

Forschungsinfrastrukturen

Auf dem (langen) Weg zu den e-Humanities

Von Andreas Henrich

Es ist eine verlockende Vision: Wenn die in zahlreichen kultur- und geisteswissenschaftlichen Einzelprojekten gesammelten Forschungsdaten miteinander verknüpft werden, könnte dies fachübergreifende Analysen erheblich vereinfachen. Wie sinnvoll der wissenschaftliche Austausch zwischen ganz unterschiedlichen Disziplinen in der Praxis sein kann, zeigt ein Beispiel: Um universelle Trends in der Literatur und Architektur des Mittelalters ableiten zu können, müssen Spezialisten aus mehreren Bereichen zusammenarbeiten – nämlich Bauhistoriker und Philologen. Aber auch innerhalb der Einzeldisziplinen ist ein umfassender Erkenntnis-austausch durchaus vielversprechend. So lassen sich beispielsweise aus mehreren, unabhängig voneinander entstandenen Arbeiten über die Portraitmalerei verschiedener Epochen ggf. neue Erkenntnisse über das Bildverständnis im Zeitverlauf gewinnen. Die Forschungsinitiative DARIAH-DE (Digital Research Infrastructure for the Arts and Humanities) leistet einen wichtigen Beitrag dazu, diese Vision schon bald Realität werden zu lassen. Im Rahmen des umfangreichen Projekts befassen sich Bamberger Informatiker mit der Datenintegration und der Etablierung von übergreifenden Suchdiensten.



Zwei Welten, die gemeinsam neue Perspektiven für die Geistes- und Kulturwissenschaften eröffnen können.

In den Kultur- und Geisteswissenschaften fallen in immer größerer Menge digitale Forschungsdaten an. Dieser Datenschatz kann erst dann sein volles Potential entfalten, wenn die Informationen automatisch verarbeitet und vor allem sinnvoll miteinander verknüpft werden. Genau an dieser Stelle setzt die Forschungsinitiative DARIAH (Digital Research Infrastructure for the Arts and Humanities) an. Mit DARIAH soll eine nachhaltige digitale Infrastruktur in Europa geschaffen werden, die die Langzeitverfügbarkeit von Forschungsdaten aus der Kulturgutforschung und den Geisteswissenschaften sicherstellt. Inhaltlich steht dabei die Verlinkung von Personenangaben, Diensten und Forschungsergebnissen im Fokus. Diese Verknüpfung ermöglicht es, die vorhandenen Synergiepotentiale umfassend zu nutzen. Bis es soweit ist, gilt es freilich einige technische Stolpersteine auf syntaktischer und semantischer Ebene aus dem Weg zu räumen.

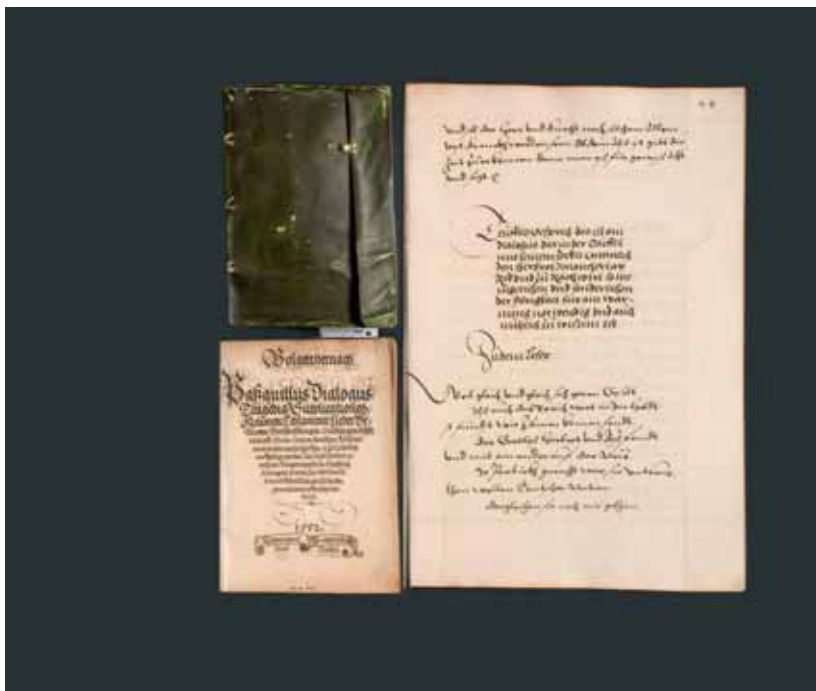
Der Projektrahmen

Das 7. und 8. Rahmenprogramm der Europäischen Kommission haben unter anderem den Aufbau nachhaltiger Forschungsinfrastrukturen (FIs) in allen Wissenschaftszweigen zum Ziel. Die Planung und Steuerung der entsprechenden Projekte liegt dabei beim European Strategy Forum for Research Infrastructures (ESFRI). Gemäß der ESFRI-Roadmap aus dem Jahr 2008 stehen zunächst 44 Projekte aus den Bereichen Social Sciences and Humanities, Environmental Sciences, Energy, Biological and Medical Sciences, Materials and Analytical Facilities, Physical Sciences and Engineering sowie e-Infrastructures im Vordergrund.

Die Schaffung einer zukunftsfähigen Forschungsinfrastruktur bezieht sich dabei nicht nur auf den Bau von neuen Forschungsschiffen oder die Erneuerung von Hochsicherheitslaboren für die Pandemieforschung, sondern umfasst eben auch die Verbesserung des Datenaustauschs im Bereich der Geistes- und Kulturwissenschaften mithilfe von DARIAH.

Der übergeordnete Entwicklungsplan für die FIs ist in drei Phasen unterteilt:

- eine Präparationsphase (preparatory phase; 2008-2010),
- eine Konstruktionsphase (construction phase; 2011-2015) und
- eine Nutzungsphase (operational phase; 2016-2020).



Eine frühneuzeitliche Handschrift...

wird über das TEI-XML-Format digitalisiert und editiert...

Die wichtigsten Partner beim Infrastrukturprojekt DARIAH in der bereits abgeschlossenen Präparationsphase waren das CeRch am King's College in London, die Max Planck Digital Library, die Niedersächsische Staats- und Universitätsbibliothek in Göttingen, die Data Archiving and Networked Services in Den Haag, das Department of Scandinavian Research an der Universität Kopenhagen sowie das Centre National de la Recherche Scientifique (CNRS) in Paris. In dieser ersten Projektphase wurden zunächst die Anforderungen an eine entsprechende Infrastruktur umfassend zusammengetragen und Architekturmodelle sowie technische Prototypen erstellt.

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <EXCERPT TEI STRUCT? http://gdl.mpi.de/tei/struct/struct.dtd?>
3 <TEI xmlns="http://www.tei-c.org/2001/XMLSchema-instance" xmlns:xsi="http://www.tei-c.org/1.0/xsi"
4 <!-->
5 <!-->
6 <!-->
7 <!-->
8 <!-->
9 <!-->
10 <!-->
11 <!-->
12 <!-->
13 <!-->
14 <!-->
15 <!-->
16 <!-->
17 <!-->
18 <!-->
19 <!-->
20 <!-->
21 <!-->
22 <!-->
23 <!-->
24 <!-->
25 <!-->
26 <!-->
27 <!-->
28 <!-->
29 <!-->
30 <!-->
31 <!-->
32 <!-->
33 <!-->
34 <!-->
35 <!-->
36 <!-->
37 <!-->
38 <!-->

```

An die primär aus dem Rahmenprogramm der Europäischen Kommission geförderte Präparationsphase schließt sich nun die von einzelnen Mitgliedsstaaten getragene Konstruktionsphase an. Im März 2011 erfolgte zum Beispiel der Startschuss für das vom Bundesministerium für Bildung und Forschung (BMBF) geförderte Vorhaben „DARIAH-DE: Aufbau von Forschungsinfrastrukturen für die e-Humanities“. Unter der Leitung der Niedersächsischen Staats- und Universitätsbibliothek Göttingen haben sich zu diesem Zweck 18 deutsche Partner zusammengefunden. In enger Abstimmung mit den weiteren DARIAH-Teilprojekten anderer europäischer Länder geht es um die konkrete Umsetzung der Zielvorgaben im Bereich der Geistes- und Kulturwissenschaften. Hierzu gibt es auf europäischer Ebene das „DARIAH-EU-Coordination Office (DCO)“ sowie thematisch definierte Virtuelle Kompetenzzentren: VCC 1 „e-Infrastructure“, VCC 2 „Research and Education Liaison“, VCC 3 „Scholarly Content Management“ sowie VCC 4 „Advocacy, Impact and Outreach“.

Die Vision

Ziel von DARIAH-EU ist der Aufbau von Forschungsinfrastrukturen und die beispielhafte Entwicklung von virtuellen Forschungsumgebungen, welche neue, kollaborative und digitale Forschungsmethoden unterstützen. Insbesondere sollen neue Interpretationen bestehenden Wissens gefördert,

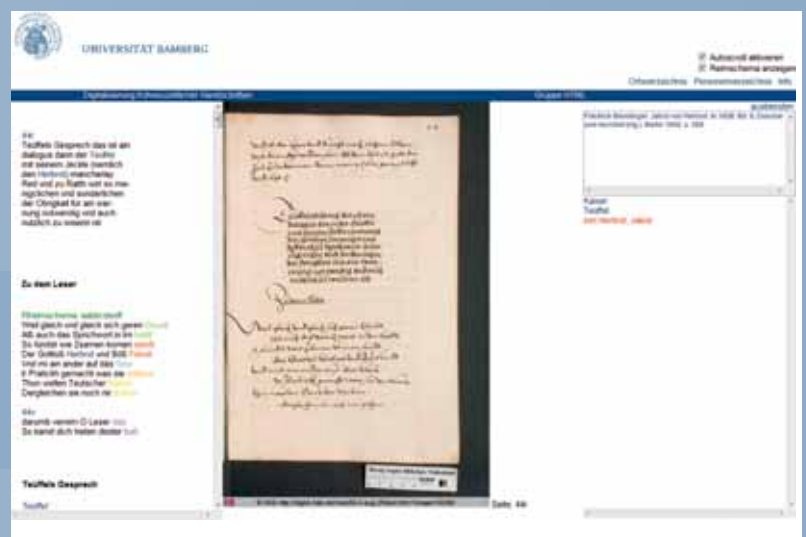
die Formulierung von neuen Forschungsfragen ermöglicht und deren Beantwortung gleichzeitig erleichtert werden. Erreicht wird dies durch

- die größtmögliche Unterstützung von Forschern und Forscherinnen und ihren Forschungsprozessen aus allen geistes-, sozial- und kulturwissenschaftlichen Disziplinen bei zunehmend komplexer werdenden Arbeitsabläufen,
- die Bereitstellung von Werkzeugen und Technologien, die von den Forschern benötigt werden, um kollaborativ, Disziplin übergreifend, international und institutionell unabhängig interagieren zu können,
- den freien Zugriff auf Ressourcen (Daten, Dienste) und auf eine umfassende technische Forschungsinfrastruktur (lokal, national, europäisch).

Der Bamberger Beitrag

Um die übergreifende Erschließung und inhaltliche Verknüpfung von geistes- und kulturwissenschaftlichen Archiven und Sammlungen zu ermöglichen, müssen zunächst die technischen Voraussetzungen geschaffen werden. An dieser Stelle kommen die Angewandten Informatiker der Universität Bamberg ins Spiel. Im Fokus der Forscher steht dabei einerseits die sinnvolle Strukturierung der Daten sowie andererseits die Implementierung von leistungsfähigen Suchalgorithmen. Da im Rahmen von DARIAH auch multilinguale und multimedie-

... und schließlich mit allen
Informationen als Webseite angezeigt.



Research Infrastructure



The (long) road to the e-humanities

It's an enticing vision: If the collected research data from numerous individual humanities projects were linked, interdisciplinary analysis could be simplified considerably. The practical value of academic exchange between various disciplines can be illustrated by the following example: In order to deduce universal trends in medieval literature and architecture, specialists in multiple fields (namely construction historians and philologists) have to work together. But comprehensive knowledge exchange in individual disciplines is also quite promising. So, by way of example, multiple independent projects dealing with portraiture in different eras can all lead to new insights into the appreciation of paintings over time. The research initiative DARIAH-DE (Digital Research Infrastructure for the Arts and Humanities) is making an important contribution to the realization of this vision. Within the scope of this extensive project, Bamberger computer scientists are working on data integration and the establishment of comprehensive search resources.

ale Daten betrachtet werden, ist neben inhaltlichen Aspekten insbesondere eine Verknüpfung und Suche auf Basis von Metadaten sinnvoll. Das Problem dabei: die entsprechenden Metadaten sind formal (unterschiedliche Schemata) und pragmatisch (unterschiedliche Nutzung der Felder) heterogen und weisen eine höchst unterschiedliche Qualität auf.

Durch eine Schema-Registry für Metadaten soll eine Infrastruktur für die übergreifende Betrachtung geschaffen werden. Dazu werden aktuelle Ansätze aus den Bereichen Schema- und Datenintegration, Schema- und Ontologie-Matching und der multilingualen Datenintegration zu einem Gesamtkonzept verknüpft. Diese Basisinfrastruktur ermöglicht bereits eine integrative und übergreifende Betrachtung von Archiven und Sammlungen ohne großen manuellen Aufwand. In einem weiteren Schritt werden dann generisch übergreifende Suchdienste realisiert, die eine reine Textsuche mit einer facettierten Suche kombinieren. Dabei werden vorhandene Schemainformationen genutzt. Zusätzlich sollen in den generischen Suchdienst nahtlos spezifische Suchfunktionalitäten, wie z. B. medien- oder quellenspezifische Vergleichsmaße, integriert werden.

Die Bamberger Informatiker befassen sich im Rahmen des Projekts mit der Konzeption und prototypischen Umsetzung der Metadaten-Registry,

der Integration wichtiger Metadaten-Standards, der Evaluierung der Leistungsfähigkeit der Schemaintegration an ausgewählten Beispielen, der Konzeption und prototypischen Umsetzung des generischen Suchdienstes, der Evaluierung der Effektivität und Effizienz des Suchdienstes und ggf. mit der Umsetzung einzelner spezifischer Suchmöglichkeiten zur Überprüfung des Erweiterungskonzeptes.

Im Hinblick auf die Nachhaltigkeit der erzielten Ergebnisse setzt das Teilprojekt hierbei auf etablierte Standards, eine einfache Erweiterbarkeit und auf Open Source. Weder für die Darstellung der Inhalte oder für die Metadaten selbst noch für die Integration von Metadaten sollen neue Formate und Standards entwickelt werden. Primäres Ziel ist vielmehr die Nutzung und Nutzbarmachung geistes- und kulturwissenschaftlicher Archive und Sammlungen auf Basis existierender Standards sowohl für Inhalte als auch Metadaten. Dazu werden in der Schema-Registry für Metadaten wesentliche Standards (wie z. B. TEI-Header oder Dublin Core) bereits als Referenzen hinterlegt. Ferner sollen zur Integration und zur Modellierung der Beziehungen zwischen den Schemata ebenfalls Standards wie OWL, RDF bzw. SKOS eingesetzt werden.

Zur Gewährleistung der Erweiterbarkeit wird insbesondere für die Suchdienste eine offene, auf bewährten Entwurfsmustern basierende Architektur genutzt. Dadurch soll auch nach Jahren die

einfache und flexible Nachnutzung ermöglicht werden. Die im Rahmen des Teilprojekts entwickelten Software-Komponenten werden aus diesem Grund mit offenen Schnittstellen ausgestattet. So ist dann z. B. auch die Nutzung als Web-Service möglich. Die Software wird sich voraussichtlich in einem hohen Maß auf frei verfügbare Komponenten wie beispielsweise Apache Solr stützen. Dies hat den Vorteil, dass sie ggf. selbst als Open Source bereitgestellt werden kann. Die eigens entwickelte Software und die erforderlichen Basissysteme können somit sehr frei und flexibel genutzt und ggf. an neue Anforderungen angepasst werden.

Zusammenfassung und Ausblick

Die Schaffung einer digitalen Infrastruktur im Bereich der Geistes- und Kulturwissenschaften ist zweifelsohne weniger spektakulär als der Bau eines naturwissenschaftlichen Forschungsschiffs – in mancher Hinsicht ist ein solches Projekt allerdings schwieriger in die Tat umzusetzen. Vor allem ist der Erfolg nicht immer sofort sichtbar. Wenn das Schiff erst einmal fährt, verfliegen viele Zweifel fast von selbst. Um hingegen Forschungsdaten miteinander vernetzen zu können, gilt es zunächst, einige kontroverse Fragen zu beantworten: Wie kann es beispielsweise gelingen, neue Forschungsmethoden und neue Kollaborationsmodelle in der

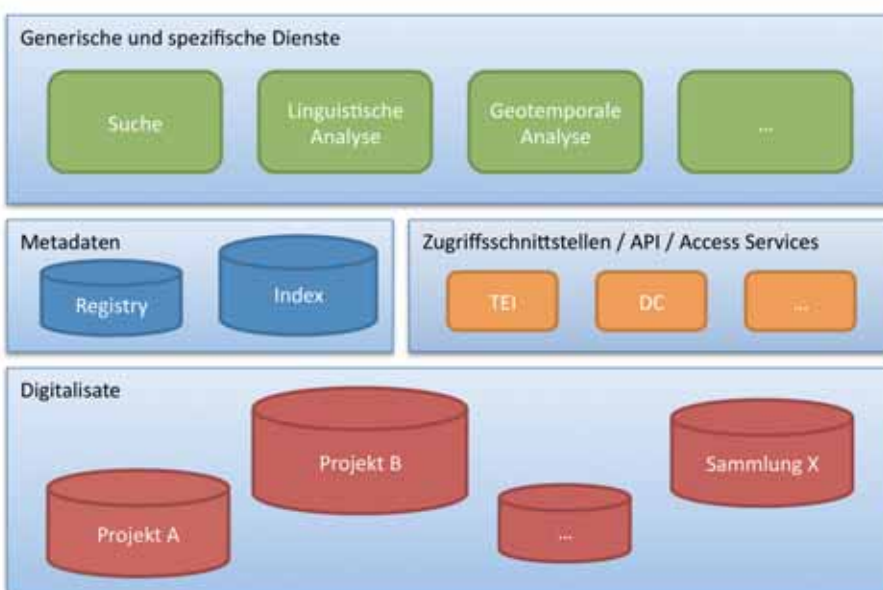
DARIAH-DE:
<http://de.dariah.eu>

DARIAH-EU
<http://dariah.eu>

ESFRI:
ec.europa.eu/research/esfri

Deutsches Portal zum
7. EU-Forschungsrahmenprogramm:
www.forschungsrahmenprogramm.de

Wissenschaft zu etablieren? Und wie bringt man dann Wissenschaftlerinnen und Wissenschaftler dazu, ihre Forschungsergebnisse in eine solche Forschungsinfrastruktur einzubringen? Die Schaffung der Voraussetzungen für e-Humanities im Sinne der Vision von DARIAH ist also mehr als nur ein technisches Problem. Aus Sicht der Bamberger Informatiker muss die Informationstechnik gerade in diesem Umfeld ein unaufdringliches und letztlich auch unauffälliges Werkzeug sein, das nichtsdestotrotz neue Möglichkeiten schafft und so im Arbeitsalltag der Wissenschaftler überzeugt.



Die Dienste realisieren auf Basis der registrierten Metadaten, der Indexdaten für den schnellen Zugriff und der Programmierschnittstellen generische und spezifische Funktionalitäten für verschiedene Zielgruppen.

Um einen übergreifenden, integrierten Zugriff zu ermöglichen werden Metadaten und Programmierschnittstellen als Dienste in einer Service-orientierten Architektur bereitgestellt.

Digitalisate werden integriert, indem ihre Metadaten registriert werden. Für den Zugriff können sowohl der Direktzugriff als auch der Zugriff über allgemeinere Dienste der höheren Ebenen genutzt werden.