

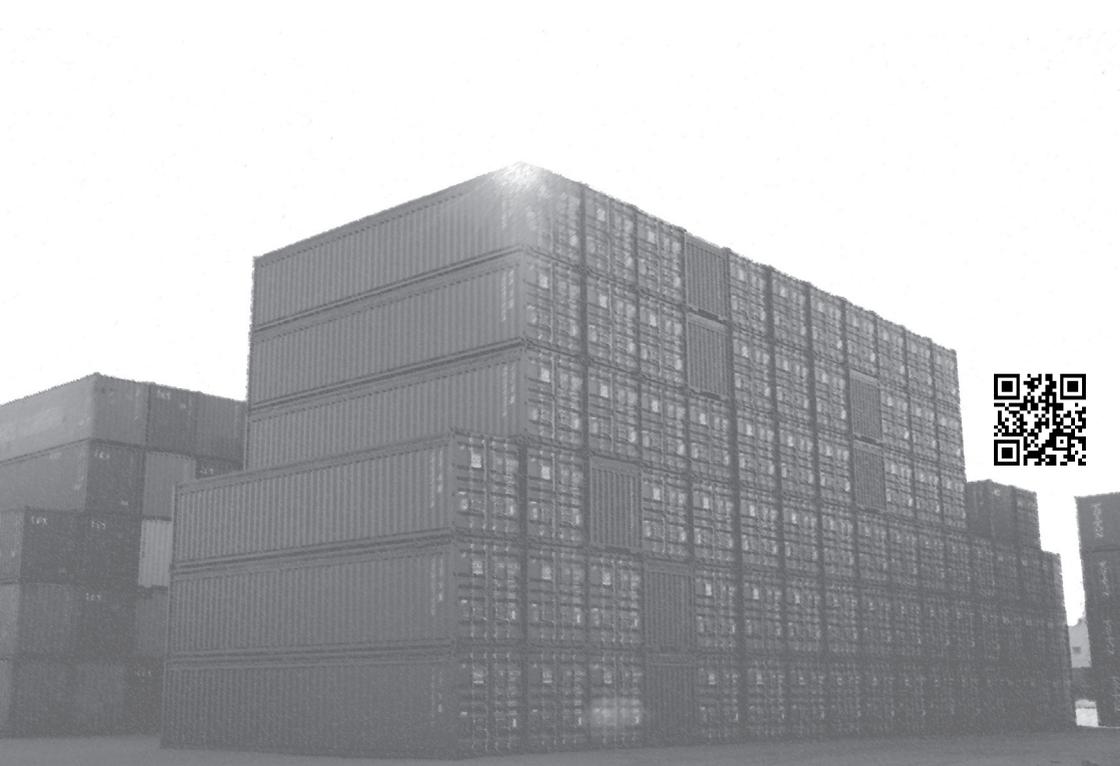
# Forschungsdaten in den **FORGE 2016**

## Geisteswissenschaften

14.-16. September 2016  
Universität Hamburg

— **Jenseits der Daten** —

Nachhaltigkeit für Forschungsanwendungen und Software



Universität Hamburg, Edmund-Siemers-Allee 1 • 14.-16. September 2016



# Mittwoch 14.09.

<b>9:30-10:00</b>	<b>Hörsaal H</b>	<b>Hörsaal K</b>
<b>10:00-10:30</b>	<p><b>[6] Workshop „Reuse or new Development: sustainability of resources and tools for multi-facetted historical data and languages“</b></p> <p>Dr. Alicia Gonzales, Dr. Peter Verkinderen, Dr. Cristina Vertan, (Organisatoren)</p>	<p><b>[7] Workshop „Nachhaltigkeit von Workflows zur Datenkuratierung“</b></p> <p>Hanna Hedeland, Timm Lehmborg, (Organisatoren)</p>
<b>10:30-11:00</b>		
<b>11:00-11:30</b>		
<b>11:30-12:00</b>		
<b>12:00-12:30</b>		
<b>12:30-13:00</b>		
<b>13:00-13:30</b>		
<b>13:30-14:00</b>	<b>Kaffepause</b>	
<b>14:00-14:30</b>	<b>Hörsaal J (Alle weiteren Vorträge)</b>	
	<p><b>Begrüßung</b> Prof. Dr. Anselm Steiger Dr. Kai Wörner</p>	
<b>14:30-15:00</b>	<p><b>[Eingeladen] Nachhaltigkeit von Forschungssoftware</b> Dr. Matthias Katerbow, DFG, Wissenschaftliche Literaturversorgungs- und Informationssysteme (LIS)</p>	
<b>15:00-15:30</b>	<b>Kaffepause</b>	
<b>15:30-16:00</b>	<p><b>[8] Infrastrukturen für wirtschaftswissenschaftliche Forschungsdaten in Deutschland</b> Ralf Toepfer, Sven Vlaeminck</p>	
<b>16:00-16:30</b>	<p><b>[9] Das Web als Plattform für wissenschaftliche Software: Erfahrungen aus der Entwicklung und Pflege einer HTML5 Anwendung</b> Felix Rau, Jonathan Blumtritt, Sebastian Zimmer</p>	
<b>16:30-17:00</b>	<p><b>[10] Entwicklung einer Internetanwendung zur griechisch-lateinischen Epigraphik unter Berücksichtigung von Datennachhaltigkeitsstrategien</b> Hagen Peukert, Franziska Weise</p>	

## Donnerstag 15.09.

<b>9:30-10:00</b>	[12] Einführung von Softwareentwicklung als Service in das Produktportfolio einer wissenschaftlichen Bibliothek - Ein Erfahrungsbericht Zeki Mustafa Dogan, Kristine Schima-Voigt
<b>10:00-10:30</b>	[14] Anwendungskonservierung und die Nachhaltigkeit von Forschungsanwendungen Sven Bingert, Jonathan Blumtritt, Stefan Buddenbohm, Claudia Engelhardt, Simone Kronenwett, Daniel Kurzawe
<b>10:30-11:00</b>	Kaffepause
<b>11:00-12:30</b>	<p style="background-color: black; color: white; padding: 2px;">Foyer vor dem Hörsaal J</p> <p>[40] POSTERSESSION</p> <p>[41] Kompendium der deutschsprachigen historischen Arbeits-, Berufs- und Amtsbezeichnungen (KABA) (Katrin Moeller)</p> <p>[43] Forschungsdaten in Repositorien auf der Basis von MyCoRe (Wiebke Oeltjen)</p> <p>[44] CoRS-O-Matic Eine freie, webbasierte Anwendungssoftware für linguistische Forschungsdaten (Nils Cordes, Anja Weingart)</p> <p>[46] Lassen sich technische Plattformen bei Forschungsanwendungen konsolidieren und gewinnt man damit etwas? (Kai Wörner)</p> <p>[47] Nutzung einer Repositoriums-anwendung zur Modellierung epigraphischer Daten (Hagen Peukert, Franziska Weise)</p> <p>[48] Wieviele Komponenten verträgt ein nachhaltiges Repository (Daniel Jettka, Kai Wörner)</p> <p>[49] Das Datenrepository Manuskriptkulturen (Karsten Helmholz, Stefan Thiemann)</p>
<b>12:30-13:00</b>	
<b>13:00-13:30</b>	
<b>13:30-14:00</b>	[17] Elemente eines Geschäftsmodells für etablierte wissenschaftliche Software Thomas Schmidt, Kai Wörner, Hanna Hedeland
<b>14:00-14:30</b>	[19] AGATE – European Academies Internet Gateway: Konzept für eine digitale Infrastruktur für die sozial- und geisteswissenschaftlichen Forschungsvorhaben der europäischen Wissenschaftsakademien Dominik Adrian, Carolin Ott, Ulrike Wuttke
<b>14:30-15:00</b>	[21] Consulting Workflow for Humanities Research Data Brigitte Mathiak, Jonathan Blumtritt
<b>15:00-15:30</b>	Kaffepause
<b>15:30-16:00</b>	[23] Digitale Nachhaltigkeit von Forschungsanwendungen durch Microservices Michael Haft, Anna Neovesky, Gabriel Reimers
<b>16:00-16:30</b>	[25] Von implizitem Wissen zu nachhaltigen Systemanforderungen - Interdisziplinäre Zugänge in der frühen Anforderungsanalyse am Beispiel der Digital Humanities Dr. Bianca Meise, Yevgen Mexin, Franziska Schloots, Björn Senft, Anastasia Wawilow, Prof. Dr. Gregor Engels, Prof. Dr. Dorothee Meister, Prof. Dr. Jörg Müller-Lietzkow, Prof. Dr. Gerd Szwillus
<b>16:30-17:00</b>	[27] Contemporary challenges and future perspectives for sustainable reuse of archived textual qualitative research data Reinhard Messerschmidt, Brigitte Mathiak

## Freitag 16.09.

<b>9:30-10:00</b>	[29] Ist das Forschung oder kann das weg? Methoden zur Reduktion und Restrukturierung von Forschungsdaten und -anwendungen Iris Vogel	
<b>10:00-10:30</b>	[30] Wir machen dann mal Digital Humanities Software — Ein Erfahrungsbericht Julia Damerow, Dirk Wintergrün	
<b>10:30-11:00</b>	Kaffepause	
<b>11:00-11:30</b>	[33] Eine interdisziplinäre Bestandsaufnahme Alexander Struck	
<b>11:30-12:00</b>	[36] Datenautorenschaft als Option für nachhaltige Datennutzung Katrin Moeller	
<b>12:00-12:30</b>	[38] Aufbau und Hosting nachhaltiger Softwarelösungen durch die Verbundzentrale Kathleen Neumann	
<b>12:30-13:00</b>	Verabschiedung	
<b>13:00-13:30</b>	(Ende der Konferenz: 12:45)	
<b>13:30-15:00</b>	Sitzungssaal des akad. Senats Treffen der DHd-Arbeitsgruppe „Datenzentren“ [Nur für Mitglieder der Arbeitsgruppe]	Raum 125 Workshop „MyCoRe Migration“ Kathleen Neumann
<b>15:00-17:00</b>		

# Workshop: Reuse or New Development: sustainability of resources and tools for multi-facetted historical data and languages

*Dr. Cristina Vertan, University of Hamburg, ERC Project TraCES*

*Dr. Alicia Gonzales, University of Hamburg, ERC Project COHBUNI*

*Dr. Peter Verkinderen, University of Hamburg, ERC Project The Early Islamic Empire at Work*

Data in humanities, especially historical data, is characterized by a strong presence of vague information and uncertainty. The available Content Management Systems and annotation tools have often disregarded the requirements of research projects dealing with fuzzy data, languages with non-concatenative morphologies and scripts of non-Latin writing systems. Additionally, data encoding standards often overstress the importance of mere standardization at the expense of human readability and efficiency in terms of storage and parsing performance. Similarly, morphological tag sets and natural language processing frameworks primarily based on Indo-European languages are presented as universal solutions, but fail to meet some of the linguistic phenomena characteristic of other languages.

This workshop will bring together scholars using annotation tools for non-Western languages with people involved in the development of such tools and content management systems, in order to exchange experiences, discuss problems, and search for ways to overcome these barriers.

Three main directions will be discussed:

## 1. Sustainability in terms of data repository

- What kind of data management framework is needed for which type of data?
- How can one store enough explanation about vague, imprecise and ambiguous data (like place and person names, historical dates, complex relationships among actors)?
- How can currently available systems handle multilingual issues?

## 2. Sustainability of (annotation) tools

Analysis of historical data analysis implies: annotation of textual material in languages which usually:

- either are no longer (actively) spoken like Ge'ez, or changed a lot from their present form (classical Arabic), or the writing system is no longer known (e.g. Maya);
- are less resourced from the point of view of digital resources;
- the linguistic structure/features for this languages differ considerable from the modern ones which are usually object of most available annotation tools;
- annotation cannot be done on the original script but on transliteration but both versions have to be kept synchronised;
- grammatical structure of the language is not completely researched. Some phenomena are discovered during the analysis, one cannot consider them in the modelling phase.

The workshop intends to tackle issues like:

- which annotation tools can be considered for annotation; which limitations do they have?
- what does the design of a new annotation tool imply and how to ensure sustainability?
- visualisation of complex annotations;
- local solutions versus general tools;
- frequent requirements of structural complex languages: multilevel annotation, correction of text during annotation, multi-level segmentation.

## 3. Sustainability of annotated objects (standards)

Nowadays XML, and TEI-XML in particular, is generally accepted as an export format. Whilst for this purpose TEI-XML is very useful, ensuring the portability of data, there are still issues to be discussed like:

- For internal processing is TEI-XML quite difficult to use. In case of intensive fine-grained interlinking between components practically impossible. Thus project-specific solutions are often required and "EXPORT"-function should ensure a TEI-XML output at the end
- TEI-XML ensures that data is readable by everyone. Can this data then be also processed easily by a third party? - Which other formats could be used (e.g. JSON)?
- Are the current existent Tag-Sets (like the Tübinger Tagset) enough specified for all other non-European languages?

<https://www.traces.uni-hamburg.de/en/forge2016.html>

# Workshop: Nachhaltigkeit von Workflows zur Datenkuratierung

*Hanna Hedeland, Universität Hamburg, CLARIN-D/HZSK*

*Timm Lehmberg, Universität Hamburg, INEL/HZSK*

Bei der Reflexion von Fragen der Nachhaltigkeit digitaler linguistischer Ressourcen stehen traditionell die Daten selbst im Vordergrund. Dementsprechend beschränken sich viele der in diesem Kontext entwickelten Lösungen im Bereich der Datenmodellierung und -aufbereitung auf die Schaffung nachhaltiger Datenformate und die Implementierung entsprechender Zugriffsplattformen. Tatsächlich jedoch stellen nachhaltig aufbereitete Daten lediglich das Endprodukt von oft sehr komplexen und aufwendigen Workflows der Datenkuratierung dar. Die Dokumentation und Wiederverwendbarkeit der in diesem Rahmen eingesetzten bzw. entwickelten Verfahren und Werkzeuge, ist jedoch ein ebenso erstrebenswertes Ziel, wie die Nachhaltigkeit der Daten selbst.

In dem Workshop wird zunächst über Arbeiten aus diesem Bereich im Rahmen mehrerer Langzeit- und Infrastrukturvorhaben referiert:

**Thomas Schmidt** (Institut für Deutsche Sprache, Mannheim) informiert über Verfahren der Datenkuratierung am Archiv für Gesprochenes Deutsch (AGD).

**Peter Bouda** und **Felix Rau** (Universität zu Köln) berichten über die Poio API: Eine im Rahmen eines CLARIN-D Kurationsprojektes entwickelte Bibliothek zur Verarbeitung und Analyse von im Bereich der Dokumentationslinguistik gebräuchlichen Datenformaten.

**Daniel Jettka** und **Tommi Pirinen** (Hamburger Zentrum für Sprachkorpora) referieren über Datenkuratationsworkflows, die in Projekten am HZSK und im Langzeitprojekt INEL zum Einsatz kommen.

Die Beiträge werden anschließend unter folgenden Aspekten der Nachhaltigkeit von Datenkuratierungsworkflows diskutiert:

- Modellierung und Dokumentation von Workflows zugunsten
  - der Optimierung und Überwachung von Projektabläufen,
  - der Möglichkeit einer (teil-)automatischen Kontrolle von Kuratierungsprozessen,
  - der Nachvollziehbarkeit von Forschungsergebnissen, die auf der Grundlage kuratierter Datensammlungen entstanden sind.
- Modularisierung von Workflows zum Zweck einer nachhaltigen Nutzbarkeit bzw. Wiederverwendbarkeit ganzer Workflows bzw. Teilen davon.
- Skalierbarkeit und Anpassung von Workflows u. a. an sich verändernde technische und wissenschaftlich/methodische Rahmenbedingungen.

<https://www.gwiss.uni-hamburg.de/gwin/ueber-uns/forge2016/programm/workshop-workflows.html>

# Infrastrukturen für wirtschaftswissenschaftliche Forschungsdaten in Deutschland

*Ralf Toepfer & Sven Vlaeminck, Deutsche Zentralbibliothek für Wirtschaftswissenschaften (ZBW)*

Forschungsdaten bilden eine immer wichtiger werdende Grundlage für wissenschaftliche Erkenntnisse. So ist beispielsweise im Bereich der Volkswirtschaftslehre in den letzten Jahrzehnten der Anteil von Forschungsartikeln, die auf der Auswertung empirischer Daten beruhen stark angestiegen. Mit der steigenden Verfügbarkeit von Daten treten jedoch auch neue Herausforderungen bezüglich des Umgangs mit diesen Daten auf den Plan: Wie soll mit Daten verfahren werden, die in Projekten generiert wurden und wem gehören diese? Wie sollten Forschungsdaten dokumentiert werden? Wo können diese Daten gespeichert und langfristig bereitgestellt werden und wer trägt die Kosten? Welche rechtlichen Regelungen zu Daten- und Persönlichkeitsschutz werden möglicherweise berührt?

Vor diesem Hintergrund entwickelt sich aktuell eine Forschungsdateninfrastruktur an der zahlreiche Akteure auf allen Ebenen (lokal, regional, national, international sowie disziplinspezifisch, interdisziplinär) beteiligt sind. Die Präsentation skizziert (1.) die wirtschaftswissenschaftliche Forschungsdatenlandschaft in Deutschland und stellt (2.) die Aktivitäten der ZBW im Bereich Forschungsdatenmanagement vor. Vorgestellt werden vor allem das im Aufbau befindliche Forschungsdatenrepositorium SowiDataNet sowie das ZBW Journal Data Archive. Die Präsentation schließt (3.) mit einigen „lessons learned“, die für die Entwicklung von Forschungsdateninfrastrukturen für die Wissenschaft unserer Auffassung nach von zentraler Bedeutung sind.

# Das Web als Plattform für wissenschaftliche Software: Erfahrungen aus der Entwicklung und Pflege einer HTML5 Anwendung

*Felix Rau, Jonathan Blumtritt, Sebastian Zimmer*

Mit der Standardisierung von HTML5 wurde das Web auch als Technologieplattform zur Entwicklung von Applikationen positioniert. Dieser Anspruch war und ist verbunden mit der großen Verheißung unkompliziert plattformübergreifende Anwendungen schaffen zu können. The „write once, deploy anywhere“ promise of HTML5 and the Open Web platform<sup>1</sup>, wie das W3C dieses zentrale Versprechen nennt.

Wir berichten in unserem Vortrag von unseren Erfahrungen bei der Entwicklung der HTML5 Applikation CMDI Maker. Dabei gibt es neben klaren Vorteilen auch viele Herausforderungen und – teilweise sehr tiefgreifende – Probleme. Wenn sich auch das Open Web als Plattform für unsere Anwendung CMDI Maker bewährt hat, so liegt dies vor allem an den sehr spezifischen Anforderungen unserer Benutzer. Am Beispiel von CMDI Maker kann man dabei sehr gut die Stärken, aber auch klar die Schwächen des Webs als Technologieplattform für wissenschaftliche Software erkennen.

CMDI Maker ist eine offline-fähige HTML5-Anwendung zur schnellen Generierung von CMDI-Metadaten. CMDI Maker erfüllt damit das Bedürfnis feldforschender Linguisten, situationsunabhängig – im Feld oder im Büro – große Sätze an Metadaten zu erstellen und den Forschern repetitive Eingaben und offensichtliche Strukturierungen abzunehmen. Die Wahl der Technologien ergab sich aus den Anforderungen der User Community an die Software. Ziel war es eine Anwendung zu schaffen, die plattformunabhängig ist, eine dem Benutzer vertraute und intuitive Benutzerschnittstelle besitzt, in Feldforschungssituationen Updates erfahren kann und auch ohne Internetverbindung voll funktionsfähig ist.

Für Anwendungen mit einem solchen Anforderungsprofil bietet das Web mit offline-fähigen HTML5-Anwendungen eine attraktive Plattform. Besonders die Unabhängigkeit vom Betriebssystem und Betriebssystemversion ist attraktiv für wissenschaftliche Software, da eine HTML5-Applikation auf allen Plattformen funktionsfähig ist, auf denen es einen HTML5-fähigen Browser gibt.

Darüberhinaus sind Benutzer mit interaktiven Webseiten und Webapplikationen vertraut, so dass Bedienkonzepte und Interfacekomponenten auch für nicht versierte Nutzer gut verständlich sind. Webapplikationen, die aus HTML, JavaScript und CSS bestehen, haben darüberhinaus einen vergleichsweise kleinen Datenumfang und können auch über schlechte Internetverbindungen aktualisiert werden. Die mit HTML5 eingeführte persistente

Datenspeicherungstechnologie Web Storage<sup>2</sup> ermöglicht eine Nutzung und Speicherung der Daten ohne Verbindung zum Internet.

Die Realität dieser Versprechen ist allerdings deutlich komplexer, so führt die relative Neuheit von HTML5 zu einer Dynamik, in der zentrale Technologien verworfen werden und durch noch nicht stabile Nachfolger ersetzt werden, wie dies gerade durch die Ablösung der Web Storage Technologie durch die Web Worker<sup>3</sup> geschieht.

In unserem Vortrag präsentieren wir die Vor- und Nachteile der Webplattform für wissenschaftliche Software und diskutieren für welche Anforderungsprofile HTML5 Anwendungen eine vielversprechende Technologie sein können.

---

<sup>1</sup> <https://www.w3.org/2014/10/html5-rec.html.en>

<sup>2</sup> <https://www.w3.org/TR/webstorage/>

<sup>3</sup> <https://html.spec.whatwg.org/multipage/workers.html>

# Entwicklung einer Internetanwendung zur griechisch-lateinischen Epigraphik unter Berücksichtigung von Datennachhaltigkeitsstrategien

Hagen Peukert, Universität Hamburg

Franziska Weise, Universität Hamburg

## Zusammenfassung

Der Vortrag beschreibt die Erstellung einer nachhaltigen Internetanwendung zur Suche, Visualisierung und Pflege epigraphischer Daten auf Basis des MyCoRe-Frameworks. Wir zeichnen die Überführung der epigraphischen Daten von einer idiosynkratischen Datenbankapplikation hin zu einer xml-basierten, generischen Datenverwaltung mit moderner Suchtechnologie nach und beschreiben die spezifischen Anforderungen, die sich aus den Erfordernissen der Inschriftenforschung ergeben.

## Einleitung

Methodische Zugänge der Digital Humanities lassen sich zunehmend in allen Bereichen geisteswissenschaftlicher Forschung beobachten. Darunter nehmen die Disziplinen, welche auf die ständige Verfügbarkeit von empirischen Daten angewiesen sind und deren Datenbeschaffenheit und -struktur es erlauben, seither eine Führungsrolle in der Entwicklung entsprechender Anwendungen ein. Die griechische wie lateinische Epigraphikforschung hat sich in Teilen bereits in einer frühen Entwicklungsstufe diesen neuen methodischen Ansätzen gestellt. Ausgangspunkt des vorliegenden Projekts sind inschriftliche Zeugnisse des antiken Kleinasien, die in einem über 12 Jahre (1993-2006) geförderten DFG-Projekt gesammelt und als edierte Volltexte in einer Online-Datenbank überführt wurden. Den damaligen technologischen Standards entsprechend, wurden die Ergebnisse des Projekts in einer SQL-Datenbank abgelegt und mittels einer skriptbasierten Anwendung zur Verfügung gestellt. Zur Weiterführung und Fortentwicklung des Projekts ist es Ziel, die dort erhobenen Inschriftendaten unabhängig von den Technologieschüben auf Dauer verfügbar und pflegbar zu halten.

## Anforderungsanalyse und Funktionsumfang

Für den Umgang mit epigraphischen Texten ist es wichtig, alle Funde einer Region möglichst vollständig zu dokumentieren und mit zusätzlichen Informationen anzureichern. Aus der Gesamtheit aller Dokumentationen und durch die geschickte Kombination und Filterung der Einzeldokumentationen lassen sich dann effizient Belege finden, die historische Hypothesen bestätigen oder verwerfen. Zu den Einzelinformationen gehören Daten über den genauen Fundort, die Art, Mono- oder Bilingualität, den Inschriftenträger sowie die Datierung des Fundes.

Der Fundort kann anhand von Geodaten relativ genau erfasst und auf jeweils einer antiken und modernen Karte fixiert werden. Die Kategorisierung des Fundes erfolgt anhand eines festen Kataloges und richtet sich nach der Art der Inschrift, z.B. Bauinschriften, Weihinschriften oder Urkunden.

Kern eines Datensatzes stellen die digitalisierte Inschrift und deren Inhaltsinterpretation dar. So ist der Inschriftentext in einer klar definierten Transkription ediert, in der unlesbare Passagen oder auf plausible Annahmen beruhende Textergänzungen separat ausgezeichnet werden. Ein kritischer Apparat erfasst Varianten der Lesung früherer Herausgeber und präsentiert diese als Anmerkung zum Text. Darüber hinaus werden die Informationen mit bibliographischen Angaben, einer Auflistung der Editionen, einem inhaltlichen Kommentar sowie einer Übersetzung ergänzt. Die Beschreibung und Kommentierung der Texte ist ein wichtiger Bestandteil der epigraphischen Forschung und wird für jede Inschrift gesondert erstellt.

Unverzichtbar ist zudem, Inschrift und Träger in ihrem Kontext zu erfassen. Das Trägermedium wird dafür genau beschrieben und in Abbildung – sofern vorhanden und bildrechtlich vereinbar – dargestellt. Typische Inschriftenträger sind Marmorstelen, Altäre oder Grabsteine.

Die beschriebenen Informationsfelder sollen konsequenterweise als eine kombinierte Suche mit unterschiedlicher Fuzziness zugänglich gemacht und deren Ergebnisse zunächst auf einer Übersichtsseite mit nur wenigen Angaben aufgelistet werden. In einer Detailansicht werden dann die weiteren Ausführungen zur Beschreibung des Inschriftenträgers, zur Datierung, zum Kommentar und kritischen Apparat sowie Karten und Abbildungen der einzelnen Inschrift sichtbar. Diese neue Anwendung soll langfristig der Öffentlichkeit zur Verfügung stehen und daher in den Workflow zur Instandhaltung geisteswissenschaftlicher Forschungsdaten des zuständigen Datenzentrums integriert sein.

### **Umsetzung mit dem MyCoRe-Framework**

Die Nachhaltigkeitsstrategie der Universität Hamburg sieht vor, dass geisteswissenschaftliche Forschungsdaten nach Möglichkeit auf Dauer zugänglich bleiben. In der konkreten Umsetzung bedeutet dies die Verwendung des MyCoRe-Frameworks, für das entsprechende Strukturen im Rechenzentrum der Universität vorgehalten werden. Daraus folgt, dass die Datenmodellierung auf die Belange des MyCoRe-Kerns angepasst wird. Die SQL-strukturierten Daten werden deshalb mit Hinblick auf die MyCoRe-interne Verarbeitung und den Anforderungen an die Anwendung analysiert und im XML-Standard modelliert. Dazu gehört, dass grundsätzlich zwischen zwei Arten von Datenfeldern unterschieden werden muss: Datenfelder, die voraussichtlich ständiger Veränderung unterliegen und Datenfelder, die sich nur wenig ändern (Klassifikationen, oft typische 1:n-Relationen in der Logik relationaler Systeme). Die zu verwendenden Klassifikationen werden wie die übrigen SQL-Datensätze im xml-Format hinterlegt und skriptbasiert ausgelesen. Sie sind Gegenstand stetiger Erweiterung.

Zusätzliche Programmlogik wird zusammen mit den entsprechenden Masken zur Suche, Darstellung und Eingabe sowie Änderung von Datensätzen in den Stylesheets gehalten. Auf diese Weise wird sichergestellt, dass die idiosynkratischen Änderungen, die nur im vorliegenden Projekt wichtig sind, von der allgemein gültigen Anwendungslogik zur Verarbeitung digitaler Inhalte im MyCoRe-Kern getrennt sind. So kommt es bei der Aktualisierung des MyCoRe-Systems zu keiner Beeinträchtigung der vorgenommenen Anpassungen und die Anwendung kann neben vielen weiteren geisteswissenschaftlichen MyCoRe-Anwendungen aufwandsminimierend verwaltet werden. Die speziellen Erfordernisse, die sich bei der mehrsprachigen Suche ergeben, lassen sich über Konfigurationsdateien und Mappings in den getrennt eingefügten Komponenten der Suchtechnologie des Apache Solr vornehmen. Dabei finden die Transkriptionen der Inschriften ebenso Berücksichtigung wie die optionalen Diakritika des Griechischen. Der Upload und die Verwaltung der Digitalisate, insofern vorhanden, und die Verwaltung von Zugriffsrechten und Rollen werden ebenfalls als gekapselte Komponenten mit der Standardtechnologie des MyCoRe-Systems eingebunden. Lediglich die Menüauswahl, Positionierung und Layout unterliegen kleineren Anpassungen.

# Einführung von Softwareentwicklung als Service in das Produktportfolio einer wissenschaftlichen Bibliothek - Ein Erfahrungsbericht

*Zeki Mustafa Dogan, Kristine Schima-Voigt,  
Niedersächsische Staats- und Universitätsbibliothek Göttingen*

Als eine der größten wissenschaftlichen Bibliotheken Deutschlands beteiligt sich die Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB) durch verschiedene Aktivitäten an Forschungsprozessen. Diese Entwicklung kann anhand der Beteiligung der SUB an ca. 30 verschiedenen wissenschaftlichen Projekten jedes Jahr sowie durch das durchschnittliche Drittmittelaufkommen in Höhe von ca. 4,5 Millionen Euro p.a. beobachtet werden. Im Zuge dieser Aktivitäten wurde 2009 der Bedarf der Gruppierung der Software-Entwickler in einer Organisationseinheit erkannt und das Team Software- und Service-Entwicklung mit zwei Entwicklern und einem Teamleiter gegründet. Inzwischen besteht die Gruppe aus neun EntwicklerInnen und vier ProjektleiterInnen und bietet Softwareentwicklung als Service für die Fragestellungen aus dem bibliothekarischen und wissenschaftlichen Umfeld. Zu den Abnehmern dieses Services gehören neben den Abteilungen der SUB verschiedene Zentren, Fachbereiche und Einrichtungen der Universität, die Akademie der Wissenschaften zu Göttingen sowie nationale und internationale Wissenschaftler und Projekte u.a.

In der Anfangsphase wurden die Entwicklungsarbeiten projektgetrieben durchgeführt. Ein Entwickler war einem Projekt zugeordnet und für mehrere Aufgabenbereiche, wie beispielsweise Architekturauswahl, Datenmodellierung, Backend- Frontendprogrammierung, Tests und Betrieb von IT-Systemen während der Projektlaufzeit zuständig. Ungeachtet dessen, dass es nahezu unmöglich war, diese Kompetenzvielfalt in einer Person zu finden, förderte diese Vorgehensweise auch nicht die Erkennung der Synergien zwischen verschiedenen Projekten mit vergleichbaren Anforderungen. Die autarke Architekturauswahl in den einzelnen Projekten führte zu einer unüberschaubaren Vielfalt

von zu betreibenden Technologien, so dass eine Überführung der Software als Projektergebnis in den Regelbetrieb mit deutlichem Mehraufwand verbunden war bzw. auf berechtigtes Unverständnis seitens der Systemadministratoren stieß. Es war nur schwer möglich die entwickelte Software zu pflegen, zu erweitern oder ein Service-Level-Agreement anzubieten, da einheitliche Policies zu Dokumentation, Codequalität, Lizenzierung, Entwicklungsmethoden und -werkzeugen fehlten. Die Entwickler haben nicht gegenseitig voneinander profitieren können - es gab keine „lessons learned“.

Nach der systematischen Analyse der Probleme wurden verschiedene Maßnahmen eingeleitet, um die Situation sowohl für die Service-Anbieter als auch für die Service-Abnehmer zu verbessern und von einer projektgetriebenen Vorgehensweise zum prozess- und skill-basierten Arbeitsmodell zu wechseln:

- Etablierung eines Entwicklungsframeworks und entsprechender Werkzeuge anhand Continuous-Integration
- Wechsel vom Wasserfall-Modell zu agilen Methoden wie Scrum beim Projektmanagement, Einführung der entsprechenden Rollen wie z.B. Product Owner
- Verbesserung der Kommunikation mit den Service-Abnehmern bzw. Stärkung der Kooperation zwischen Systemadministratoren und Entwicklern (s. auch DevOps)<sup>1</sup>
- Pooling von Entwicklern und zeitweise Zuordnung dieser zu Projekten anhand der Projektanforderungen

---

<sup>1</sup> A. Arooni, "Keynote Presentation: From Dashboard Management to an Improvement Index for the Teams," Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2014 Joint Conference of the International Workshop on, Rotterdam, Netherlands, 2014, pp. 135-135

- Konsolidierung der Middleware-Architektur und Platform-as-a-Service für Betrieb und Entwicklung<sup>2</sup>
- Organisationsweite Formalisierung der Überführung der digitalen Dienste in den Regelbetrieb

Wie bei jedem Änderungsprozess war die Umsetzung der einzelnen Maßnahmen mit verschiedenen Herausforderungen organisatorischer, technischer und finanzieller Art verbunden und wird teilweise noch immer fortgesetzt. Es wurden dennoch messbare Erfolge verzeichnet wie z.B. in der Zufriedenheit der Service-Anbieter und -Abnehmer, Wiederverwendbarkeit der Quellcode, Skalierbarkeit der Architektur<sup>3</sup> u.a.

Seit der Umstellung wird darauf geachtet, dass die Entwickler auch projektübergreifend zusammensitzen, um einen steten Austausch zu ermöglichen. Jedes Projektteam besteht aus mindestens drei Experten. Zu den Teams gehören auch Systemadministratoren, die sich an Architekturentscheidungen beteiligen, Übernahmekriterien festlegen und eine Entwicklungsumgebung durch Betriebssystemvirtualisierung (z.B. via Docker<sup>4</sup>) bereitstellen, welche vergleichbare Spezifikationen wie die spätere Produktivumgebung besitzt.

In zweiwöchigen Intervallen werden Funktionalitäten entwickelt, die in einer abschließenden Ergebnisbesprechung präsentiert werden. Die Wissenschaftler sind als Serviceabnehmer eingeladen, an diesen Ergebnisbesprechungen teilzunehmen und dort mit dem Entwicklerteam in Kontakt zu treten. Sie werden auf diese Weise aktiv in den Entwicklungsprozess mit einbezogen und haben die Möglichkeit, sich die neuen Funktionalitäten anzuschauen und direkt Feedback zu geben. Während der Entwicklungsintervalle steht den Wissenschaftlern ein Koordinator - der so genannte Product Owner - als ständiger Ansprechpartner und „Übersetzer“ zur Verfügung.

Die kurzgetaktete Zusammenarbeit mit den Wissenschaftlern ermöglicht es, flexibel auf die wechselnden Anforderungen der Wissenschaftler zu reagieren und erspart allen Beteiligten die Probleme des Modells Big-Design-Up-Front<sup>5</sup>. Darüber hinaus gewährleistet das agile Modell, dass der Entwicklerpool mit jedem Projekt an Expertise gewinnt und das Erlernte nicht mit Projektende und einem Mitarbeiter verschwindet. Sollte ein Entwickler durch Krankheit, Urlaub etc. ausfallen, ermöglicht die Teamkonstellation und der ständige Austausch, dass Aufgaben von einem anderen Entwickler übernommen werden können und die Risiken in Bezug auf den Projekterfolg minimiert werden können.

Nach erfolgreicher Abnahme der Anwendung anhand der vordefinierten Kriterien wird die Anwendung in den Regelbetrieb überführt. Sollte dieser Schritt technische oder organisatorische Auswirkungen haben, wie z.B. die Einführung einer neuen Middleware-Technologie oder Bedarf an zusätzlichen Systemadministratoren-Ressourcen, so werden diese im Rahmen des formalisierten Prozesses berücksichtigt und an die Organisationsleitung vorgelegt. An dieser Stelle findet der Wechsel von Softwareentwicklung als Service zu Software als Service statt.

---

<sup>2</sup> Y. Zheng, J. Pang, J. Li and L. Cui, "Business Process Oriented Platform-as-a-Service Framework for Process Instances Intensive Applications," Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2012 IEEE 26th International, Shanghai, 2012, pp. 2320-2327

<sup>3</sup> C. A. Ardagna, E. Damiani, F. Frati, D. Rebecani and M. Ughetti, "Scalability Patterns for Platform-as-a-Service," Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on, Honolulu, HI, 2012, pp. 718-725

<sup>4</sup> R. Dua, A. R. Raja and D. Kakadia, "Virtualization vs Containerization to Support PaaS," Cloud Engineering (IC2E), 2014 IEEE International Conference on, Boston, MA, 2014, pp. 610-614.

<sup>5</sup> T. J. Lehman and A. Sharma, "Software Development as a Service: Agile Experiences," 2011 Annual SRII Global Conference, San Jose, CA, 2011, pp. 749-758.

# Anwendungskonservierung und die Nachhaltigkeit von Forschungsanwendungen

*Sven Bingert, Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen mbH*

*Jonathan Blumtritt, Data Center for the Humanities, Universität zu Köln*

*Stefan Buddenbohm, Max-Planck-Institut zur Erforschung multireligiöser und multiethnischer Gesellschaften  
Göttingen*

*Claudia Engelhardt, Niedersächsische Staats- und Universitätsbibliothek Göttingen*

*Simone Kronenwett, Data Center for the Humanities, Universität zu Köln*

*Daniel Kurzawe, Niedersächsische Staats- und Universitätsbibliothek Göttingen*

## Executive Summary

Der Vortrag beleuchtet die Bewahrung von Anwendungsumgebungen, die mit ihr verbundenen Herausforderungen und Lösungsansätze aus der Perspektive zweier geisteswissenschaftlicher Datenzentren.

## Ausgangslage

Die Verfügbarkeit von Forschungsdaten und -ergebnissen ist ein Schlüsselfaktor für erfolgreiche Forschung. Um dies auch in den Geisteswissenschaften zu gewährleisten müssen allerdings nicht nur die sog. Primärdaten, sondern auch die damit verknüpften Forschungsanwendungen und Software erhalten und gesichert werden, da nur so eine sinnvolle Nachnutzung der geisteswissenschaftlichen Forschungsdaten möglich ist.<sup>1</sup> Datenzentren wie das Data Center for the Humanities (DCH)<sup>2</sup> in Köln oder das Humanities Data Centre (HDC) in Berlin/Göttingen adressieren diesen Zusammenhang und rücken dabei insbesondere geisteswissenschaftliche Forschungsdaten und ihre speziellen Charakteristika in den Mittelpunkt.

## Theoretischer Hintergrund

Der theoretische Ausgangspunkt für diese Herangehensweise bildet die von Thibodeau<sup>3</sup> formulierten drei Ebenen digitaler Objekte: die physische (Spuren auf einem Datenträger), die logische (eine von einer Software interpretierte und dargestellte Programmlogik) und die konzeptuelle (eine vom Menschen als sinnvoll erkannte Einheit, z.B. ein Buch). Das WissGrid-Projekt<sup>4</sup> definierte korrespondierend dazu drei aufeinander aufbauende Arten der Langzeitarchivierung: Bitstream Preservation, Content Preservation und Data Curation. Um ein digitales Objekt verstehen und nutzen zu können, reicht Bitstream Preservation nicht aus, es müssen auch Maßnahmen der Content Preservation zur Bewahrung des logischen Objekts bzw. der technischen Nachnutzbarkeit sowie der Data Curation zur Erhaltung des konzeptuellen Objekts bzw. der intellektuellen Nachnutzbarkeit durchgeführt werden. Bei beiden spielen kontextspezifische Aspekte eine Rolle, deren Bearbeitung im Falle von Forschungsdaten eine zentrale Aufgabe disziplinspezifischer Forschungsdatenzentren ist.

---

<sup>1</sup> Vgl. Simon Hettrick, Knowledge Exchange: Research Software Sustainability. Report on a Knowledge Exchange Workshop, February 2016, S. 5,

[http://repository.iisc.ac.uk/6332/1/Research\\_Software\\_Sustainability\\_Report\\_on\\_KE\\_Workshop\\_Feb\\_2016\\_FINAL.pdf](http://repository.iisc.ac.uk/6332/1/Research_Software_Sustainability_Report_on_KE_Workshop_Feb_2016_FINAL.pdf).

<sup>2</sup> Vgl. Patrick Sahle, Simone Kronenwett: Jenseits der Daten. Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner 'Data Center for the Humanities', in: LIBREAS. Library Ideas, 23/2013, <http://libreas.eu/ausgabe23/09sahle/>.

<sup>3</sup> Vgl. Kenneth Thibodeau: Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years 1. Online, 2002,

<http://www.clir.org/pubs/reports/pub107/thibodeau.html>.

<sup>4</sup> Vgl. WissGrid: Generische Langzeitarchivierungsarchitektur für D-Grid. Version - 14. Januar 2010, S. 48ff.,

<http://www.wissgrid.de/publikationen/deliverables/wp3/WissGrid-D3.1-LZA-Architektur-v1.1.pdf>.

Welche konkreten Herausforderungen damit verbunden sind und wie ein praktischer Lösungsansatz für die dem Bereich der Content Preservation zuzuordnende Bewahrung von Anwendungsumgebungen aussehen kann, wird im Vortrag aus der Perspektive des DCH und HDC skizziert.

### **LAZARUS-Projekt des DCH**

Das DCH wird in diesem Kontext die Ergebnisse des LAZARUS-Projektes ("Langfristige Aufbereitung, Zugänglichkeit und Archivierung universitärer Sammlungen") näher vorstellen, welches von Dezember 2013 bis Januar 2015 zusammen mit mehreren Projektpartnern am Kölner Datenzentrum für die Geisteswissenschaften durchgeführt wurde.<sup>5</sup>

Im Fokus des Pilotprojektes stand der "Relaunch" bzw. die "Wiederbelebung" dreier unterschiedlicher Forschungsressourcen, die für die gewünschte Wieder- bzw. Weiternutzung durch das DCH exemplarisch in technischer, methodischer und konzeptioneller Hinsicht neu aufbereitet, nachhaltig gesichert und schließlich dauerhaft bereitgestellt wurden.<sup>6</sup>

Im Vortrag berichtet und reflektiert das DCH über den Aufwand, den es bedeutet, Projekte regelmäßig aufzubereiten, damit die Anwendung benutzbar und für eine inhaltliche Aktualisierung offen bleibt. In der Diskussion um die Erhaltung von Anwendungen wird vorausgesetzt, dass diese Vorgehensweise besonders schwer zu institutionalisieren ist, da sie einen mutmaßlich unkalkulierbaren Einsatz von Arbeitsstunden und finanzielle Aufwendungen generiert und eine breite Aufstellung der technischen Kompetenzen in einem Zentrum erfordert.

Die Erfahrungen des LAZARUS-Projektes des DCH verdeutlichen, dass die Zugänglichkeit von Forschungsanwendungen bisweilen nur sehr aufwändig wiederhergestellt werden kann, teilweise kann mit Berechtigung von Datenarchäologie gesprochen werden. Es liegt auf der Hand, dass bereits von Beginn eines Vorhabens die perspektivische langfristige Aufbewahrung, Präsentation und Nachnutzung berücksichtigt werden muss, um den Aufwand zu minimieren. Dies ist aus verschiedenen Gründen jedoch nicht der Regelfall in Forschungsprojekten. Dennoch muss diskutiert werden, ob ein Forschungsdatenzentrum nicht auch Angebote für Szenarien einer nicht optimal vorbereiteten Verstetigung und unzureichenden Dokumentation unterbreiten sollte, wenn das Interesse an der Aufbewahrung und Präsentation der Anwendungsumgebungen besonders hoch ist.

### **HDC-Anwendungskonservierung**

Die HDC-Anwendungskonservierung adressiert diesen Präsentations- und Referenzierungsaspekt, indem sie vollständige Anwendungsumgebungen für einen Zeitraum bereitstellt. Unter Anwendungsumgebungen sind komplexe, mehrschichtige Datenstrukturen wie Datenbankanwendungen, Digitale Editionen, Virtuelle Forschungsumgebungen oder Datenvisualisierungen zu verstehen. Bei der Konservierung werden sie in einen geschützten und abgeschirmten Bereich überführt, in dem sie weiterhin sicher betrieben werden können. Durch eine Zugriffsverwaltung kann genau gesteuert werden, wie auf die Umgebung zugegriffen wird. Dadurch werden mit der Überalterung von Software verbundene Risiken minimiert. Bei der Übergabe werden die Forschungsdaten oder -präsentationen mitsamt ihrer spezifischen Konfigurationsparameter, ergänzender Softwarekomponenten und einer Client-Umgebung, bspw. einem Browser, übernommen.

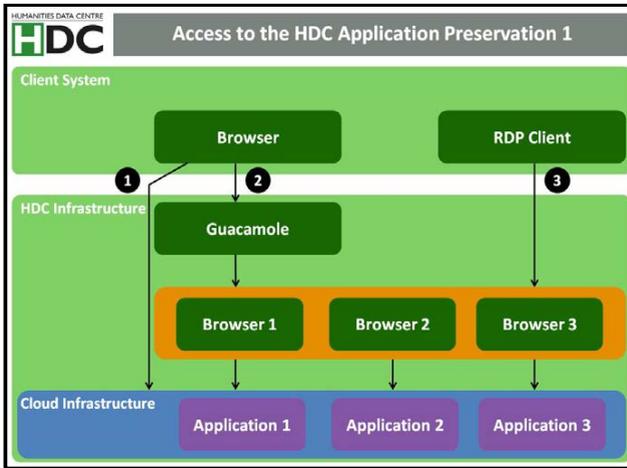
Die abgeschirmte Umgebung wird über eine Authentifizierungsschranke zugänglich gemacht und bietet dann eine langfristige Nutzung. Ein weiterer Mehrwert ergibt sich durch die Referenzierung und Zitierbarkeit

---

<sup>5</sup> Weitere Informationen zum LAZARUS-Projekt vgl. <http://dch.phil-fak.uni-koeln.de/lazarus.html>.

<sup>6</sup> Bei den drei Teilprojekten handelt es sich (1) um das Meister-Eckhart-Archiv, (2) um das Cologne Sanskrit Lexicon (CSL) sowie (3) um Paläographie-Online. Während das Meister-Eckhart-Archiv (1) ein physisches Archiv ist, zu dem bislang nur gering strukturierte textliche Beschreibungen vorliegen, besteht das Sanskrit-Lexikon (2) bereits aus digitalen Volltexten mit einem in die Jahre gekommenen Benutzer-Interface, und Paläographie-Online (3) aus einer Sammlung von didaktischen Materialien, die ebenfalls in ihrer Benutzbarkeit überarbeitet wurden.

bestimmter Systemzustände mittels Fragment-PIDs.<sup>7</sup> Hierdurch lassen sich gezielt Zustände innerhalb von Anwendungen referenzieren.



### Schlussfolgerung

Die Erfahrungen aus dem LAZARUS-Projekt des DCH und die Anwendungskonservierung des HDC beschreiben jenseits einer theoretischen Diskussion pragmatische Herangehensweisen von geisteswissenschaftlichen Forschungsdatenzentren. Die zuletzt vorgestellte Anwendungskonservierung ist zwar keine Langzeitarchivierung, aber eine Möglichkeit, digitale Forschung über einen begrenzten Zeitraum nach Projektende zu präsentieren und referenzieren zu können. Damit stellt sie u.E. eine attraktive Angebotskomponente eines Datenzentrums dar und kann dazu beitragen, insbesondere methodische Innovationen zu fördern, da sie Forschung in einem Gesamtkontext darstellt und nicht nur dateibasierte Primärdaten oder finale Forschungsergebnisse.

### Literaturauswahl

Bingert, Sven/Buddenbohm, Stefan/Kurzawe, Daniel: Referencing of complex software environments as representations of research data, Elpub Conference 2016, DOI: 10.3233/978-1-61499-649-1-58

Hettrick, Simon/Knowledge Exchange: Research Software Sustainability. Report on a Knowledge Exchange Workshop, February 2016,

<http://www.knowledge-exchange.info/event/software-sustainability>

Sahle, Patrick/Kronenwett, Simone: Jenseits der Daten. Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner 'Data Center for the Humanities', in: LIBREAS. Library Ideas, 23/2013, <http://libreas.eu/ausgabe23/09sahle/>

Thibodeau, Kenneth: Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years<sup>1</sup>, online 2002, <http://www.clir.org/pubs/reports/pub107/thibodeau.html>

WissGrid: Generische Langzeitarchivierungsarchitektur für D-Grid, Version - 14. Januar 2010, <http://www.wissgrid.de/publikationen/deliverables/wp3/WissGrid-D3.1-LZA-Architektur-v1.1.pdf>

<sup>7</sup> Vgl. Vortrag von Sven Bingert, Stefan Buddenbohm, Daniel Kurzawe: Referencing of Complex Software Environments as Representations of search Data, Elpub Conference 2016, DOI 10.3233/978-1-61499-649-1-58.

# Elemente eines Geschäftsmodells für etablierte wissenschaftliche Software

*Hanna Hedeland, hzsk, Universität Hamburg*

*Thomas Schmidt, Institut für Deutsche Sprache, Mannheim*

*Kai Wörner, gwin, Universität Hamburg*

## Einleitung

Zur FORGE-Tagung im vergangenen Jahr hatten wir erste Überlegungen zu einem Geschäftsmodell für das Softwarepaket EXMARaLDA (<http://www.exmaralda.org>, Schmidt/Wörner 2014) angestellt, das bislang inklusive vielfältiger Support-Leistungen der wissenschaftlichen Gemeinschaft kostenlos zur Verfügung gestellt wurde. Wachsende Nutzerzahlen einerseits und eine deutliche Verringerung der personellen Kapazitäten andererseits haben dazu geführt, dass das Entwicklerteam den steigenden Anforderungen an Pflege, Weiterentwicklung und Support nicht mehr im Rahmen der bisherigen Modalitäten gerecht werden kann. An die Stelle der bisher rein auf Grundausrüstung und eigenen Drittmitteln basierenden Finanzierung soll daher künftig ein Geschäftsmodell treten, das Nutzer der Software in transparenter Weise an den anfallenden Kosten beteiligt. Wir sind zu dem Schluss gekommen (Schmidt/Wörner 2015), dass ein solches Geschäftsmodell idealerweise...

- ... weiterhin eine kostenlose, quelloffene Distribution der Software zulässt,
- ... weiterhin Entscheidungen zur Weiterentwicklung nach Kriterien der beteiligten wissenschaftlichen Communities erlaubt,
- ... einer Leitung ausreichend Raum für kurz-, mittel- und langfristige Planungen lässt,
- ... finanzielle Mittel bedarfsgerecht generiert,
- ... anteilig finanzielle Mittel für eine „Grundfinanzierung“ (insbesondere für Instandhaltungsarbeiten) generiert,
- ... erlaubt, Software-Entwickler als solche zu bezahlen, und
- ... erlaubt, Support-Mitarbeiter als solche zu bezahlen.

In der Zwischenzeit haben wir ausgehend vom Hamburger Zentrum für Sprachkorpora (HZSK, <https://corpora.uni-hamburg.de/>) begonnen, erste Bestandteile eines solchen Geschäftsmodells zu skizzieren. In unserem Beitrag stellen wir erstens diesbezügliche Grundlagen dar, und berichten zweitens über Erfahrungen, die wir bei ersten Versuchen der Umsetzung dieses Modells gewinnen konnten.

## Grundlagen

Zu den Grundlagen eines Geschäftsmodells, die wir im Folgenden darstellen, gehören neben der Definition von Leistungen und Zielgruppen auch Überlegungen zu möglichen Verrechnungsmodi.

Wie oben beschrieben soll das Softwarepaket weiterhin über ein offenes Lizenzmodell zugänglich bleiben und somit kostenlos distribuiert werden können. Aus diesem Grund können Kosten für die Anpassung und Weiterentwicklung des Systems nicht über den Vertrieb von Lizenzen, sondern ausschließlich über ergänzende Leistungen gedeckt werden. Diese ergänzenden Leistungen lassen sich in zwei Bereiche unterteilen:

Unterstützende Leistungen wie Beratung zum Einsatz der Software, Nutzersupport und Schulungen können unter dem Begriff Wissenstransfer zusammengefasst werden.

Im Bereich der Softwareentwicklung kann Nutzern angeboten werden, über eine Kostenbeteiligung bestimmte Funktionalitäten bevorzugt oder überhaupt erst entwickeln zu lassen.

Für die Umsetzung müssen Beratungsleistungen, Supportpläne und Schulungskonzepte genau definiert sowie hinsichtlich des erforderlichen personellen und finanziellen Aufwandes eingeschätzt werden. Umfangreichere Rahmenverträge, wie etwa Support- und Schulungspläne für Studierende einer Fakultät oder eines Fachbereichs, könnten dabei für eine gewisse Planungssicherheit sorgen.

Die jeweiligen Zielgruppen für die hier angesprochenen Leistungen stammen in aller Regel aus der akademischen Community. Dabei kann es sich um Privatpersonen (bspw. Studierende oder Promovierende), drittmittelgeförderte Forschungsprojekte oder akademische Institutionen in öffentlicher Trägerschaft handeln.

Bei einem Transfer von Leistungen und Vergütungen zwischen universitären oder universitätsähnlichen Einrichtungen stellt sich zunächst die Frage des institutionellen Rahmens für eine Abwicklung. Eine Leistungsverrechnung ist dabei prinzipiell sowohl innerhalb der eigenen, als auch zwischen verschiedenen akademischen Institutionen in Deutschland, der EU oder weltweit möglich. Im erstgenannten Fall besteht die Möglichkeit einer hochschulinternen Verrechnung und zwischen Institutionen im deutschen Raum können auf öffentlich-rechtlicher Grundlage sogenannte Beistandsleistungen erbracht werden. Beide Verrechnungsmodi erlauben zwar eine einfache Verrechnung, sie unterliegen aber auch entsprechend strengen Bestimmungen, und die Eignung für das Geschäftsmodell muss in dieser Hinsicht genau geprüft werden.

### **Umsetzung**

Im vergangenen Jahr gab es im Umfeld der Software mehrere Anfragen, die potentiell für eine Anwendung des Geschäftsmodells in Frage gekommen wären. Bei den Anfragenden handelte es sich größtenteils um Projekte an anderen Universitäten, die Bedarf an spezifischen Erweiterungen der Software hatten. Teilweise ging es auch um grundständigen Support für geplante Projekte oder um die Kuration vorhandener Datenbestände, die für die Verarbeitung mit der Software nutzbar gemacht werden sollten.

Im Einzelfall wäre es dabei aber sehr schwierig gewesen, die Verrechnungsmodi für den Transfer zwischen Forschungseinrichtungen anzuwenden. Die betreffenden Nutzer der Software zeigen prinzipiell Verständnis für die Notwendigkeit neuer Finanzierungskonzepte, ziehen aber vertrautere und administratorisch erprobte Verrechnungsmodi dem neuen und weitestgehend unbekanntem Konzept vor. Bislang wurden die Leistungen entweder als Werkverträge zwischen externen Drittmittelprojekten und Einzelpersonen abgerechnet, oder das Zentrum wurde als Mit Antragsteller oder Partner mit entsprechenden Personalmitteln in Drittmittelanträge aufgenommen. Auf diese Weise werden zwar immerhin finanzielle Mittel für die Arbeit an der Software generiert, wegen der unmittelbaren Zweckbindung fehlt aber weiterhin die angestrebte Flexibilität, diese Mittel z.B. anteilig auch für die Wartung der Software einsetzen zu können. Auch wenn sich die Möglichkeit, den Betrieb im Rahmen einer akademischen Institution zu führen, grundsätzlich vielversprechend zeigt, bleibt also das grundlegende Problem der Planungsunsicherheit zunächst bestehen.

Einige mit der weiteren Entwicklung des Geschäftsmodells verbundene Fragen – wie etwa die Beurteilung der (Nicht-)Wirtschaftlichkeit der Leistungen, die Chancen eines solchen Geschäftsmodells auf breite Akzeptanz in der Community, sowie die Rolle, die weitere Geldquellen wie Drittmittelförderung neben diesem Geschäftsmodell noch spielen können oder müssen – möchten wir bei der FORGE 2016 diskutieren.

### **Referenzen**

Schmidt, Thomas & Wörner, Kai (2014): EXMARaLDA. In: Jacques Durand, Ulrike Gut, and Gjert Kristoffersen (eds.): The Oxford Handbook of Corpus Phonology. Oxford: OUP 2014, pp. 402-419.

Schmidt, Thomas & Wörner, Kai (2015): Überlegungen zu einem Betriebsmodell für etablierte wissenschaftliche Software. Beitrag zur FORGE 2015. [<https://www.gwiss.uni-hamburg.de/gwin/ueber-uns/forge2015/forge2015abstracts.pdf>] und [<https://lecture2go.uni-hamburg.de/veranstaltungen/-/v/18298>]

# AGATE – European Academies Internet Gateway: Konzept für eine digitale Infrastruktur für die geistes- und sozialwissen- schaftlichen Forschungsvorhaben der europäischen Wissenschaftsakademien

*Dominik Adrian, Union der deutschen Akademien der Wissenschaften*

*Carolyn Ott, Union der deutschen Akademien der Wissenschaften*

*Ulrike Wuttke, Union der deutschen Akademien der Wissenschaften*

## Zusammenfassung

Ziel des von der Union der deutschen Akademien der Wissenschaften (Akademienunion)<sup>1</sup> koordinierten Projekts „Aufbau eines europäischen Akademienportals“ (Laufzeit Oktober 2015-März 2017)<sup>2</sup> ist das Ausloten des inhaltlichen, organisatorischen und technischen Rahmens für ein europäisches Akademienportal für die Geistes- und Sozialwissenschaften (European Academies Internet Gateway, kurz AGATE).<sup>3</sup> Im Rahmen dieser digitalen Infrastruktur sollen Informationen zu den geistes- und sozialwissenschaftlichen Forschungsaktivitäten an den europäischen Wissenschaftsakademien gebündelt zur Verfügung gestellt und die digitalen Forschungsergebnisse und -daten besser auffindbar und zugänglich gemacht werden. Darüber hinaus sollen Informationen zu nachhaltigen digitalen Forschungs- und Publikationspraktiken bereitgestellt bzw. auf bereits bestehende Informations- und Serviceangebote verwiesen werden. Im Vortrag werden Details der geplanten digitalen Infrastruktur vorgestellt und Herausforderungen und Lösungsszenarien skizziert.

## Abstract

Wissenschaftsakademien gibt es überall in Europa.<sup>4</sup> Sie unterscheiden sich jedoch stark in ihrer Größe und Ausrichtung. Während manche Akademien ausschließlich als Gelehrtenesellschaften agieren, sind viele Akademien wichtige nationale Forschungszentren, gerade im Bereich der Geistes- und Sozialwissenschaften (SSH). Trotz ihrer großen Bedeutung für die jeweilige nationale Wissenschaftslandschaft und ihrer langen Tradition, die sich insbesondere in der Langfristigkeit ihrer Forschungsvorhaben niederschlägt, sind Informationen über die an den europäischen Akademien durchgeführten geistes- und sozialwissenschaftlichen Forschungsprojekte in vielen Fällen schwer online auffindbar. Deshalb wurden in der SASSH-Umfrage (Survey and Analysis of Basic Social Science and Humanities Research at the Science Academies and Related Research Organisations of Europe, 2013-2015) erstmals über 600 Forschungsvorhaben an europäischen Akademien und ähnlichen Forschungsinstitutionen systematisch zu verschiedenen Themengebieten befragt.

In der Abschlusspublikation der SASSH-Umfrage werden verschiedene Maßnahmen identifiziert, die in ihrer Gesamtheit zur Stärkung der geistes- und sozialwissenschaftlichen Forschung an den europäischen Wissenschaftsakademien beitragen würden.<sup>5</sup> Insbesondere zeigten sich bislang ungenutzte Potentiale in den

---

<sup>1</sup> Die Union der deutschen Akademien der Wissenschaften (<http://www.akademienunion.de/>) ist die Dachorganisation von acht deutschen Wissenschaftsakademien.

<sup>2</sup> <http://www.akademienunion.de/arbeitsgruppen/kooperationsprojekte/agate/>.

<sup>3</sup> Das Projekt wird vom Bundesministerium für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 01UG1503 gefördert.

<sup>4</sup> Mehr als 50 europäische Akademien sind Mitglieder von ALLEA (ALL European Academies, [www.allea.org](http://www.allea.org)), dem Zusammenschluss der europäischen Akademien der Wissenschaften, mit dem das Projekt kooperiert.

<sup>5</sup> Die Ergebnisse der SASSH-Studie sind dargelegt in: Camilla Leathem, Dominik Adrian: "Bestandsaufnahme und Analyse geistes- und sozialwissenschaftlicher Grundlagenforschung an den europäischen Wissenschaftsakademien und

Teilbereichen Kooperationen und Erfahrungsaustausch, digitale Infrastrukturen sowie digitale Methoden und Publikationspraktiken. Aus der Studie ging hervor, dass Kooperationen bzw. der Erfahrungsaustausch mit Forschungsvorhaben an anderen Akademien oftmals an mangelnden Informationen über potentielle Partnernvorhaben scheitern. Auch zeigte sich ein Nachholbedarf bezüglich des Wissensstandes über den Auftrag, die Angebote und Kooperationsmöglichkeiten mit europäischen SSH-Infrastrukturen (wie zum Beispiel CLARIN, DARIAH und Europeana). Des Weiteren wurde festgestellt, dass die geistes- und sozialwissenschaftliche Forschung der Akademienvorhaben im Internet und für die breitere Öffentlichkeit kaum sichtbar ist, wobei insbesondere eine stärkere Umsetzung von Prinzipien wie Open Access und Open Data die Verbreitung, Sichtbarkeit und Nachnutzung der digitalen Forschungsergebnisse erhöhen würde. Gerade in diesen Bereichen würde eine generell bessere Zusammenarbeit mit europäischen Infrastrukturanbietern und -initiativen beziehungsweise die verstärkte Nutzung ihrer Angebote zu einer Win-Win Situation für alle Beteiligten führen. Den benannten Herausforderungen soll mit einer paneuropäischen digitalen Infrastruktur für die geistes- und sozialwissenschaftliche Forschung begegnet werden (AGATE).

Zum jetzigen Stand der Planung wird sich die Plattform aus zwei grundlegenden Komponenten zusammensetzen, die nach und nach mit zunehmend differenzierten Funktionen ausgestattet werden können: eine Projektdatenbank und ein Service and Information Hub, die die drei oben beschriebenen Potentiale besser nutzbar machen sollen.

Die Projektdatenbank soll detaillierte Informationen über die Forschungsaktivitäten der an Akademien angesiedelten Projekte aufnehmen können und idealerweise die dazugehörigen verfügbaren digitalen Ressourcen aufzeigen, damit sie über eine gemeinsame Einstiegsseite bzw. Suchmaschine recherchiert werden können. Die Datenbank sollte es nicht nur erlauben, Projekte klassisch fachwissenschaftlich zu kategorisieren (wie Forschungsgegenstand, Epoche, etc.) sondern sie auch digitalen Methoden und Formaten zuzuordnen, um den Wissensaustausch in diesen Bereichen zu befördern. Neben der europaweiten Bündelung von Informationen zu Forschungsvorhaben böte sich durch die Datenbank zudem erstmalig die Möglichkeit einer Vernetzung der dezentralen SSH-Ressourcen der europäischen Wissenschaftsakademien. Dies führt zu einer gesteigerten allgemeinen Sichtbarkeit der Fülle der geistes- und sozialwissenschaftlichen Akademienforschung und der Beförderung des interdisziplinären fachlichen Austausches über Sprachgrenzen hinweg sowie zu einer besseren Auffindbarkeit und Zugänglichkeit der digitalen Ressourcen.

Die geplante digitale Infrastruktur für die geistes- und sozialwissenschaftliche Akademienforschung (AGATE) würde einen gemeinsamen Rahmen schaffen, um den besonderen Anforderungen der Forschungsvorhaben der europäischen Akademien in der Grundlagenforschung vor dem Hintergrund der zunehmenden Digitalisierung zu begegnen. Angesichts der langen Laufzeit vieler Akademienvorhaben sowie der langen Halbwertszeit der Forschungsergebnisse spielt die Gewährleistung der Nachhaltigkeit sowohl bei der Wahl der digitalen Methoden als auch den digitalen Publikationsformen eine zentrale Rolle und muss bei der Konzipierung neuer Vorhaben, bei der Digitalisierung laufender Vorhaben und bei der Sicherung abgeschlossener Vorhaben berücksichtigt werden.

Um die Nachhaltigkeit der digitalen Forschung durch wissenschaftlichen Erfahrungsaustausch und Kooperationen zwischen den Einzelakademien und darüber hinaus zu befördern, ist als zweite AGATE-Komponente ein sogenannter Information and Service Hub angedacht. In diesem sollen Informationen zu relevanten europäischen Infrastrukturen und ihren Angeboten bereitgestellt werden, aber auch redaktionell betreute Informationen zu Spezialthemen wie Open Access oder Forschungsdatenmanagement. Ein weiterer Schwerpunkt des Service und Information Hub soll die Förderung des Erfahrungsaustausches über digitale Tools und Methoden sein sowie die Unterstützung bei virtueller Projektarbeit. Im Vortrag werden Details der geplanten digitalen Infrastruktur vorgestellt und Herausforderungen und Lösungsszenarien skizziert.

---

ähnlichen Forschungseinrichtungen", Union der deutschen Akademien, 2015:  
[https://edoc.bbaw.de/files/1902/2015Projektpublikation\\_SASSH\\_deutsch\\_A1b.pdf](https://edoc.bbaw.de/files/1902/2015Projektpublikation_SASSH_deutsch_A1b.pdf)

# Consulting Workflow for Humanities Research Data

*Brigitte Mathiak, Data Center for the Humanities, Universität zu Köln*

*Jonathan Blumtritt, Data Center for the Humanities, Universität zu Köln*

In this contribution we outline a strategy for offering central support for research data preservation at our Faculty for Humanities. Our findings suggest that research data management in the Humanities cannot be solely reduced to preserving data sets, but that the main challenges revolve around the support and maintenance of dedicated applications.

Although we are not able to offer solutions for every problem presented, we approach the increasing number of enquiries by offering consulting to the faculty members and developing a structured and documented workflow to handle them. This talk will present the cornerstones of this strategy and exemplify them with real life use cases.

## Introduction

Many funding agencies demand the long-term preservation of research data, yet, for the Humanities this opens some fundamental questions. What is research data? And what does preservation mean, in this context? We found that these questions cannot be answered in isolation and the answers depend heavily on the methodology in question.

On one end of the spectrum there is the “classical” research data, experimental data, in many disciplines of the Humanities, such as Archeology, Linguistics and Educational Science. How this data should be handled is somewhat understood. The current paradigm is to separate both metadata and data and store both in common standards. A plethora of options for storage on both national and international level have been established in recent years, so that the problem

The other end of the spectrum are living systems which stand between software, research and social community. Imagine trying to preserve an active discussion board. While it is perfectly possible to dump all the text to XML, add some metadata and put it in long-term storage, the usefulness of the forum to the fellow scholars has immediately dropped to zero. Although this may be an extreme case, it exemplifies how a meaningful utilization of particular resources can be bound to a specific environment. We found that living systems are an everyday reality at our institution and that they play a prominent role in digital research in the humanities.

## Consulting procedure

To identify the problems and tackle them in the most efficient way, we have instituted a system to offer consultation to all members of the Faculty. Faculty come to us in one of three stages. Ideally, they are in the planning phase of the project. This has the advantage that we can still influence strategic decisions ranging from technology choices to early contacts with potential data archives. Many come while the project is still running, usually towards the end, but we also have a lot of consultees that come after the project has finished.

We have developed a semi-structured questionnaire to cover all the aspects that might be relevant to the data archiving. This ranges from general information about the project, type of research data and metadata, data quality, software, and technology. Additionally, we ask for the expectations for the data preservation, both ideal and realistic. This includes information on both real and targeted users of the work products. Usually, one member of the crew does the interview and then proceeds to produce a project description.

The project descriptions are discussed at the group meetings. While we start from the premise that all research data from the Faculty is valuable and worth preserving, expectations often have to be adjusted. For

“classical” research data, we try to find the best-fitting archives and contact them on behalf of our Faculty members. Less conventional work products, like databases, web applications and other software, typically require some sort of technical support and funding.

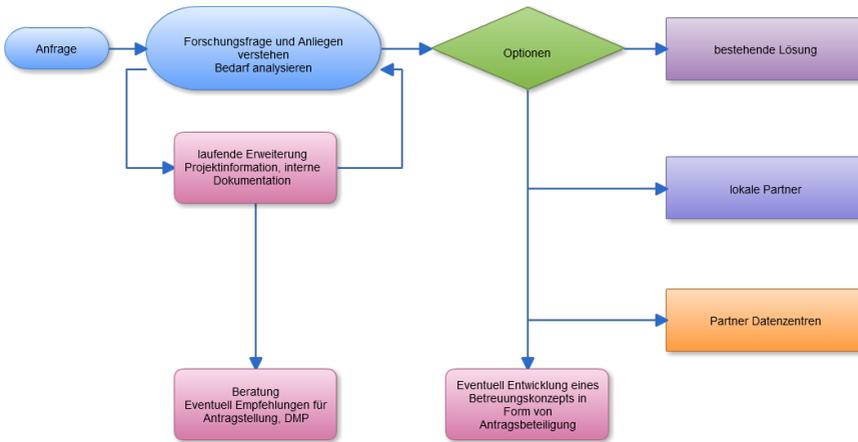


Fig. 1: A flow chart illustrating the consultation workflow

Our priority system for solutions works in the following way.

1. Are there synergies with our own projects and systems?
2. Do we have local partners, e.g. at the computer center or library that could be helpful?  
Often a re-framing of the problem, or a shift in the technology stack can transform an unmaintainable system into something that is routinely maintained by a local partner.
3. Are there external partners, e.g. archives or projects that could be helpful?  
There is an increasing number of projects and institutions out there, trying to help scholars with preserving their data and may be better equipped to deal with specific requirements.
4. How can we develop and maintain new systems that solve the problems at hand?  
We do not introduce new systems lightly, in particular, when new technologies are involved.

### Outline for our talk

In this talk, we will present the cornerstones of our program: our questionnaire to gather crucial information about a project to evaluate viable options, a typology of enquiries and a schema how to handle them. We will exemplify our strategies with real life use cases.

# Digitale Nachhaltigkeit von Forschungsanwendungen durch Mikroservices

*Michael Haft, Digitale Akademie, Akademie der Wissenschaften und der Literatur, Mainz*

*Anna Neovesky, Digitale Akademie, Akademie der Wissenschaften und der Literatur, Mainz*

*Gabriel Reimers, TU Berlin*

Nachnutzung bereits vorhandener Komponenten, Schnittstellen und Infrastrukturen stellt einen wesentlichen Aspekt moderner und nachhaltiger Software-Entwicklung in den Digitalen Geisteswissenschaften dar. Durch Generalisierung und Modularisierung können optimale Voraussetzungen für die Nachnutzbarkeit von Neuentwicklungen geschaffen werden. Der softwarearchitektonische Ansatz der „Microservices“ bedeutet eine Aufteilung in einzelne Komponenten, die jeweils eine konkrete Aufgabe übernehmen und auch einzeln voll funktional sind und erweitert werden können. Das Konzept von Microservices sowie deren Einsatz in der Anwendungsentwicklung wird anhand einer musikwissenschaftlichen Rechercheplattform erläutert und die Möglichkeiten hinsichtlich der Nachhaltigkeit digitaler Forschungsapplikationen dargestellt.

Die Konzeption von generischen Softwarekomponenten und deren Bezug auf die geisteswissenschaftliche Fragestellungen und Anwendungen ist ein Kerngebiet der Digitalen Geisteswissenschaften. Einen wesentlichen Aspekt moderner und nachhaltiger Software-Entwicklung stellt neben Neuentwicklungen auch die Nachnutzung bereits vorhandener Komponenten, Schnittstellen und Infrastrukturen dar. Dies erlaubt nicht nur eine effizientere Entwicklung, sondern ermöglicht es auch externen Beteiligten auf bestehende Kenntnisse und Schnittstellen aufzubauen.

Die Verwendung verbreiteter Standards für die Datenhaltung und für die Programmierschnittstellen (API) erlaubt es Daten und Dienste im Sinne einer offenen Wissenschaft zu nutzen und weiterzuführen. Hierfür sind jedoch nicht nur die lizenzrechtlichen Rahmenbedingungen und die dokumentierte Bereitstellung nötig, sondern auch eine Generalisierung der spezifisch für ein Forschungsvorhaben entwickelten Anwendungen.

Dies kann am besten erreicht werden indem einzelne Funktionalitäten herauslöst werden, die dann einzeln genutzt werden können und somit auch in einem breiteren Kontext anwendbar werden. Dieser Ansatz der Softwarearchitektur wird als „Microservices“ bezeichnet. Grundsätzlich ist das Konzept des „Teilen und Herrschens“ in der Softwareentwicklung nicht neu. Der leichte, flexible und günstige Zugang zu Server- und Netzinfrastruktur erlaubt es jedoch heute in praktisch jedem Projekt verteilte Systeme zu entwerfen, die aus mehreren kleinen, unabhängigen Diensten bestehen, welche untereinander über Standard-Protokolle wie REST kommunizieren. Der Vorteil dieser Mikro-Dienste ist, dass Sie einen geringen (Code-) Umfang haben und damit für Andere leichter nachvollziehbar und änderbar sind. Bei einer Nachnutzung der Anwendung können die jeweils benötigten Komponenten einzeln verwendet und um Dienste für spezifische Anforderungen ergänzt werden.

Der Einsatz von Microservices in der praktischen Entwicklungs- und Forschungsarbeit wird durch ein Beispiel veranschaulicht. Ausgangspunkt für die Entwicklung der vorgestellten Anwendung war der Bedarf zur Erweiterung bestehender Recherchemöglichkeiten einer musikwissenschaftlichen Plattform. Zusätzlich zur Suche über Metadaten sollten die Stücke über ihre charakteristischen Melodien (Incipits) auffindbar gemacht werden. Die Incipits werden mit dem Kodierungsstandards zur Musikannotation Plaine & Eaise ausgezeichnet.

Die Anwendung nutzt Open Source Technologien nach und steht selbst unter der MIT Lizenz auf github zur Verfügung (<https://github.com/annaneo/incipitSearch>). Die Funktionalitäten sind in mehrere Microservices unterteilt, die das Abgreifen der Daten aus dem Repositorium über eine definierbare Schnittstelle, das Erstellen des Suchindexes und Einbindung einer Suchmaschine, einer Nutzerschnittstelle mit Klaviatur für die Noteneingabe sowie die Rückgabe der Ergebnisse realisieren. Die einzelnen Komponenten der Anwendung können in Kombination für ähnliche Repositorien genutzt werden oder einzeln in andere Projekte eingebunden werden. Schnittstellen können entsprechend konfiguriert, neue benötigte Funktionen ergänzt werden. Entsprechend ist der Einsatz nicht nur auf musikwissenschaftliche Inhalte beschränkt.

Diese generische Anwendbarkeit führt nicht nur zu breiteren Einsatzmöglichkeiten sondern erhöht auch die Zahl der Entwickler, die mit der Software arbeiten und sie weiterentwickeln. Dadurch kann eine langfristige Nutzbarkeit leichter gewährleistet werden und der entstandene Code nachhaltig in anderen Forschungsapplikationen eingesetzt werden.

# Von implizitem Wissen zu nachhaltigen Systemanforderungen – Interdisziplinäre Zugänge in der frühen Anforderungsanalyse am Beispiel der Digital Humanities

*Dr. Bianca Meise*

*Yevgen Mexin*

*Franziska Schloots*

*Björn Senft*

*Anastasia Wawilow*

*Prof. Dr. Gregor Engels*

*Prof. Dr. Dorothee Meister*

*Prof. Dr. Jörg Müller-Lietzkow*

*Prof. Dr. Gerd Szwillus*

In den Digital Humanities wird die Softwareentwicklung vor besondere Herausforderungen gestellt, da hochspezialisierte Software für wissenschaftliche Forschung entwickelt wird. Damit sind die Nutzer sehr spezifisch und deren Expertenwissen häufig implizit. Komplexitätssteigernd wirkt, dass sich die Forschung mit digitalen Daten und im Digitalen von der traditionellen Forschungsarbeit unterscheidet und stetig weiterentwickelt. Diese Problematiken müssen von allen beteiligten Forschungsdisziplinen in DH-Kontexten in Grundzügen erfasst werden, um Forschungsanteile bearbeiten zu können. Der Beitrag diskutiert diese Themen beispielhaft anhand der interdisziplinären Projektarbeit der Medienforschung und der Informatik zur Weiterentwicklung der Software für digitale Musikeditionen und somit der Verarbeitung von textuellen und nichttextuellen Daten. Hierbei steht die Methode zur Erforschung impliziten Wissens und deren Transfer in die Anforderungen zur Entwicklung der Softwaremodelle im Mittelpunkt der Diskussion.

Um Software nachhaltig zu entwickeln müssen vielfältige Prozesse durchlaufen und systematisch miteinander verbunden werden. Dennoch gestaltet sich Softwareentwicklung nicht reibungslos. So sind bei der Softwareentwicklung in Unternehmen immer noch unvollständige, verborgene und sich verändernde Anforderungen die am häufigsten genannten Probleme (vgl. Fernández / Wagner 2014, The Standish Group 1995). Innerhalb der Digital Humanities (DH) sind die Unwägbarkeiten noch dramatischer, denn die Herausforderungen sind immanent im Gegenstand verankert. Das resultiert zum einen aus dem innovativen Charakter der zu entwickelnden Lösungen wie es Burdick et al. (2012) charakterisieren:

„Digital Humanities work embraces iterative, in which experiments are run over time and become object to constant revision. Critical design discourse is moving away from a strict problem-solving approach that seeks to find a final answer: Each new design opens up new problems and – productively – creates new questions. (ebd.: S. 22)“

Durch die Tatsache, dass Forscher auf ihrem Gebiet durch Software unterstützt werden sollen, ist zum anderen eine enge Verzahnung von Technologien und geisteswissenschaftlichen Forschungsmethoden und Inhalten erforderlich. Hierfür ist sowohl Expertenwissen auf Seiten der Forschung in den Geisteswissenschaften als auch auf der technologischen Seite notwendig. Allerdings bedeutet Expertenwissen, dass es sich nicht nur um explizites Wissen, sondern vielfach um implizites Wissen handelt, welches nicht ohne Weiteres zugänglich ist. Insofern sind Strategien notwendig, um insbesondere das implizite Wissen einschließlich der habitualisierten Arbeitsweisen offen zu legen (vgl. Gervasi et al. 2013). Zudem gilt es zu bedenken, dass Methoden des Software Engineerings und des Usability Engineerings auf die Erstellung eines Produkts ausgerichtet sind. Dies fokussieren beide Disziplinen unter der Prämisse, dass die Nutzer klar formulieren können wie ein Produkt aussehen soll. Allerdings führen Frohlich und Sarvas (2011) an, dass Technologie sowohl durch kulturelle Einflüsse auf die Entwickler, als auch durch die potentiellen Nutzer

und deren kreativen Ansätze der Handhabung, geformt wird. Die Identifizierung und Interpretation solcher sozialen Konstrukte, wie auch die Offenlegung impliziten Wissens, sind Prozesse, die zumeist durch sozialwissenschaftliche Verfahren erforscht werden können. Diese haben aber in den meisten Fällen keinen Bezug zu Fachwissen in der Informatik. Somit sind vielfältige diskursive Prozesse notwendig, um in dieser Situation aus Nutzerbedürfnissen Anforderungen generieren zu können.

Das vom BMBF geförderte Digital Humanities Projekt „Zentrum Musik - Edition - Medien“ erfährt eben diese Herausforderungen und entwickelt hierfür Lösungen in einem interdisziplinären Kontext. Im Vordergrund steht u. a. die Unterstützung von Musikwissenschaftlern bei ihrer Arbeit als Editoren mit der Forschungssoftware Edirom<sup>1</sup>. Diese erlaubt es Faksimiles, Digitalisate und digitale Daten von Notentexten oder anderen Quellen zu speichern, zu organisieren, zu kollationieren, zu annotieren und zu analysieren. Musikwissenschaft, Software Engineering, Usability Engineering, quantitative und qualitative empirische Medienforschung forschen hier gemeinsam an der Erarbeitung neuer Modelle und Werkzeuge für die Softwareentwicklung. An dieser werden wir im Folgenden unser Vorgehen und die damit verbundenen Herausforderungen exemplarisch verdeutlichen. Darüber hinaus besitzt die Problemstellung weitreichendere Relevanz, da grundsätzliche Fragestellungen der Unsicherheiten in der Softwareentwicklung bearbeitet werden. Zudem legen wir dar, wie die Bereiche durch Synthese der fachdisziplinären Zugänge, Methoden und Ergebnisse diese Problemstellungen im Rahmen der frühen Anforderungsanalyse konzipieren und umsetzen. So verschafft die quantitative Forschung einen Überblick über den gesamten potentiellen Nutzerkreis, die qualitative Forschung analysiert das implizite Wissen des individuellen Nutzers, das Software Engineering erarbeitet durch die Erhebung von stabilen Anforderungen neue Modelle für die Softwareentwicklung und das Usability Engineering optimiert Oberflächengestaltung und darunterliegende Softwarearchitektur. Erst die Trennung der Lösung (Software) vom Problem erlaubt es, systematisch eine funktionierende Lösung zu identifizieren. Die wissenschaftliche Aufarbeitung des Problemkontexts durch die quantitative und qualitative Medienforschung führt zu einer genaueren Beschreibung und Verständnisses des Problemkontexts, wodurch gezielter nachhaltige Anforderungen abgeleitet werden können, die sich auf stabilere Faktoren stützen als ein lösungsbasiertes Vorgehen. Die Verbindung der Perspektiven wird in Anlehnung an die Grounded Theory Methodologie (vgl. Strauss/Corbin 1996; Strübing 2004) systematisch bearbeitet, um so die Kollaboration und Synthese zwischen den Disziplinen zu strukturieren.

## Literatur

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., & Schnapp, J. (2012). DIGITAL HUMANITIES. MIT Press.

Fernandez, D. M., & Wagner, S. (2014). Naming the Pain in Requirements Engineering: A Design for a Global Family of Surveys and First Results from Germany. *Information and Software Technology* (57).

Frohlich, D. M. & Sarvas, R. (2011). HCI and Innovation in CHI'11 Extended Abstracts on Human Factors in Computing Systems. ACM.

Gervasi, V., Gacitua, R., Rouncefield, M., Sawyer, P., Kof, L., Ma, L., Piwek, P., de Roeck, A., Willis, A., Yang, H. & Nuseibeh, B. (2013). Unpacking Tacit Knowledge for Requirements Engineering. In Maalej, W. & Thurimella A. K. (Hrsg.), *Managing Requirements Knowledge*. Springer-Verlag.

The Standish Group (1995). *Chaos'94*. Technical report.

Strauss, A. & Corbin, J. (1996). *Grounded Theory. Grundlagen qualitativer Sozialforschung*. Weinheim, Beltz.

Strübing J. (2004). *Grounded Theory. Zur theoretischen und epistemologischen Fundierung des Verfahrens der empirisch begründeten Theoriebildung*. Wiesbaden, VS.

---

<sup>1</sup> <http://www.edirom.de>

# Contemporary challenges and future perspectives for sustainable reuse of archived textual qualitative research data<sup>1</sup>

*Reinhard Messerschmidt, Cologne Center for eHumanities, Universität zu Köln*

*Brigitte Mathiak, Cologne Center for eHumanities, Universität zu Köln/ GESIS – Leibniz-Institut für Sozialwissenschaften*

## Summary

Reuse of research data is one of the key arguments of establishing an archiving landscape. One of the things that make qualitative research data so interesting for digital humanities is that, in theory, it allows to conduct large scale research using its coding frame and annotations. In practice, however, there are still various hurdles towards that goal, which we outline in the following text. The purpose of this contribution is to raise a discussion about potential strategies to overcome present challenges and consequently to increase sustainability with respect to data and QDA software.

When it comes to sustainable research data management, archiving constitutes a parallel process to the entire life cycle of data and should become part of the research methodology (Smioski, 2013; Jacobs and Humphrey, 2004). Extensive efforts made by the recently emerged qualitative data archive landscape provide a fundamental premise. However, there are still challenges on this path. On the one hand, the variety of QDA software and its different formats raises compatibility and therefore future accessibility issues due to current deficits of import and export capabilities. On the other hand, essential parts of the raw data as well as the analysis are frequently not easily accessible, often due to privacy concerns and anonymization issues. While in some cases adequate anonymization is impossible without losing essential information, it always means more time consuming work either for the researchers or the archive's staff. Furthermore, datasets containing legally protected material require a more restricted distribution. Moreover, some communities are reluctant to share data due to a specific conception of qualitative research as a practiced art (Broom et al., 2009). They also feel that the data is not very useful for others (Kuula, 2009) as it is impossible to achieve the ideals of rigorous and self-reflective research with archived data (Parry and Mauthner, 2004). Despite all skepticism, a large number of national and international archives have been established (Huschka et al., 2013), containing thousands of documents. The same issues that drive the skepticism, make these data particularly interesting for secondary analyses and methodological research, if insights into the data generation process are granted.

Problems of secondary analysis have been discussed in the qualitative archiving community already more than ten years ago (Corti et al., 2005). Within the digital humanities landscape, Bischoff et al. (2014) have identified three general challenges for secondary analysis in context of possible support for data access and exploration: accessibility, findability, and interpretability. Digital availability is as important as the ability to find the appropriate data for a given purpose. In the heterogeneous archive landscape, the latter is not always an easy task. With regard to interpretability, context information is crucial, e.g. for interview interpretation.

Our recent pilot study of the international archive landscape showed a low quota of datasets containing code systems and annotations of the data. Although we managed to find few annotated corpora from various archives and could additionally use personal contacts to researchers for constructing a first test corpus, low case numbers limit the current possibilities for research and fundamentally challenge availability and accessibility. In the recently emerging German speaking qualitative data archive landscape, currently there are mostly datasets without code systems available, typically transcripts of interviews. The situation is similar in other European as well as in the English speaking countries. For example, the Australian Data Archive currently hosts only five mostly older qualitative studies and authors need to be contacted for getting access. Others archives are still under construction. In the US, there is an additional challenge for non-US researchers

---

<sup>1</sup> English language is used due to the international scope of the topic.

due to the necessity of an IRB approval. Although the search term “qualitative” leads to several hundred studies in some of the archives, many results only contain links to specific university archives or contacting authors is required. Currently, UK Data Service hosts the promising amount of over 900 qualitative or mixed methods datasets. Unfortunately, there is (generally in all archives) no possibility for queries based on file types. Consequently, there was no alternative to manually checking all entries. This time consuming activity lead finally to only 4 datasets stored in NVivo format. One had no code system and only contained the transcripts, so the effective result was much less than we expected. Such obstacles might also be related to the different stages of data availability in the recently emerging landscape of qualitative data archives respectively integrating such data into existing infrastructure. While some projects are still work-in-progress, other archives formerly established for quantitative data only host few datasets or in case of larger amounts only raw data lacking coding frames.

To an outsider it could seem irritating that the most interesting part of research data is routinely missing, like a critical edition without the critique. But it becomes much clearer when observing the individual incentivization systems. Researchers have nothing to gain by sending in annotations. If the data is required as a supplement for a journal, methodological debates surrounding the actual coding are easily avoided by not adding annotations. The latter also makes the texts more widely readable and is much less work. Archivists similarly might have no incentive to insist on annotations. The most common reason for retrieving a dataset from the archive is teaching qualitative methods (Bishop, 2012) for which the annotations are not always required. How to deal with this challenging situation in order to provide sustainability in future qualitative data archiving?

## Bibliography

Bishop, L. “Using archived qualitative data for teaching: practical and ethical considerations.” *International journal of social research methodology: theory & practice*, Vol. 15 (2012) Nr. 4, 341-351.

Bischoff, K. , Niederée, C. , Tran, N. , Zerr, S. , Birke, B. , Brückweh, K. and Wiede, W. “Exploring Qualitative Data for Secondary Analysis: Challenges, Methods, and Technologies.” *Digital Humanities DH2014* (2014), 436-438.

Corti, L.; Witzel, A. & Bishop, L. (2005) “On the Potentials and Problems of Secondary Analysis. An Introduction to the FQS Special Issue on Secondary Analysis of Qualitative Data.” *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(1), Art. 49, <http://nbn-resolving.de/urn:nbn:de:0114-fqs0501495>.

Kuula, A. “Methodological and ethical dilemmas of archiving qualitative data.” *IASSIST Quarterly* 34.3/4 (2010): 35.

Broom, A., Cheshire, L. & Emmison, M. (2009) “Qualitative researchers understandings of their practice and the implications for data archiving and sharing”, *Sociology*, 43(6), 1163-1180.

Jacobs, J. A. & Humphrey, C. K. (2004). Preserving research data. *Communications of the ACM*, 47(9), 27-29.

Huschka, D. (ed.). “Forschungsinfrastrukturen für die qualitative Sozialforschung.” Berlin: Scivero Verlag, 2013.

Parry, O. and Mauthner, N. (2004) “Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data.” *Sociology*, 38(1), 139-152.

Smioski, A. (2013). Archivierungsstrategien für qualitative Daten [54 Absätze]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 14(3), Art. 5, <http://nbn-resolving.de/urn:nbn:de:0114-fqs130350>.

# Ist das Forschung oder kann das weg? Methoden zur Reduktion und Restrukturierung von Forschungsdaten und -anwendungen

*Iris Vogel, gwin, Universität Hamburg*

Die Kuration von Forschungsdaten abgeschlossener Projekte ist ein Teil des Service-Angebotes des gwin-Projekts der Fakultät für Geisteswissenschaften an der Universität Hamburg. Ziel der Kuration ist in den meisten Fällen, die Daten und Funktionalität aus abgelaufenen Forschungsprojekten für einen weiteren Lebenszyklus zu sichern. Meist erfolgt im Zuge dessen eine Migration in die Forschungsdateninfrastruktur des gwin-Projektes. Neben der physischen Übertragung der Daten auf ein anderes Medium erfordert die Überführung meist nicht nur eine Format-Konversion sondern auch eine Re-Strukturierung der Daten.

Im einfachsten Fall werden Daten und Funktionalität auf die Bedürfnisse des Zielsystems angepasst. Setzt man eine erfolgreiche Konversion und Migration damit gleich, dass der intellektuelle Inhalt bzw. Wert der Objekte erhalten bleibt, ist das Ziel in einer idealen Forschungsdatenwelt eine verlustfreie Konvertierung ohne störende Artefakte. Die Durchführung und Qualitätssicherung ist in diesem Fall weitgehend automatisierbar.

In der praktischen Arbeit muss man jedoch feststellen, dass die Realität von Projektabläufen, die jeweiligen technischen und personellen Gegebenheiten und Veränderungen, unweigerlich Spuren in den Daten und Anwendungen hinterlassen haben. So treten in der Regel bei der Aufbereitung der Daten Inkonsistenzen in Erscheinung, die durch die Anwendung mehr oder weniger geschickt maskiert wurden. Striktere Validierungsregeln im Zielformat oder -system können eine Restrukturierung der Vokabulare und Referenzen erfordern.

Hinzu kommt, dass sich durch die zunehmende Digitalisierung in den letzten Jahrzehnten auch im Bereich der Forschungsdaten nicht nur der Anspruch an die Strukturierung der Daten sondern auch an deren Präsentation gewandelt hat. Durch die Open Access Initiativen im öffentlichen Bereich wird nun zusätzlich der Anspruch abgeleitet, nicht nur Forschungsergebnisse in Form von Publikationen, sondern auch die Daten selbst in Form von Applikationen oder digitalen Sammlungen der allgemeinen Öffentlichkeit zur Verfügung zu stellen. Diese Öffnung der Daten und Applikationen für ein breiteres Publikum involviert auch ein Umdenken im Bereich der Zugänglichkeit.

Der Beitrag widmet sich den Fragen, was die Kuration von Forschungsdaten und -anwendungen im Bereich Digital Humanities leisten kann beziehungsweise wo sie ihre Grenzen hat. Es werden Möglichkeiten und Probleme aus der Kurationspraxis der letzten Jahre aufgezeigt. Dabei geht es neben der Reorganisation von Daten und Überarbeitung der Funktionalität auch um die Frage der potenziellen Reduktion und nach dem Umgang mit Forschungsdaten, die sich im Laufe der Migration als potenziell fehlerbehaftet oder unvollständig entpuppt haben.

# Wir machen dann mal Digital Humanities Software

## Ein Erfahrungsbericht

*Julia Damerow, Arizona State University*

*Dirk Wintergrün, Max-Planck-Institut für Wissenschaftsgeschichte*

Die Entwicklung von Softwareanwendungen für die Auswertung, Erstellung und Verwaltung von Forschungsdaten in den Geisteswissenschaften rückt mit den „Digital Humanities“ zunehmend in das Zentrum der Diskussion nicht nur in den Textwissenschaften, wo dies eine vergleichsweise lange Tradition hat, sondern auch in die den historisch arbeitenden Wissenschaften. Wir arbeiten an zwei Forschungseinrichtungen, dem Max-Planck-Institut für Wissenschaftsgeschichte (MPIWG)<sup>1</sup> in Berlin und der Digital Innovation Group (DigiNG)<sup>2</sup> im Center for Biology and Society, Arizona State University. Die unterschiedliche institutionelle Verankerung in einem reinen Forschungsinstitut auf der einen Seite, angebunden an universitäre Forschung und Lehre auf der anderen Seite führen zu unterschiedlichen jedoch komplementären Ansätzen zur Sicherung der Langzeitstabilität vorgenommener Entwicklungen. Unser Vortrag ist ein Erfahrungsbericht, wie diese Ansätze aussehen, welche Herausforderungen sie stellen, und für wie erfolgreich wir sie einschätzen.

Wir werden uns auf die Darstellung der Arbeitsformen und ihrer Herausforderungen konzentrieren. Bezüglich Auswirkungen auf die Infrastruktur sei auf Überlegungen zur Sicherung von Daten im Forschungsdatenkreislauf hingewiesen wie sie beispielsweise im Rahmen von DARIAH und CLARIN angestellt werden (Puhl et. al, 2015). Im Kontext von DARIAH-DE wurde ein beispielhafter Forschungsdatenkreislauf entwickelt, der sich an den „scholarly primitives“ von John Unsworth orientiert (Unsworth 2000, Leganovic 2015) (siehe Abb. 1). Die Projekte der beteiligten Institutionen decken fast alle Aspekte dieses Kreislaufs ab. Im Folgenden stellen wir eine Auswahl unserer Projekte kurz vor.

Eines der ersten digitalen Projekte des MPIWGs war die Veröffentlichung des Manuskripts MS72 von Galileo Galilei, welches seit 1999 im Netz verfügbar ist.<sup>3</sup> Den damaligen technischen Möglichkeiten gemäß ist die Präsentation ein statisches HTML, welches jedoch seit fast 20 Jahren stabil verfügbar ist. Das Archimedes Projekt<sup>4</sup> hingegen stellt seit den frühen 2000er Jahren Texte der Mechanik der frühen Neuzeit durch Sprachtools unterstützt ins Netz. Neben Texteditionen ist eine weitere Herausforderung, strukturierte Daten<sup>5</sup> zusammen mit den Quellen, auf denen diese basieren, verfügbar zu machen. Projekte, wie die Cuneiform-Digital-Library und das Virtuelle Laboratorium für Physiologie zeigen exemplarisch, wie eine solche Datenpublikation erfolgen kann.<sup>6</sup> Aus diesen Einzelerfahrungen erwuchs 2001 das Projekt „European Cultural Heritage Online“ (ECHO). Hier wurden Tools und Quellen von der Anthropologie bis hin zur Wissenschaftsgeschichte zusammengefasst und Online publiziert. Für Veröffentlichung von Forschungsergebnissen in Form von virtuellen Ausstellungen wurde 2008 Virtual Spaces entwickelt. In dieser Anwendung werden semantische Modelle von virtuellen Ausstellungen erstellt, aus denen dann statische HTML Seiten generiert werden. Quadriga hingegen ist eine Anwendung zur Verwaltung von „Quadruple“-Annotationen. Quadruple sind kontextualisierte Tripel, die Quadriga in Graphdatenbanken speichert.

---

<sup>1</sup> Siehe: <http://www.mpiwg-berlin.mpg.de/>

<sup>2</sup> Siehe: <http://diging.asu.edu/>

<sup>3</sup> Siehe: <http://www.imss.fi.it/ms72/INDEX.HTM>

<sup>4</sup> Siehe: <http://archimedes.fas.harvard.edu/>

<sup>5</sup> Beispielsweise in Form von Datenbanken, Spreadsheets oder XML-Dateien.

<sup>6</sup> Zum Einstieg in diese Projekte siehe die Links auf <http://www.mpiwg-berlin.mpg.de/en/resources/websites>

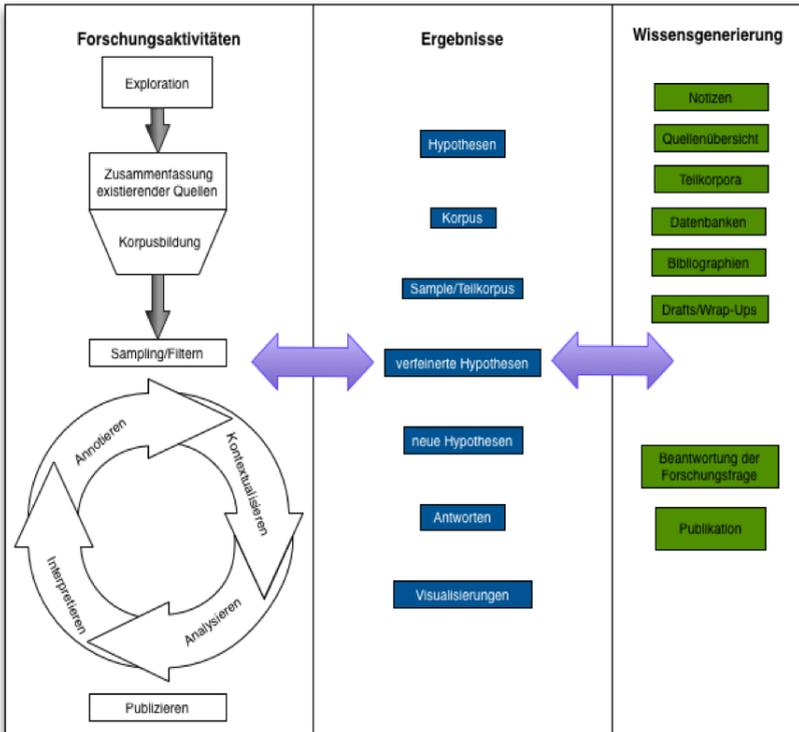


Abb. 1: Forschungskreislauf

Gemeinsames Ziel aller Projekte ist es durch die Datenaggregation und -erzeugung eine Wissensbasis bereitzustellen, die es erlaubt neue Verbindungen innerhalb von digitalisierten Quellen zu finden und wissenschaftliche Kollaboration zu unterstützen. Darüberhinaus zielen unsere Projekte darauf ab durch Wiederverwendung von Daten, Anwendungen und Infrastruktur deren Stabilität zu erreichen. Die grundlegende Idee ist, dass die Nutzung von existierenden Services deren Langfristigkeit fördert und die offene Bereitstellung von Daten deren Reproduktion ermöglicht. Während manche unserer Projekte dies bereits vollständig umsetzen, ist das nicht bei allen der Fall. Ohne Forschungsprojekte mit konkreten Anforderungen an eine Anwendung, sind jedoch häufig nicht genügend Ressourcen für die Generalisierung und Weiterentwicklung verfügbar. Unser Vortrag wird dies exemplarisch am Beispiel Virtual Spaces diskutieren.

In Bezug auf Nachhaltigkeit gehen MPIWG und DigiNG verschiedene Wege, die sich gut ergänzen. Als Ergebnis der ersten Erfahrungen mit digitalen Projekten am MPIWG wurde bereits im Jahr 2000 eine wissenschaftliche IT Gruppe mit drei Mitarbeitern eingerichtet, die digitale geisteswissenschaftliche Arbeit unterstützen sollte. Zu diesem Zeitpunkt bestand die IT eines geisteswissenschaftlichen Institutes in der Regel aus 1-2 Personen, die für das Netzwerk und die Betreuung der PCs am Arbeitsplatz zuständig waren. Die IT-Gruppe des MPIWGs war von Anfang an in die Forschungsgruppen am Institut eingebunden und unterstützte Projekte mit Softwareentwicklungsbedarf. Der zentrale Charakter der IT-Gruppe schafft ideale Voraussetzungen für die Entwicklung und Instandhaltung projektübergreifender Anwendungen.

Im Gegensatz dazu setzt DigiNG größtenteils auf den Beitrag von Studenten. Geleitet wird DigiNG von einem Team aus Wissenschaftshistorikern und Softwareentwicklern, das eine Gruppe von studentischen

Hilfskräften koordiniert, die die Datenerhebung, Datenauswertung, und Softwareentwicklung unterstützen. Darüberhinaus wurde ein Kurs zum Thema Softwareentwicklung in den Digital Humanities angeboten, in welchem Informatikstudenten an DigiInG Projekten mitarbeiten. Die auf Studenten gestützte Entwicklung ist zwar langsamer als die Arbeit mit vollausgebildeten Softwareentwicklern, hat jedoch Vorteile. Anstelle von einem Softwareentwickler können beispielsweise mehrere studentische Hilfskräfte angestellt werden, auf die die Softwareentwicklung verteilt wird. Dadurch wird das Risiko gesenkt, dass die Umsetzung einer Anwendung an einem einzigen Entwickler hängt. An beiden Institutionen wird darüberhinaus größtenteils Open Source Software entwickelt; in unseren Augen eine unbedingte Voraussetzung für nachhaltige Softwareentwicklung.

Das MPIWG und DigiInG setzen bezüglich Nachhaltigkeit auf die Trennung von Datenspeicherung und -nutzung. Während sich zur Verfügung stehende Technologien stetig weiterentwickeln und Anwendungen zur Analyse und Visualisierung schnell Gefahr laufen zu veralten, basieren gerade im "Big Data Zeitalter" immer mehr Projekte auf der digitalen Verfügbarkeit von Datensätzen. Eine konsequente Bereitstellung von Daten unabhängig der Anwendungen durch die sie erstellt oder analysiert wurden, ist daher von zentraler Bedeutung für nachhaltige und langlebige Daten- und Anwendungsentwicklung.

## Referenzen

Leganovic, Oona, Viola Schmitt, Juliane Stiller, Klaus Thoden und Dirk Wintergrün. „Anforderungen Und Bedürfnisse von Geisteswissenschaftlern an einen Digital Gestützten Forschungsprozess.“ In Book of Abstracts - DHD, 2015.

Puhl, Johanna, Peter Andorfer, Mareike Höckendorff, Stefan Schmunk, Juliane Stiller und Klaus Thoden. „Diskussion Und Definition Eines Research Data LifeCycle Für Die Digitalen Geisteswissenschaften“ (2015). <http://resolver.sub.uni-goettingen.de/purl/?dariah-2015-4> (Zuletzt besucht: 22. April 2016)

Unsworth, John. „Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This?“ (2000) <http://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html> (Zuletzt besucht: 22. April 2016).

# Eine interdisziplinäre Bestandsaufnahme

Alexander Struck, Humboldt-Universität zu Berlin

## Motivation

Der Autor betreibt Plattformen für eine im Rahmen der Exzellenzinitiative geforderte Einrichtung und berät Forschende in der Softwareentwicklung und Langzeitverfügbarkeit der entstehenden Forschungsdaten und Infrastruktur (Produkte). Es besteht eine Kooperation mit der lokalen Universitätsbibliothek im Rahmen neuer Publikationsformen, mit dem Rechenzentrum für Infrastrukturdienste & Informationsaustausch und den Sammlungsbeauftragten im Kontext von Nachhaltigkeit.

## Problemstellung

### 1. Bewusstsein

Bei in Forschungsprojekten vertretenen Naturwissenschaften hat sich bereits teilweise ein Bewusstsein für Forschungsdaten entwickelt; auch der Tatsache geschuldet, dass für Publikationen bei ausgewählten Verlagen die Daten bereits zum Zeitpunkt des Reviews in einem Repository verfügbar sein müssen. Die Geisteswissenschaften verneinen überwiegend das Vorhandensein von Daten und die vertretenen Gestaltungsdisziplinen nähern sich erst langsam dem Konzept einer Publikation im Sinne einer Forschungsleistung. Obwohl der Exzellenzcluster u. a. zur Verstärkung der Forschungsprojekte aufgerufen ist, fehlt es bei den Forschenden und in den Entscheidungsgremien an Kompetenz und Bewusstsein für Daten-, Software- und Plattformmanagement.

### 2. Recht

Rechtsfragen stellen alle Beteiligten vor Probleme. Forschende sehen teilweise keine Notwendigkeit, die "ihnen gehörenden" Daten zu veröffentlichen und haben kein Verständnis für Nachvollziehbarkeit von Ergebnissen im Sinne guter wissenschaftlicher Praxis. Anderen fehlen rechtliche Grundlagen, um die Möglichkeit der Veröffentlichung abzuschätzen. Es werden im Universitätsrepository bereits veröffentlichte Daten gelöscht, weil die Forschenden als Ansprechpartner nicht (mehr) zur Verfügung stehen (können). In solchen Umgebungen sind Nachhaltigkeitszusagen schwer zu treffen. Es herrscht zudem Unklarheit, zu welchen Bedingungen Software, die im Forschungsumfeld entsteht, lizenziert werden soll oder kann und wie die Daten anderer nutzbar sind.

### 3. Management

Die mit interdisziplinärem Forschungsauftrag versehenen Projekte sollen Software entwickeln, die auch plattformbasierte innovative (Nach)nutzung von Daten vorsieht. In vielen bekannten Fällen weist das (Software)Projektmanagement Lücken auf, die entweder Daten schwer nutzbar zurücklassen oder Plattformen aufgrund ihrer Architektur kaum langfristig betrieben werden können. In Projektanträgen ist nicht vorgesehen, technische Aspekte wie erforderliche Infrastruktur, Datenmanagement oder Entwicklerkapazität anzugeben. Aus Kostengründen wird nicht die vorhandene Infrastruktur (DB- und Webserver) genutzt, sondern eine weitere VM als langfristig zu pflegende Umgebung erzeugt. Obwohl Anwendungsentwicklung als Dienstleistung bereits in Anspruch genommen wird ist das bei externem Hosting nicht der Fall. Es sind am RZ keine Ressourcen vorgesehen, Forschungsinfrastruktur in Form von VMs zu pflegen. Wenn die Forschenden keine Updates einspielen, wird die Plattform abgeschaltet. In für Sicherheitslücken bekannten Sprachen geschriebene Module sind Teil des Problems. Forschungsförderer verlangen die Veröffentlichung von Daten und Software nicht. Zu Plattformen fehlt es an Empfehlungen. Die fehlenden Mittel für die zumindest zehnjährige Bereitstellung sind nicht bei der Vergabe vorgesehen. Es existieren für Forschungsdaten vorgesehene Repositories (teilweise leider ohne Persistent Identifier). Dort

können jedoch nur Dumps und keine in Plattformen eingebettete Daten hinterlegt werden. Software Repositories sind nicht breit verfügbar und werden nicht im Kontext von FDM gesehen.

#### **4. Datenschutz**

Personenbezogene Forschungsdaten und Softwareprodukte werden bei kommerziellen Cloud Storage Anbietern gehostet. Projektmanagementplattformen werden ebenfalls bei kommerziellen Anbieter subskribiert (obwohl lokal vorhanden). Dort werden auch Forschungsmittel investiert, um z. B. den notwendigen Speicherplatzbedarf oder Nutzeranzahl einzukaufen. In Ermangelung von Ressourcen am Rechenzentrum sind Forschungseinrichtungen gezwungen, eigenständig solche Plattformen langfristig zu betreiben, damit Nutzung eben dieser innerhalb der akademischen Einrichtung überhaupt erst möglich wird. Wegen möglicher Überwachung durch den Arbeitgeber darf das am RZ vorhandene Software Repository nicht produktiv ausgerollt und verpflichtend gemacht werden.

### **Diskussion möglicher Best practice Ansätze**

#### **1. Bewusstsein**

Studierende müssen an Einführungen in die Universitätsbibliotheken teilnehmen, was erfolgreich die Werkzeuge und Ressourcen vermittelt. Ähnlich können die Forschungsfoederer sowohl das administrative Management wie auch die Forschenden der geförderten Einrichtungen zur Teilnahme an Weiterbildungsveranstaltungen zu Forschungsdatenmanagement verpflichten. Es ist zu klären, inwieweit die Freiheit der Forschung eingeschränkt werden kann. Die Zitierung als Wahrung der Forschungsreputation wird auch für Daten und Plattformen etabliert und Angaben wie von <http://rightsstatements.org/en/> verpflichtend gemacht.

#### **2. Recht**

Forschungsfoederer kooperieren mit Heimateinrichtungen und koordinieren für geförderte Projekte die notwendige rechtliche Unterstützung. In Projekten werden aus der Einrichtung Verantwortliche für die langfristige Verfügbarkeit von Daten und Software einbezogen.

#### **3. Management**

Bei geförderter Entwicklung neuer Plattformen ist im Antrag auf einer Marktanalyse vorhandener Werkzeuge zu bestehen und ggfs. die Konsultation durch geeignete Marktbeobachter verpflichtend zu machen. So sollte auch erreicht werden,

- dass existierende Standards Berücksichtigung finden. (Bsp: <https://www.w3.org/TR/annotation-model/>)
- kontrollierte Vokabulare für die Auszeichnung benutzt werden (für bessere Nachnutzung aus Aggregation)
- existierende Werkzeuge nachgenutzt werden
- eine Weiterförderung vom Erreichen relevanter Zertifikate abhängig gemacht werden kann z. B. <https://wiki.dnb.de/display/DINIAGKIM/Zertifikat+Digitale+Sammlungen>
- eine realistische Abschätzung zur Community-Bildung für langfristige Pflege von Softwareplattformen vorliegt, wonach Geld für Community Developer veranschlagt wird

Für den langfristigen Betrieb von VMs oder Plattformen sollten am Rechenzentrum Ansprechpartner für sichere Softwareentwicklung und langfristiges Hosting zur Verfügung stehen. Das muss auch die Kapselung von Anwendungen in VMs oder Containern beinhalten. Es sind langfristige Strategien und Kompetenzen für die Bereitstellung von Plattformen entweder in VMs oder Containern zu entwickeln. Derzeit verspricht Containerisierung leichtgewichtiges ressourcenschonendes Handling, ist jedoch in Bezug auf Langfristigkeit

und Sicherheit noch zu evaluieren. Insbesondere die Zitierbarkeit z. B. ueber DataCite muss sichergestellt werden. Plattformen werden in relevanten Zuständen, z. B. Docker Image oder VM Snapshot, gesichert und versioniert zitierfähig gehostet.

<https://www.software.ac.uk/blog/2016-03-29-reproducible-research-citing-your-execution-environment-using-docker-and-doi>

#### **4. Datenschutz**

In Anträgen ist das Vorhandensein oder Fehlen lokaler Infrastruktur zu beschreiben und ggfls Alternativen fuer Projektmanagement, Cloud Storage, Datentransfer etc listen.

# Datenauforschung als Option für nachhaltige Datennutzungen

*Katrin Moeller, Historisches Datenzentrum Sachsen-Anhalt, MLU Halle-Wittenberg*

Die Bedeutung von Forschungsdatenmanagement ist mittlerweile umfänglich in der Wissenschaftskultur angekommen. Die Vorteile der Open Data Sciences überzeugen, auch wenn die jüngst vom Rat für Informationsinfrastrukturen (RfII) pointiert formulierten Herausforderungen nach wie vor bestehen. Sie betreffen neben vielen anderen Arbeitsfeldern vor allem die Entwicklung einer neuen „Forschungsdatenkultur“.

Die neue Wissenschaftspraxis ist insbesondere den Denkansätzen des Open Access und den Regeln guter wissenschaftlicher Praxis verpflichtet, die einerseits einen ungehinderten Zugang und eine effiziente Nachnutzung wissenschaftlicher Inhalte fordern, andererseits kooperative Forschungsstrategien und transparente Partizipation betont. Open Access-Initiativen stärken daher die eigenen Rechte und Pflichten von Autoren und fördern über die leichte Erreichbarkeit und Zitierbarkeit den Reputationsgewinn von Wissenschaftler/innen. In diesem Sinne etablierten vor allem die STM-Fächer (Science, Technology, Medicine) ein analoges Konzept der Open Access Data. Analog zur wissenschaftlichen Textpublikationen sollen Daten künftig publiziert und veröffentlicht werden. Andere Autoren können diese Daten anschließend herunterladen und je nach Lizenzierungsmodell mehr oder weniger frei nachnutzen und in eigene Versuchsreihen integrieren. Indem sie die nachgenutzten Datensätze zitieren, weisen sie die Herkunft der Daten nach und mehren die Reputation des Datengebers, deren Anwachsen über verschiedene Formen von Open Metrics Verfahren in Zitiersystemen nachzuvollziehen ist. Das Verfahren besticht letztlich durch die Einfachheit, weil es auf etablierte Standards aufsetzt und mit relativ wenig Aufwand in der Forschungspraxis eingeführt werden kann.

Allerdings bringen diese konzeptionell angedachten Formen des Reputationsgewinns vor allem in den Geistes- und Kulturwissenschaften keine praktisch messbaren Ergebnisse hervor. Die Gründe dafür sind vielfältig und ich möchte sie an dieser Stelle nur kurz nennen:

1. Die Datengewinnung in den Geisteswissenschaften ist kostenintensiv, zeitaufwändig und weit überwiegend eine eigenständige wissenschaftliche Leistung mit ausreichender Schöpfungshöhe.
2. Bisher hat die Geisteswissenschaft kein wissenschaftliches Reputationssystem entwickelt, das die Messbarkeit der Forschung über Zitationsraten, Impact Factor(en) oder anderen Evaluationsmetriken ermöglicht.
3. In den Geisteswissenschaften dominieren stattdessen Publikationsformen der Monografie und des Sammelbandes. Eine Zweit- oder Co-Autorenschaft wie in den STM-Fächern ist weitgehend unüblich.
4. Das Prestige von Publikationen auf Open Access Plattformen und in Datenrepositorien besitzt in den Geisteswissenschaften bisher nicht den gleichen Stellenwert wie Publikationen in renommierten Buchreihen, Zeitschriften und Verlagen.

Die Datenpublikation bringt dem Autor daher über die eigentliche Publikation der Daten keinen messbaren wissenschaftlichen Mehrwert. Der Aufwand zur Aufbereitung durch Dokumentation, Metadatenproduktion, Beschreibung und Qualitätssicherung dürfte daher den Nutzen für den Autor häufig deutlich übersteigen.

Auch in den STM-Fächern bringt das bereits eingespielte Modell der Co-Autorenschaft Problemlagen mit sich. Diskutiert wird über die Platzierung von Autoren und die Berechtigung zur Autorenschaft, die in den Regeln "Guter wissenschaftlicher Praxis" definiert wurden. Mehr Transparenz darüber, welchen Anteil einzelne Forschende an Publikationen innehalten, leisten die darin gemachten Vorschläge noch nicht. Grundlegendes Prinzip dieser Formen der Autorenschaft ist eine aktive Beteiligung an dem entstehenden Forschungsergebnis. Das Prinzip der Open Data entkoppelt diesen unmittelbaren Zusammenhang. Daten können auch weit nach ihrer Entstehung in neue Forschungsergebnisse einmünden. Entscheidender Nachteil

dieser Praxis bleibt, dass diese Altdaten zwar im Rahmen der neuen Veröffentlichung vielleicht alle Kriterien für eine Autorenschaft der Urheber erfüllen, diese Urheber jedoch unter Umständen keinen Einfluss mehr auf die Publikation nehmen wollen oder auch mit den daraus abgeleiteten Forschungsergebnissen nicht einverstanden sein müssen.

Daher ist für solche Nachnutzungen ein Zitat vorgesehen. Allerdings ist diese Form rechtlich gesehen an verschiedene Bedingungen gebunden, die meines Erachtens nach nicht mit den Nachnutzungswünschen vollständig kompatibel sind.

Das Grundprinzip des Zitats ist die Belegfunktion. Ein Zitat ist nach dem Urheberrechtsgesetz dann zugelassen, wenn es eigene Ideen oder Gedanken unterstützt bzw. Ideen anderer aufgreift und in den eigenen Text integriert. Das Urheberrecht regelt im § 52, dass in der wissenschaftlichen Forschung "kleine Teile eines Werkes, Werke geringen Umfangs sowie einzelne Beiträge aus Zeitungen oder Zeitschriften" für die wissenschaftliche Forschung genutzt werden können. Auch wenn es kein festgesetztes Limit für den Umfang eines Zitats gibt, gilt als Faustregel maximal etwa ein Drittel eines Werkes. Die Zitation ganzer Datensätze ist mit dem üblichen Kleinzitat eher nicht vereinbar. Auch die Verwendung von Großzitationen - also unter Umständen auch die Zitation eines ganzen Werkes - bestimmen eine strikte logische Verknüpfung von Zweck und Umfang. Die Zitation ist also ungeeignet, um ganze Datensammlungen vollständig für weiterführende Forschungen zu verwenden.

Eine Lösung für diese komplexe Problemlage würde die Trennung von Text- und Datenautorenschaft erbringen. Ein bereits existierendes Beispiel für eine solche Trennung liegt bei der Urheberschaft durch eine Übersetzungsarbeit mit Schöpfungshöhe vor. Diese Übersetzungen werden regelmäßig von der eigentlichen Textautorenschaft separiert ausgewiesen. Dieses Verfahren schafft eine erhebliche Transparenz über die Formen der Urheberschaft und der Zumessung wissenschaftlicher Reputation. Textautoren sollten daher künftig getrennt von den Datenautoren ausgewiesen werden, wobei erstere die Verantwortung für die Publikation des aktuellen Forschungsergebnisses in Textform tragen, alle anderen jedoch Anteil an der Schaffung der Datengrundlage besitzen. Die rechtlichen Voraussetzungen dafür liegen mit dem aktuellen Urheberrechtsgesetz bereits vor.

# Aufbau und Hosting nachhaltiger Softwarelösungen durch die Verbundzentrale

*Kathleen Neumann, Verbundzentrale des GBV (VZG)*

Im Rahmen dieses Vortrages werden die Erfahrungen zum Aufbau und Hosting von Repository-Anwendungen aus der VZG vorgestellt. Danach werden die unterschiedlichen Anforderungen und Gemeinsamkeiten von Repository und Forschungsdatensystem und deren Umsetzung diskutiert. Anschließend wird ein Konzept vorgestellt, wie eine langfristige Datenerhaltung durch einen neuen Dienst der VZG garantiert werden kann, ohne den vollen Umfang einer Langzeitarchivierungslösung umzusetzen.

## **Aufbau und Hosting von Repository-Anwendungen**

Bereits heute können sich viele Bibliotheken den Betrieb eigener Repository-Anwendungen nicht mehr leisten - zu umfangreich ist die notwendige Infrastruktur. Hier unterstützen die Verbundzentralen in ihrer Rolle als Dienstleistungs- und Entwicklungszentrum und bieten Aufbau und Hosting von Softwarelösungen an.

Die Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG) stellt seit 2013 einen Service zum Aufbau und Hosting einer auf MyCoRe basierten Repository-Lösung bereit. Ohne spezielle Anpassungen an Datenmodell, Erfassung oder Präsentation der Daten ist dieser Dienst innerhalb weniger Tage aufgesetzt und gewährleistet trotzdem ein gewisses Maß an Individualität. Dies wird u.a. durch die Verwendung von MODS (Metadata Object Description Schema) als zugrundeliegendes Datenmodell zur Beschreibung der Metadaten erreicht. Die einzelne Anwendung nutzt nur einen Teil der Mächtigkeit des Datenmodells, die die Anwendung ermöglicht.

Des Weiteren wurde die Repository-Basisanwendung MIR (MyCoRe MODS Institutional Repository) so konzeptioniert, dass sie alle typischen Repository-Anforderungen (vgl. DINI-Zertifikat) erfüllt und dabei einen gewissen Baukasten-Charakter behält. So gibt es beispielsweise einzelne Templates für jedes Metadaten-Formularfeld, wodurch die Formulare in Reihenfolge und Inhalt je nach Bedarf zusammengesetzt werden können. Weiterhin ist es dadurch möglich verschiedene Dokumenttypen (Zeitschrift, Artikel, Hochschulschrift, usw.) zu spezifizieren und zu beschreiben.

Die Verbundzentrale baut den Dienst seit 5 Jahren auf und aus und bietet selbigen seit 3 Jahren Kunden an. Mittlerweile gibt es 7 produktive Anwendungen und für das aktuelle Jahr 2016 sind bereits 4 weitere Anwendungen geplant. Zusammenfassend kann man also sagen, dass der Bedarf eines solchen Angebotes da ist und auch genutzt wird. Schafft man es für das Forschungsdaten-management eine ähnliche Basis durch einen gemeinsamen Metadatenstandard und Bausteine für eine Standardanwendung bereitzustellen, könnte man analog dazu ein Serviceangebot aufsetzen.

## **Eine Frage der Nachhaltigkeit**

Derzeit arbeitet die VZG an einem Konzept zur langfristigen Datenerhaltung. Im Zuge dessen steigen die Anforderungen an die einzelnen Anwendungen bzw. der dort eingestellten Daten. So kann beispielsweise nur dann der Datenerhalt garantiert werden, wenn mit MD5-Summen sichergestellt wird, dass es sich jederzeit um die korrekte Datei handelt. Dies müsste somit bereits beim Einstellen der Daten in die Anwendung geschehen. Alternativ prüft der Nutzer die eingestellte Datei und die MD5-Summe wird in der Anwendung erzeugt.

Weiterhin sollte die Anwendung zumindest über die Langzeitarchivierungs-Tauglichkeit der eingespielten Daten informieren. Dies schafft die Voraussetzung um die Daten nach Bedarf an LZA-Dienstleister wie die DNB oder die TIB zu liefern. Dazu ist es weiterhin notwendig, dass Daten und Metadaten in OAIS-standardkonformen Transferpaketen aus der Anwendung exportiert werden können.

Daneben ist es natürlich zwingend notwendig die Dienste selbst nachhaltig zu gestalten. Im Falle des Reposis-Dienstes wurde daher mit MyCoRe als Softwarebasis auf eine in Deutschland weit verbreitete Open-Source-Lösung mit aktiver Community gesetzt.

# Poster und Systempräsentationen

# Kompendium der deutschsprachigen historischen Arbeits-, Berufs- und Amtsbezeichnungen

*Katrin Moeller, Leiterin des Historischen Datenzentrums Sachsen-Anhalt, Halle*

Die Berufsangabe ist in vielen Quellengruppen und Forschungsdisziplinen eine analytisch wichtige Information. Um einen vergleichbaren, interoperablen Standard zur Auswertung dieser Information zu schaffen, wurden für moderne Berufsbezeichnungen mehrere Klassifikationen geschaffen. Es werden Modelle verwendet, die (a) Tätigkeitsprofile und Branchen systematisieren, die (b) Anforderungsniveaus, Bildungs- und Qualifikationsaspekte berücksichtigen oder (c) soziales Prestige und Sozialstrukturen messen. Während für die Berufswelten des 20./21. Jahrhunderts damit internationale Standards existieren, kann die interdisziplinäre Forschung für deutschsprachige, historische Berufe nicht auf ein Normsystem zurückgreifen. Immerhin haben sich zwei internationalen Systematiken (HISCO, PST) für historische Berufsbezeichnungen etabliert. Bisher führen diese jedoch sehr wenig deutschsprachige Berufsamen. Zudem zeichnen sie sich durch eine klare Ausrichtung auf das 19./20. Jahrhundert aus und sind in ihren Methodiken weitgehend auf die sozialstatistische Analyse konzentriert.

Die Ontologie deutschsprachiger Berufe möchte diese Lücke schließen und auf Grundlage von weit über 100.000 Berufsvarianten von (normiert) ca. 20.000 Berufsamen ein Klassifikationssystem frühneuzeitlicher Berufe entwickeln. Dabei setzen wir neue Schwerpunkte.

1. Im Unterschied zu allen anderen Vorgehensweisen behält die Systematik den Bezug zur historischen Schreibvariante, die erst in einem zweiten Schritt normiert wird. Dies ermöglicht nicht nur einen breiteres wissenschaftliches Analysespektrum (bspw. für sprachwissenschaftliche oder etymologische Fragestellungen), sondern bildet eine entscheidende Grundlage für Text Modeling, Textmining und Annotation in ganz unterschiedlichen Anwendungsgebieten.
2. Entwickelt das Projekt in direkter Anknüpfung an bestehende Systematiken ein theoretisches Konzept zur Kategorisierung von Tätigkeitsprofil und Anforderungsniveau entlang der modernen Definition. Entsprechend internationalen Anforderungen wird die Systematik damit nachvollziehbar und konsistent konstruiert. Um eine adäquate historische Analyse zu erzielen und langfristig auch den zeitlichen und räumlichen Wandel von Berufsamen stärker in Analyseinstrumente einzubeziehen, nehmen wir für eine klar begrenzte Stichprobe von Berufsamen eine umfassende induktive Inhaltsanalyse von Berufsbezeichnungen vor. Dabei werden ausgewählte Berufsamen in ca. 20 frühneuzeitlichen Lexika verglichen und hinsichtlich der Elemente frühneuzeitlicher Beruflichkeit ausgewertet. Auf diese Weise möchten wir einerseits analysieren, welche Lexika genaue Tätigkeitsbeschreibungen bieten. Andererseits ermitteln wir neben den modernen Definitionskriterien (Tätigkeitsbeschreibungen, Anforderungsniveau etc.) auch weitere Elemente frühneuzeitlicher Beruflichkeit, die über Lexika erfassbar sind. Auf Grundlage des für die jeweilige Berufsgruppe am besten geeigneten Lexikon werden anschließend für 20.000 Berufsamen historische Tätigkeitsprofile erstellt. Erstmals erfolgt eine quellenspezifische Zuordnung eines historischen Berufes zu modernen Tätigkeitsfeldern, Berufsgruppen und anderen Systematisierungskriterien, die über xml/TEI-konforme Auszeichnungen webbasiert nachvollziehbar sind.
3. Die Zuweisung eines Kodierungssystems zu Berufsamen bleibt relativ unelastisch. Um dies zu umgehen, splitten wir die Berufsinformation regelgeleitet in Informationseinheiten auf. Dies ermöglicht die individuelle Auswahl verschiedener Wortfelder bzw. ihrer Kombinationen, um Systematiken flexibler zu nutzen und auszuwerten. Zudem besitzen frühneuzeitliche Berufsbezeichnungen häufig Informationen, die Systematiken bisher überwiegend nicht erfassen

(bspw. Geschlecht, Herrschaftsbeziehungen, Verwandtschaftsbeziehungen etc.), für frühneuzeitliche Berufssystematiken aber wichtige Informationen enthalten.

4. Die Berufsbezeichnungen werden sowohl mit der Klassifikation der Berufe der Agentur für Arbeit (KldB 2010) wie mit HISCO kodiert. Dies ermöglicht eine Schnittstelle zu anderen bestehenden modernen/historischen Klassifikationssystemen. Als Teil der langfristigen Infrastruktureinrichtung DARIAH soll die Ontologie der Berufe durch die Niedersächsische Staats- und Universitätsbibliothek auf Dauer als offener Webservice angeboten werden.

# Forschungsdaten in Repositorien auf der Basis von MyCoRe

*Dr. Wiebke Oeltjen, Universität Hamburg, RRZ, MyCoRe-Geschäftsstelle*

*Hinnerk Rümenapf, Universität Hamburg, RRZ*

## Zusammenfassung

Mit einem Poster und einer Systempräsentation sollen verschiedene Repositorien vorgestellt werden, die Forschungsdaten in den Geisteswissenschaften erfassen, verwalten und präsentieren. Die Repositorien basieren auf dem Software-Framework „MyCoRe“. Die Open-Source-Software wird seit 2001 von einer bundesweiten Community bereitgestellt, kontinuierlich weiterentwickelt und an neue Anforderungen angepasst.

Die Open-Source-Software MyCoRe<sup>1</sup> dient als Framework zum Erstellen von Repositorien zum Erfassen, Verwalten und Präsentieren von Forschungsdaten und ihren beschreibenden Metadaten. Sie wird bundesweit an Universitäten, in Bibliotheken und Rechenzentren sowie öffentlichen Einrichtungen eingesetzt. Bereitgestellt wird das Framework von einer Community, die MyCoRe kontinuierlich weiterentwickelt und an neue Anforderungen angepasst. MyCoRe-Anwendungen werden je nach Anforderungen und Datenmaterial von den jeweiligen Projektpartnern individuell entwickelt, angepasst und gepflegt. Dazu stellt die Community ein Anwendungsgerüst zur Verfügung, das „Skeleton“, das als Basis zum Entwickeln einer eigenen MyCoRe-Anwendung dient. Weiter steht eine Beispiel-Applikation als Softwarepaket zur Nachnutzung zur Verfügung, das „MODS Institutional Repository MIR“. Die MIR-Anwendung ist mit allen typischen Repositoriums-Funktionen ausgestattet. Dazu gehört die Erfassung von Metadaten in standardisierten Datenmodellen, wie z.B. dem „Metadata Object Description Schema“ (MODS), das von der Library of Congress (LoC) veröffentlicht wird. In MIR werden die Metadaten im MODS-Standard erfasst und gespeichert. Forschungsdaten jeder Art können in MyCoRe-Anwendungen wie MIR erfasst werden. Das können Forschungsdaten in Form von Digitalisaten sein, wie z.B. eingescannte Urkunden, Akten, Bücher, Handschriften, Kataloge, Zeitschriften, Zeitungen etc., die mittels integriertem Bildbetrachter präsentiert werden können. Ein integrierter PDF-Viewer erlaubt das Zeigen erfasster Publikationen oder digitaler Dokumente. Ebenso können Audiodateien oder Videos integriert werden. Zugriffsrechte können mittels ACL gesteuert werden. Darüber hinaus sind Datenexporte in verschiedenen Formaten möglich (z.B. BibTeX, MODS, RIS, ISI, MS Word bibliography oder Dublin Core). Schnittstellen und Protokolle, wie das OAI Protocol for Metadata Harvesting, ermöglichen den automatisierten Datenaustausch. Auch können externe Informationsressourcen, wie z. B. die „Gemeinsame Normdatei“ (GND) zur Datenvernetzung genutzt werden (Linked Open Data).

In der MyCoRe-Entwicklung werden bewährte Software-Komponenten eingesetzt, die ein Mindestmaß an Nachhaltigkeit gewährleisten. Allen MyCoRe-Anwendungen ist gemeinsam, dass sie als Webapplikationen in einer Infrastruktur betrieben werden, die laufend aktualisiert und gepflegt werden muss, um einen sicheren und verlässlichen Betrieb sicherstellen zu können.

Exemplarisch sollen MyCoRe-Anwendungen mit Forschungsdaten gezeigt werden. Neben den produktiven Webanwendungen können auch Anwendungen in einer lokalen virtuellen Maschine präsentiert werden.

---

<sup>1</sup> <http://www.mycore.de>

# CoRS-O-Matic

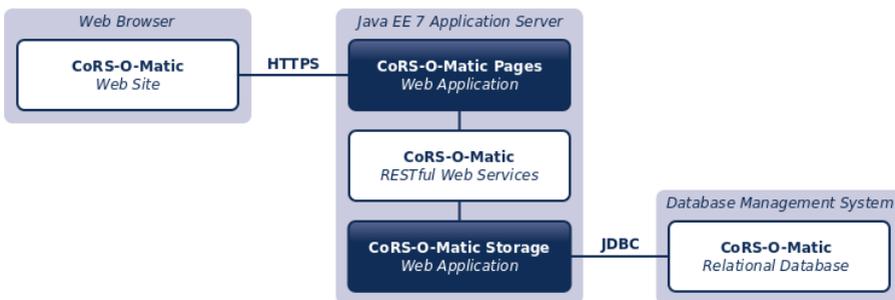
## Eine freie, webbasierte Anwendungssoftware für linguistische Forschungsdaten

Anja Weingart, Georg-August-Universität Göttingen  
Nils Cordes, binfort GmbH

Die Systemvorführung mit Poster stellt die freie, webbasierte Anwendungssoftware CoRS-O-Matic vor, deren Entwicklung die nachfolgenden Ziele verfolgte. Erstens sollen die Forschungsdaten des Projekts „Components of Romance Syntax (CoRS)“<sup>1</sup> durch CoRS-O-Matic für eine Nachnutzung zur Verfügung gestellt und eine computerunterstützte Weiterführung des Projekts ermöglicht werden. Zweitens soll CoRS-O-Matic für andere Forschungsprojekte so leicht wie möglich wiederverwendbar und anpassbar sein. Der Fokus des Posters liegt auf der komponentenbasierten Softwaresystemarchitektur von CoRS-O-Matic. Diese Architektur sowie die Einhaltung von offenen Standards und die Verwendung freier Software ermöglichen eine leichte und aufwandsarme Anpassung, wie zum Beispiel die Modifikation der Benutzeroberfläche oder die Adaption des Datenmodells an spezifische Anforderungen anderer Forschungsprojekte.

Das Projekt CoRS strebte eine minimalistische Analyse der zentralen syntaktischen Phänomene der Standardvarietäten des Spanischen, Italienischen, Rumänischen, Französischen und Europäischen Portugiesisch an (vgl. Mensching & Remberger 2011, Giurcea & Remberger 2012 und Mensching & Weingart 2016). Die Datensammlung besteht aus einzelnen Sätzen, die syntaktische Phänomene repräsentieren und unterscheidet sich von anderen Sprachkorpora insbesondere dadurch, dass auch ungrammatische Sätze aufgenommen und Sprecherurteile bezüglich der Wortstellung und der Informationsstruktur annotiert werden. Des Weiteren sind Metadaten mit funktionalen, morpho-syntaktischen, semantischen und informationsstrukturellen Merkmalen auf Wort-, Phrasen- und Satzebene zusammen mit Positionsangaben nötig, um gezielt Abfragen wie „Alle Sätze, die ein präverbales, definites Subjekt mit kontrastiven Fokus erlauben“ formulieren zu können.

Die webbasierte Anwendungssoftware CoRS-O-Matic basiert primär auf dem offenen Standard Java EE 7 und ist daher sowohl plattform- als auch anbieterunabhängig. Die Softwaresystemarchitektur von CoRS-O-Matic ist in der nachfolgenden Abbildung skizziert.



<sup>1</sup> Das Gemeinschaftsprojekt CoRS der Freien Universität Berlin (Guido Mensching und Anja Weingart) und der Universität Konstanz (Eva Remberger and Ion Giurcea) wurde von der Deutschen Forschungsgemeinschaft von 2008 bis 2011 gefördert.

Wie dargestellt, verwenden die Nutzer einen Webbrowser zur Bedienung der Anwendungssoftware, die als Website präsentiert wird. Diese Website kommuniziert mit einem Applicationserver, der die Webapplication CoRS-O-Matic Pages bereitstellt. Zur Abarbeitung von Anfragen und zur Erstellung der Antworten nutzt die Webapplication RESTful Webservices, die von der Webapplication CoRS-O-Matic Storage bereitgestellt werden. CoRS-O-Matic Storage nutzt das Protokoll JDBC, um die Daten in einem Database Management System persistent zu speichern. Hierfür kann jedes relationale Datenbankmanagementsystem, das den Standard JDBC unterstützt verwendet werden.

Kenntnisse des verwendeten Datenbankmanagementsystem sind nicht nötig, weil zur Programmierung von CoRS-O-Matic Storage überwiegend POJOs und JPA Annotationen verwendet wurden.

Die CoRS-O-Matic RESTful Webservices entkoppeln CoRS-O-Matic Storage und CoRS-O-Matic Pages so, dass diese beiden Komponenten unabhängig voneinander modifiziert oder durch alternative Implementierung ersetzt werden können.

Bei der Implementierung der CoRS-O-Matic RESTful Webservices wurde das Entwurfsmuster DTO verwendet. Dadurch ist es möglich, die gespeicherten Daten nicht nur via JAXB in den Datenformaten JSON oder XML, sondern auch in weiteren Datenformaten, wie z.B. RDF und Tex zu repräsentieren. Hierdurch wird die Möglichkeit der Wiederverwendung der Forschungsdaten vereinfacht.

Die Implementierung der Benutzerschnittstelle in CoRS-O-Matic Pages basiert auf HTML5 und folgt den Entwurfsmustern Responsive Web Design und Progressive Enhancement, was eine Verwendung auf vielen unterschiedlichen Endgeräten ermöglicht.

Die Softwaresystemarchitektur von CoRS-O-Matic stellt alle notwendigen Bausteine der webbasierten Datenspeicherung, -eingabe, -abfrage, und -darstellung zur Verfügung. Diese Bausteine können flexibel kombiniert und unabhängig voneinander modifiziert werden. Diese Eigenschaften tragen zur guten Wartbarkeit der Software bei, eine wesentliche Voraussetzung, um die nachhaltige Nutzbarkeit von CoRS-O-Matic und seiner Derivate sicherzustellen.

## **Bibliographie**

Mensching, Guido & Eva-Maria Remberger (2011): "Syntactic Variation and Change in Romance: A Minimalist Approach". In: Peter Siemund (Hrsg.): Linguistic Universals and language Variation. Berlin /New York. Mouton de Gruyter, 361–403.

Mensching, Guido & Anja Weingart (2016): "The Null Subject Parameter and the lexicon in minimalist syntax". In: Olga Fernandez, Amaya Mendikotxea, Luis Eguren (Hrsg.): Rethinking Parameters. Oxford University Press, 291-328.

Giurgea, Ion & Eva-Maria Remberger (2012): "Zur informationsstrukturellen Konfiguration und Variation postverbaler Subjekte in den romanischen Null-Subjekt-Sprachen." In: Zeitschrift für Sprachwissenschaft 31/1, 43–99.

# Lassen sich technische Plattformen bei Forschungsanwendungen konsolidieren und gewinnt man damit etwas?

*Kai Wörner, gwin, Universität Hamburg*

ForscherInnen sind in der Wahl der Methoden, die sie zur Erlangung ihrer Forschungsergebnisse einsetzen, frei. Für Methoden der digitalen Erschließung bedeutet dies, dass WissenschaftlerInnen grundsätzlich keine Vorschriften oder Vorgaben bezüglich der Wahl der Werkzeuge oder Technologien gemacht werden können. In der Praxis hängt diese Wahl häufig stärker von existierenden Präferenzen und Kenntnissen der beteiligten ProjektmitarbeiterInnen ab als von der besonderen Eignung des Werkzeugs für den spezifischen Zweck; Erschließungs-, Annotations-, Auswertungs- und Präsentationswerkzeuge werden ad-hoc in den Technologien entwickelt, in der sich der technische Mitarbeiter am besten auskennt.

Aus dieser Praxis heraus entstehen an den unterschiedlichsten Orten Softwarekomponenten auf einer unüberschaubaren Vielzahl technischer Plattformen, darunter unterschiedliche Betriebssysteme, Laufzeitumgebungen, Datenbanken, Programmiersprachen etc., die in einigen Fällen nahezu identische Aufgaben erfüllen. Im Falle von befristeten Drittmittelprojekten stellt sich zum Ende der Projektlaufzeit die Frage, wie die entstandenen technischen Lösungen ohne die personellen Ressourcen aus dem Projekt weiterbetrieben werden können, und die enttäuschende Antwort lautet in vielen Fällen: Gar nicht.

Eine Strategie auf dem Weg zu mehr Nachhaltigkeit bei existierenden und entstehenden Forschungsanwendungen ist es, die Zahl der bei ihrer Erstellung verwendeten technischen Lösungen zu beschränken. Die prinzipielle Idee dahinter ist: Je weniger Plattformen in einem zuständigen Datenzentrum existieren, desto einfacher lassen sich diese von einer überschaubaren Anzahl Menschen in Betrieb halten. Wenn – zum Beispiel – eine gewisse Anzahl von Anwendungen ein Fedora Commons-Repository als Grundlage nutzt und damit auf die gleichen technischen Basiskomponenten (Tomcat oder Jetty-Applikationsserver, Java EE) setzt, so können mehrere Anwendungen mit Updates an diesen Komponenten und an der Fedora-Installation auf einmal auf einen aktuellen Stand gebracht werden, wenn beispielsweise Sicherheitslücken geschlossen werden müssen. Auf einer gegebenen technischen Basis (Tomcat, Java EE, ein Datenbanksystem wie PostgreSQL, Apache Solr) ließen sich auch noch weitere Repositoriensysteme wie DSpace oder MyCoRe installieren und eine übersichtlichere Struktur schaffen.

Mit der Perspektive, dass durch Auswahl bestimmter Komponenten die längerfristige Pflege einer Forschungsanwendung durch ein Dienstleistungs-/Datenzentrum gewährleistet werden kann, lässt sich so aus unserer Erfahrung durchaus ein gewisser Lenkungseffekt erzielen. Die Festlegung auf eine vom Datenzentrum unterstützte Softwarelösung bedeutet dabei häufig zwar keine Einschränkung in Bezug auf die Leistungsfähigkeit oder Angemessenheit für die anvisierte Aufgabe, kann aber durchaus zu Problemen bei der Auswahl eines geeigneten technischen Mitarbeiters machen.

Das Poster soll die Vor- und Nachteile eines vorgegebenen, eingeschränkten Softwareangebotes mit „Pflegezusage“ aufzeigen: Lässt sich eine solche Umgebung wirklich leichter pflegen und administrieren? Wie „generisch“ sind derartige Lösungen wirklich? Welche Anwendungsfälle lassen sich so behandeln, wo sind die Grenzen?

# Nutzung einer Repositoriums-anwendung zur Modellierung epigraphischer Daten

*Franziska Weise, Historisches Seminar/ Alte Geschichte, Universität Hamburg,  
Hagen Peukert, gwin, Universität Hamburg*

Schätzungen für epigraphische Funde des Altertums reichen allein auf dem Gebiet Kleinasien derzeit auf eine Zahl von 60.000 bis 80.000 Artefakte. Um diese recht große Zahl in der recht kleinen Gemeinschaft von Epigraphikern systematisch zu sammeln, organisieren und analysieren zu können, ist die Unterstützung von Informationstechnologie wichtig. Dabei erfordern epigraphische Daten erstens einige Besonderheiten, die sich in der Modellierung der Daten niederschlagen, und zweitens spezielle Anforderungen in der engen Zusammenarbeit der Wissenschaftler berücksichtigen. So ist beispielsweise die möglichst detailgetreue Betrachtung des Objekts im kontextuellen wie auch singulären Umfeld wichtig und erfordert annotierbares Bildmaterial. Andererseits fließt in den Datensatz eines Fundes sehr unterschiedliche Expertise von vielen Wissenschaftlern und Quellen ein (Datierung versus inhaltliche Interpretation des Fundes).

Die griechische wie lateinische Epigraphikforschung hat sich in Teilen bereits in einer frühen Entwicklungsstufe neuen methodischen Ansätzen gestellt. Ausgangspunkt des vorliegenden Projekts sind inschriftliche Zeugnisse des antiken Kleinasien, die in einem über 12 Jahre (1993-2006) geförderten DFG-Projekt gesammelt und als edierte Volltexte in einer Online-Datenbank überführt wurden. Den damaligen technologischen Standards entsprechend, wurden die Ergebnisse des Projekts in einer SQL-Datenbank abgelegt und mittels einer skriptbasierten Anwendung zur Verfügung gestellt. Zur Weiterführung und Fortentwicklung des Projekts ist es Ziel, die dort erhobenen Inschriftendaten unabhängig von den Technologieschüben auf Dauer verfügbar und pflegbar zu halten.

Die Nachhaltigkeitsstrategie der Universität Hamburg sieht vor, dass geisteswissenschaftliche Forschungsdaten nach Möglichkeit auf Dauer zugänglich bleiben. In der konkreten Umsetzung bedeutet dies im vorliegenden Fall die Verwendung des MyCoRe-Frameworks, für das entsprechende Strukturen im Rechenzentrum der Universität vorgehalten werden. Daraus folgt, dass die Datenmodellierung auf die Belange des MyCoRe-Kerns angepasst wird.

Das Poster beschreibt die Erstellung einer nachhaltigen Internetanwendung zur Suche, Visualisierung und Pflege epigraphischer Daten auf Basis des MyCoRe-Frameworks. Wir zeichnen die Überführung der epigraphischen Daten von einer idiosynkratischen Datenbankapplikation hin zu einer xml-basierten, generischen Datenverwaltung mit moderner Suchtechnologie nach und beschreiben die spezifischen Anforderungen, die sich aus den Erfordernissen der Inschriftenforschung ergeben.

## Die Kehrseite der Modularität.

### Wie viele Komponenten verträgt ein nachhaltiges Repositorium?

*Daniel Jettka, hzsk, Universität Hamburg*

*Kai Wörner, gwin, Universität Hamburg*

Das Korpusrepositorium des Hamburger Zentrums für Sprachkorpora (hzsk) wurde entwickelt um die nachhaltige Organisation, Verwaltung und Publikation von linguistischen Korpora zu ermöglichen. Während der beiden Projektphasen (2011-2013, 2014-2017) des von der DFG geförderten Projekts „Etablierung eines Schwerpunkts Mehrsprachigkeit und Gesprochene Sprache am Hamburger Zentrum für Sprachkorpora“ wurde eine modulare, digitale Infrastruktur aufgebaut, die zunächst die Nachnutzbarkeit der Korpora des Sonderforschungsbereichs 538 „Mehrsprachigkeit“ (Universität Hamburg, 1999-2011) gewährleistete, mittlerweile aber auch zur Speicherung und Auslieferung einer Vielzahl weiterer linguistischer Ressourcen dient.

Das im Zentrum dieser Infrastruktur stehende „HZSK Repository“ wurde auf der Basis von Fedora Commons 3 entwickelt und zweimal mit dem „Data Seal of Approval“ als „vertrauenswürdigen, digitales Langzeitarchiv“ zertifiziert. Es bietet – neben der eigentlichen Speicherung der Korpusdaten – eine Fülle an Schnittstellen, an die Oberflächen und Dienste (insbesondere aus dem CLARIN-Kontext) andocken und Nutzern verschiedenartige Zugänge zu den Daten eröffnen.

Die verschiedenen Bausteine, insbesondere die Vielzahl an (standardisierten) Schnittstellen und Webservices, sind auf der einen Seite ein Faktor, der zu einer besseren Nachhaltigkeit und Vernetzung des Repositoriums beiträgt: Sollten Dienste, die den Zugang zu den Daten gewähren, ausfallen, so stehen immer noch eine Reihe von anderen Zugriffsmöglichkeiten zur Verfügung. Sollte beispielsweise die über CLARIN angebotene, repositorienübergreifende Inhaltssuche (Federated Content Search) nicht mehr zur Verfügung stehen, so ließe sich über die repositoriumsinterne Suche immer noch auf den Inhalten der vorgehaltenen Ressourcen suchen. Der modulare Aufbau erlaubt folglich die Kompensation ausfallender Komponenten sowie den Einsatz neuer Schnittstellen und Technologien.

Auf der anderen Seite erwächst aus einer zunehmenden Anzahl technischer Komponenten ein größerer Pflegebedarf und es ergibt sich ein steigender personeller und finanzieller Aufwand. Schon das Kernrepositorium setzt mit Fedora Commons, Drupal und Islandora drei Software-Komponenten ein, bei denen größere Versionsupdates keineswegs trivial sind. Obwohl sich in der Vergangenheit selten Probleme daraus ergeben haben, muss auch die zugrundeliegende Serverarchitektur (Betriebssystem, Web-, Datenbank- und Applikationsserver) regelmäßig aktualisiert werden. Die Schnittstellen, die dem Repositorium als Webservices zur Seite gestellt wurden, müssen ohnehin händisch gepflegt werden – dasselbe gilt für die angeschlossenen Dienste. Mit Blick auf den Langzeitbetrieb des Repositoriums ergibt sich daher zwangsläufig die Frage, wie viele Komponenten ein nachhaltiges Repositorium verträgt.

Das Poster zeigt die verwendeten Bausteine des HZSK Repository mit ihren potentiellen Problemen und Schwachstellen.

# Das Datenrepositorium Manuskriptkulturen

*Karsten Helmholz, SFB 950, Universität Hamburg*

*Stefan Thiemann, SFB 950, gwin, Universität Hamburg*

Das Ziel des Serviceprojektes „Datenrepositorium Manuskriptkulturen“ ist es, die systematische Sicherung und langfristige Nutzung aller Forschungsdaten zu gewährleisten. Durch den Einsatz eines zentralen Repositoriums für alle Teilprojekte des SFB 950 „Manuskriptkulturen in Asien, Afrika und Europa“ wird ein effektives Wissens- und Datenmanagement unterstützt.

Die wichtigste Aufgabe des INF Projektes war zunächst die Entwicklung und Implementierung einer Forschungsumgebung mit Repositorium. Auf der Grundlage von Erfahrungen aus früheren Projekten, den Projektskizzen und einer genauen Bedarfsanalyse bei allen beteiligten Personen entstand im Lauf der ersten Phase des SFB das „Datenrepositorium Manuskriptkulturen“. In dessen Mittelpunkt steht die Manuskriptdatenbank, in der die Teilprojekte alle für sie relevanten Digitalisate, Texte, Forschungsergebnisse, Publikationen und anderen Materialien speichern. Dafür wurde ein Datenmodell entwickelt, das einerseits allen Anforderungen der sehr unterschiedlich ausgerichteten Teilprojekte gerecht wird, dennoch aber leicht zu handhaben ist und das man zudem an die sich verändernden Anforderungen anpassen kann. Diese werden in einer fortlaufenden Bedarfsanalyse erhoben und technisch umgesetzt.

Die Oberfläche des Repositoriums wird durch eine Reihe weiterer Funktionen ergänzt, die über die gemeinsame Datenbank gespeichert und vernetzt werden: Der Bereich Ressourcen bietet die wichtigsten, immer wieder verwendeten Inhalte in strukturierter Weise an, ein Forum bietet die Möglichkeit zum virtuellen Gedankenaustausch und ein Kalender dient dem gesamten SFB zur Terminplanung. Außerdem gibt es, angelehnt an kommerzielle Dienste, die Möglichkeit, ungewöhnlich große Datenmengen, wie z. B. Digitalisate über die Forschungsumgebung zu senden und zu empfangen, was sich besonders zur schnellen Datensicherung während Feldforschungen als sinnvoll erwiesen hat.

Durch die Verknüpfung von Inhalten, Strukturen und einer Zeitebene in Form des Kalenders dient das Repositorium nicht nur zur Speicherung und Bereitstellung von Daten und Informationen, sondern dient gleichzeitig als Planungstool und Dokumentation des gesamten SFB und zur virtuellen Zusammenarbeit.

Parallel zur Entwicklung des Repositoriums als geschlossene, interne Oberfläche wurde eine öffentliche Webseite für den SFB entwickelt, die zum einen allgemein über den SFB informiert und zum anderen, im Sinne aktiver Öffentlichkeitsarbeit, ständig wechselnde Elemente enthält, wie Neuigkeiten, Veranstaltungshinweise und eine eigene Publikation, das „Manuskript des Monats“, die sich bewusst fachfremdes Publikum ansprechen soll. Das INF Projekt pflegt die Webseite des SFB und ist für die Aufbereitung von Material u. a. für die Veröffentlichung des Manuskript des Monats verantwortlich.

## Teilnehmer (Stand: 08. September)

<b>Bingert, Sven</b> sven.bingert@gwdg.de	Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
<b>Blumtritt, Jonathan</b> jonathan.blumtritt@uni-koeln.de	Data Center for the Humanites, Universität zu Köln
<b>Christ, Andreas</b> christ@ub.uni-kiel.de	Universitätsbibliothek Kiel
<b>Cordes, Nils</b> nils.cordes@binfort.de	binfort GmbH
<b>Dahnke, Michael</b> fedja_anatevka@web.de	
<b>Damerow, Julia</b> jdamerow@asu.edu	Arizona State University
<b>Daxenberger, Johannes</b> daxenberger@ukp.informatik.tu-darmstadt.de	Universität Darmstadt
<b>Dogan, Zeki Mustafa</b> dogan@sub.uni-goettingen.de	Georg-August-Universität Göttingen
<b>Druskat, Stephan</b> stephan.druskat@hu-berlin.de	Humboldt-Universität zu Berlin
<b>Engelhardt, Claudia</b> claudia.engelhardt@sub.uni-goettingen.de	Niedersächsische Staats- und Universitätsbibliothek Göttingen
<b>Feige, Tillmann</b> tillmann.feige@uni-hamburg.de	Universität Hamburg
<b>Gietz, Peter</b> karin.gietz@daasi.de	DAASI International
<b>Goedel, Martina</b> martina.goedel@textloop.de	textloop Martina Gödel
<b>Gonzalez, Alicia</b> alicia.gonzalez@uni-hamburg.de	Universität Hamburg
<b>Haft, Michael</b> Michael.Haft@adwmainz.de	Die Akademie der Wissenschaften und der Literatur, Mainz
<b>Hasler, Tim</b> hasler@zib.de	Zuse Institut Berlin
<b>Hedeland, Hanna</b> hanna.hedeland@uni-hamburg.de	HZSK, Universität Hamburg
<b>Helmholz, Karsten</b> karsten.helmholz@uni-hamburg.de	Universität Hamburg
<b>Hennicke, Steffen</b> hennicke@leibniz-gei.de	Georg-Eckert-Institut
<b>Hütter, Heiko</b> heiko.huetter@daasi.de	DAASI International
<b>Imeri, Sabine</b> sabine.imeri.1@ub.hu-berlin.de	Humboldt-Universität zu Berlin
<b>Jettka, Daniel</b> daniel.jettka@uni-hamburg.de	INEL, LIS, Universität Hamburg
<b>Krause, Thomas</b> krauseto@hu-berlin.de	Humboldt-Universität zu Berlin
<b>Kronenwett, Simone</b> simone.kronenwett@uni-koeln.de	Universität zu Köln
<b>Lange, Felix</b> flange@mpiwg-berlin.mpg.de	Max-Planck-Institut für Wissenschaftsgeschichte
<b>Lehmberg, Timm</b> timm.lehmberg@uni-hamburg.de	INEL, Universität Hamburg
<b>Mathiak, Brigitte</b> reinhard.messerschmidt@uni-koeln.de	Universität zu Köln
<b>Meise, Bianca</b> bianca.meise@upb.de	Universität Paderborn

<b>Messerschmidt, Reinhard</b> rmessers@smail.uni-koeln.de	Universität zu Köln
<b>Moeller, Katrin</b> katrin.moeller@geschichte.uni-halle.de	Martin-Luther-Universität Halle-Wittenberg
<b>Müller, Lars</b> l.mueller@dipf.de	DIPF
<b>Neumann, Kathleen</b> neumann@gbv.de	Verbundzentrale des GBV (VZG)
<b>Oeltjen, Wiebke</b> wiebke.oeltjen@uni-hamburg.de	Universität Hamburg
<b>Oltmanns, Elias</b> oltmanns@zib.de	Zuse-Institut Berlin (ZIB)
<b>Peukert, Hagen</b> hagen.peukert@uni-hamburg.de	Universität Hamburg
<b>Rau, Felix</b> f.rau@uni-koeln.de	Universität zu Köln
<b>Rümenapf, Hinnerk</b> hinnerk.ruemenapf@uni-hamburg.de	Universität Hamburg
<b>Schima-Voigt, Kristine</b> schima-voigt@sub.uni-goettingen.de	Niedersächsische Staats- und Universitätsbibliothek Göttingen
<b>Schloots, Franziska</b> franziska.schloots@uni-paderborn.de	Universität Paderborn
<b>Schmidt, Thomas</b> thomas.schmidt@ids-mannheim.de	Institut für Deutsche Sprache, Mannheim
<b>Siegert, Wenke</b> wenke.siegert@sub.uni-hamburg.de	Staats- und Universitätsbibliothek Hamburg
<b>Sindt, Ruth</b> sindt@ub.uni-kiel.de	UB Kiel
<b>Stanzel, Arnost</b> arnost.stanzel@bsb-muenchen.de	Bayerische Staatsbibliothek
<b>Strauch, Annette</b> Annette.Strauch@uni-siegen.de	ZIMT, Universität Siegen
<b>Struck, Alexander</b> Alexander.Struck@hu-berlin.de	Humboldt-Universität zu Berlin
<b>Thiemann, Stefan</b> stefan.thiemann@uni-hamburg.de	Fakultät für Geisteswissenschaften, Universität Hamburg
<b>Toepfer, Ralf</b> r.toepfer@zbw.eu	ZBW
<b>Vertan, Cristina</b> cristina.vertan@uni-hamburg.de	Universität Hamburg
<b>Vieten, Anne</b> vieten@uni-mainz.de	Johannes Gutenberg-Universität Mainz
<b>Vogel, Iris</b> iris.vogel@uni-hamburg.de	Universität Hamburg
<b>von dem Bussche, Ruth</b> rbussche@fotostoria.de	Fotostoria
<b>Wagner, Cosima</b> cosima.wagner@fu-berlin.de	Freie Universität Berlin
<b>Weingart, Anja</b> aweinga@gwdg.de	Georg-August-Universität Göttingen
<b>Weise, Franziska</b> franziska.weise@uni-hamburg.de	Universität Hamburg, FB Geschichte, AB Alte Geschichte
<b>Wonke-Stehle, Jens</b> jens.wonke-stehle@sub.uni-hamburg.de	Staats- und Universitätsbibliothek Hamburg
<b>Wörner, Kai</b> kai.woerner@uni-hamburg.de	Gwin, Universität Hamburg
<b>Wuttke, Ulrike</b> wuttke@akademienunion-berlin.de	Union der deutschen Akademien der Wissenschaften

**[uhh.de/forge2016](http://uhh.de/forge2016)**