digitalen Editionen, zu den formalen Charakteristika von Editionsprojekten sowie zu den dabei eingesetzten Technologien. Schon anhand einer ersten Auswertung der Umfrage, die unmittelbar vor dem Workshop vorgenommen wurde und bei der die Angaben zu 31 verschiedenen Editionsprojekten berücksichtigt wurden, ließen sich deutliche Tendenzen in Bezug auf die eingesetzten Nachnutzungs- und Nachhaltigkeitsstrategien ablesen. So zeigen die in Abbildung 1 wiedergegebenen Angaben bezüglich der Datenmodellierung sehr deutlich, dass mit TEI-XML mittlerweile ein weit verbreiteter Standard zur nachhaltigen Sicherung von Daten vorliegt, der für speziellere Bedarfe durch weitere standardisierte Formate ergänzt wird.

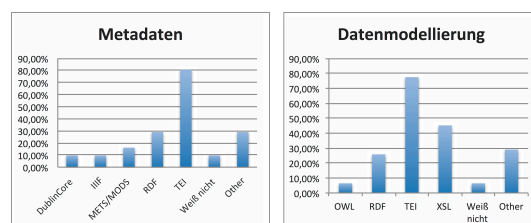


Abbildung 1: Nutzung von Standards für Metadaten und für die Modellierung von Primärdaten.

Aufseiten der Implementierungen dagegen fehlen solche Standards, was entsprechend zu einer deutlich höheren

Heterogenität der eingesetzten Technologien führt (siehe Abbildung 2). Aus der Zusammenschau der verschiedenen Fragekategorien lässt sich ersehen, dass es zwar durchaus auch in Bezug auf die genutzten Technologien und Systemkomponenten Favoriten gibt, jedoch wird unsere Annahme einer technologischen und methodischen Heterogenität des Feldes dennoch klar bestätigt.

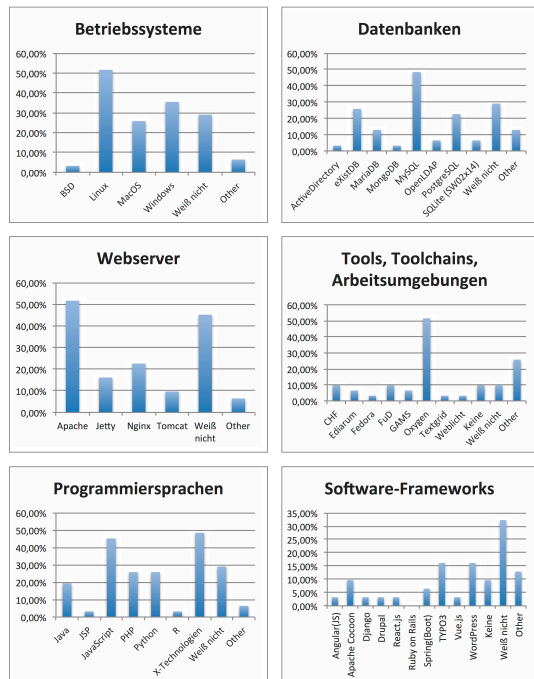


Abbildung 2: Technologienutzung im Kontext Digitaler Editionsprojekte.

Während sich also im Bereich der Primärdaten in Gestalt verschiedener TEI-Dialekte eine zunehmende Standardisierung abzeichnet, so gilt dies nicht für das Layout und die Präsentation von Editionen. Hinsichtlich Aussehen, Funktionalitäten und technischer Architekturen besteht hier weiterhin eine sehr große Heterogenität. Zudem deuten sich hier auch bereits verschiedene Technologie-Stacks an, die sich aus typischen Kombinationen von Systemkomponenten ergeben. Allerdings scheinen sich solche Stacks v. a. innerhalb verschiedener Trägerinstitutionen herauszubilden, was sich in Teilen auf personelle Überschneidungen zwischen Projekten oder auf institutionelle Lösungsansätze zurückführen lässt. Diese Ergebnisse spiegelten sich auch in den Diskussionen auf dem Workshop wider. Während die Nachhaltigkeit der Daten weitestgehend gelöst zu sein scheint, brachte die Frage nach der Nachhaltigkeit der Software eine Reihe von Lösungsansätzen hervor. Die Vor- und Nachteile von verschiedenen Technologie-Stacks wurden kontrovers diskutiert. Weitestgehende Einigkeit

bestand darin, dass eine Digitale Edition aus Daten plus Software besteht, dass also eine alleinige Konservierung der Daten für die Konservierung der Digitalen Edition an sich nicht ausreicht.

Fazit und Ausblick

In unserem Beitrag möchten wir die bisherigen Ergebnisse unserer Anforderungsanalyse an lebende Systeme in den Digital Humanities zur Diskussion stellen. Neben den hier gezeigten Kennzahlen fokussieren wir dabei verschiedene Technologie-Stacks, die typische Kombinationen von Systemkomponenten widerspiegeln. Die hier vorgestellten Ergebnisse der Anforderungsanalyse dienen uns als Basis für die weitere Projektarbeit. Unser Ziel ist es, häufig genutzte Schlüsselkomponenten und typische Anwendungsstrukturen digitaler Erkenntnisträger im Bereich der Digital Humanities zu identifizieren, sie anschließend in TOSCA zu modellieren und zu einer Komponentenbibliothek zusammenzufassen sowie in Form von Anwendungsvorlagen für weitere Modellierungen bereitzustellen, um dadurch konsistente und zukunftssichere Standards und Nachhaltigkeitsstrategien für Forschungssoftware zu etablieren. Über die konkrete Projektarbeit hinaus sind die hier vorgestellten Ergebnisse auch von generellem Interesse für die Community. So sind die erhobenen Daten u. a. auch für Datenzentren wie das DCH Köln von hohem Nutzen, um beispielsweise Strategien und Betreuungskonzepte für Forschungssoftware zu entwerfen.

Bibliographie

- Barzen, J. / Blumtritt, J. / Breitenbücher, U. / Kronenwett, S. / Leymann, F. / Mathiak, B. / Neufeind, C. (2018):** "SustainLife - Erhalt lebender, digitaler Systeme für die Geisteswissenschaften." In: Book of Abstracts der 5. Jahrestagung der Digital Humanities im deutschsprachigen Raum (DHD 2018), Köln 26.2.–2.3.2018, S. 471–474. <https://kups.ub.uni-koeln.de/8085/1/boa-DHD2018-.pdf> [letzter Zugriff 8.1.2019].
- Binz, T. / Breitenbücher, U. / Haupt, F. / Kopp, O. / Leymann, F. / Nowak, A. / Wagner, S. (2013):** "OpenTOSCA - A Runtime for TOSCA-based Cloud Applications". In: ICSOC, 2013, S. 692–695.
- Breitenbücher, U. / Barzen, J. / Falkenthal, M. / Leymann, F. (2017):** "Digitale Nachhaltigkeit in den Geisteswissenschaften durch TOSCA: Nutzung eines standardbasierten Open-Source Ökosystems", in: DHD 2017: Digitale Nachhaltigkeit, S. 235–238. http://www.dhd2017.ch/wp-content/uploads/2017/01/Abstractband_def_24.1.17-1.pdf [letzter Zugriff 8.1.2019].
- Neufeind, C. / Harzenetter, L. / Schildkamp, P. / Breitenbücher, U. / Mathiak, B. / Barzen, J. / Leymann, F. (2018):** "The SustainLife Project – Living Systems in Digital Humanities". In: Proceedings of the 12th Advanced

Summer School on Service-Oriented Computing, 2018 (IBM Research Report RC25681), S.101-112.

OASIS (2013): „*Topology and Orchestration Specification for Cloud Applications Version 1.0*“. 25 November 2013. OASIS Standard. <http://docs.oasis-open.org/tosca/TOSCA/v1.0/os/TOSCA-v1.0-os.html> . [letzter Zugriff 8.1.2019]

OASIS (2016): „*TOSCA Simple Profile in YAML, Version 1.0*“. Edited by Derek Palma, Matt Rutkowski, and Thomas Spatzier. 21 December 2016. OASIS Standard. <http://docs.oasis-open.org/tosca/TOSCA-Simple-Profile-YAML/v1.0/os/TOSCA-Simple-Profile-YAML-v1.0-os.html> . [letzter Zugriff 8.1.2019]

OASIS (2018): „*Darwin Information Typing Architecture (DITA) Version 1.3 Errata 02*“. Edited by Robert D. Anderson and Kristen James Eberlein. 19 June 2018. OASIS Approved Errata. <http://docs.oasis-open.org/dita/dita/v1.3/errata02/os/dita-v1.3-errata02-os.html> [letzter Zugriff 8.1.2019].

Parnas, D. L. (1994): „*Software Aging*“. In: Proceedings of the 16th International Conference on Software Engineering (ICSE 1994). IEEE, Mai 1994, S. 279–287.

Sahle, P. und 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, S. 76–96.

TEI Consortium, eds. (2018): „*TEI P5: Guidelines for Electronic Text Encoding and Interchange*.“ Version 3.4.0 vom 23.7.2018. <http://www.tei-c.org/Guidelines/P5/> [letzter Zugriff 8.1.2019].