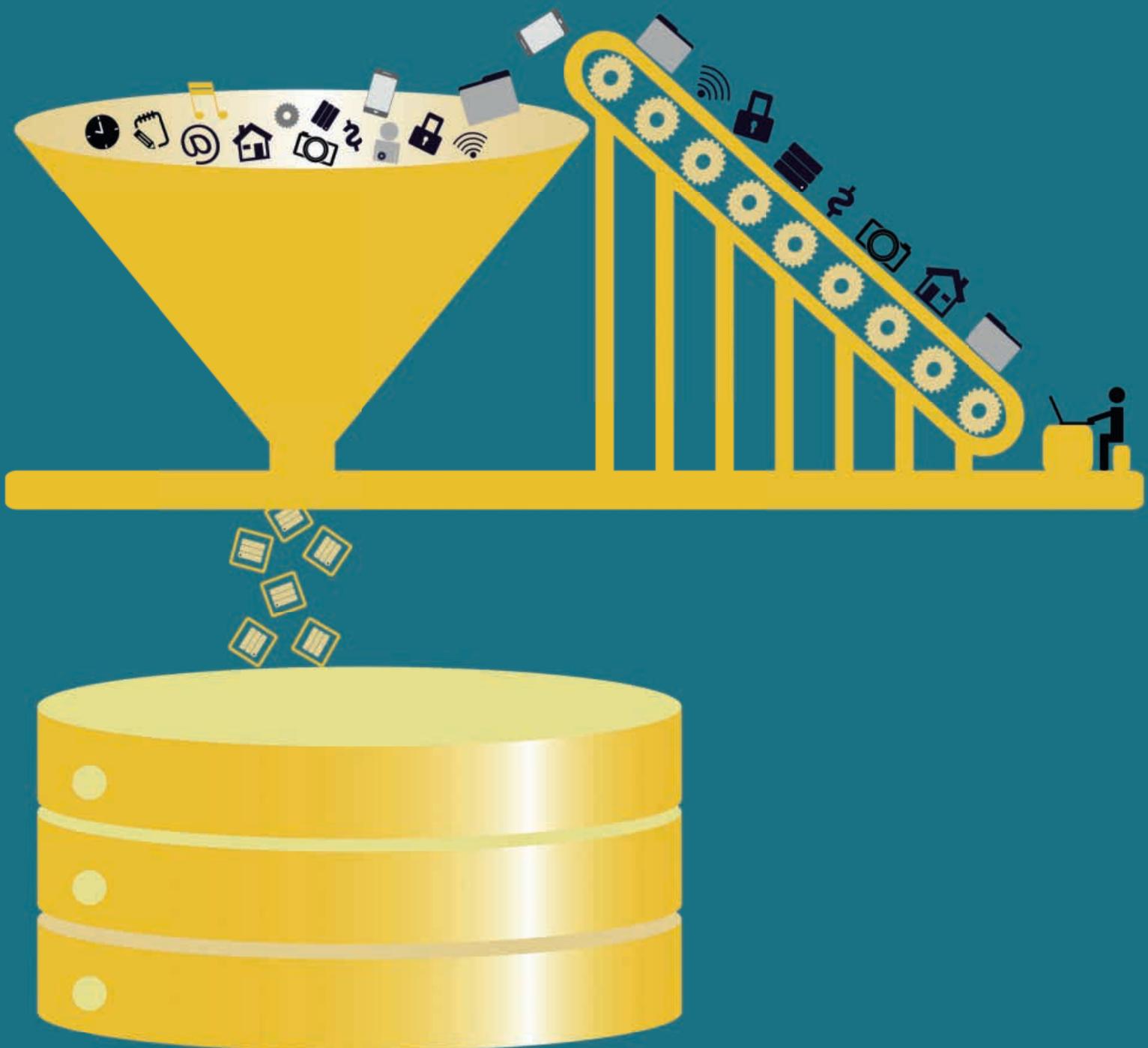


---

# Dossier

## Geisteswissenschaftliche Forschung und ihre Daten



(ib) Forschungsdaten und der adäquate Umgang damit beschäftigen die Wissenschaftlerinnen und Wissenschaftler im Zeitalter von Big Data mehr denn je. Die SAGW hat dieses zentrale Thema deshalb auf ihre Tagungsagenda gesetzt: Am 20. November 2015 findet in Bern die Veranstaltung «Geisteswissenschaftliche Forschung und ihre Daten: Stand, Herausforderungen und Perspektiven im digitalen Zeitalter» statt.

Im folgenden Dossier kommen viele Aspekte der SAGW-Tagung zu den Forschungsdaten zur Sprache: Patrick Sahle gibt einen Überblick zur Frage, was denn eigentlich geisteswissenschaftliche Forschungsdaten sind; Pierre-Yves Burgi erläutert das gerade für neue Forschungsprojekte immer wichtiger werdende Datenmanagement; Marianne Hundt und Eva Hanke referieren die zentralen Standards für digitale Editionen und Sammlungen, während Tristan Weddigen und Matej Ďurčo das Potenzial vernetzter Daten aufzeigen. Hans Bickel und Wolfgang Klein demonstrieren gelungene Umsetzungen digitaler Forschungsinfrastrukturen am Beispiel von sprachwissenschaftlichen Ressourcen. Lukas Rosenthaler berichtet über das Dienstleistungsangebot des SAGW-Projekts «Data and Service Center for the Humanities DaSCH» sowie Metagrid. Eröffnet wird das Dossier mit einem Interview mit Gabi Schneider zum SUK-Programm P-2 «Wissenschaftliche Information: Zugang, Verarbeitung und Speicherung».

## Inhaltsverzeichnis Dossier

- 37** Geisteswissenschaftliche Forschung und ihre Daten
- 38** Digital Humanities – ein Thema der SAGW
- 40** Eine nationale «Service-Drehscheibe» für die Wissenschaften
- 43** Forschungsdaten in den Geisteswissenschaften  
*Patrick Sahle*
- 46** Geisteswissenschaftliche Daten sichern und verfügbar halten. *Lukas Rosenthaler*
- 48** Data Life-Cycle Management:  
The Swiss Way. *Pierre-Yves Burgi*
- 51** Standards bei digitalen Editionen. *Marianne Hundt*
- 53** Digitale Sammlungen in Bibliotheken. *Eva Martina Hanke*
- 55** Bilddatenbanken: Geisteswissenschaften im Medienwandel. *Tristan Weddigen*
- 57** ortsnamen.ch – Portal der schweizerischen Ortsnamenforschung. *Hans Bickel, Martin Graf*
- 59** Das Digitale Wörterbuch der deutschen Sprache  
*Wolfgang Klein*
- 61** LOD entzaubert: Was semantische Technologien wirklich können. *Matej Ďurčo*
- 63** Tagungsprogramm DH15:  
Geisteswissenschaftliche Forschung und ihre Daten: Stand, Herausforderungen und Perspektiven im digitalen Zeitalter

## Weitere Informationen

### SAGW-Tagung

«Geisteswissenschaftliche Forschung und ihre Daten: Stand, Herausforderungen und Perspektiven im digitalen Zeitalter»

20. November 2015, Bern

Weitere Informationen: <http://dh15.sagw.ch>



## Digital Humanities – ein Thema der SAGW

38

(ib) *In den Geisteswissenschaften sind die Digital Humanities gegenwärtig omnipräsent: Es reihen sich Tagungen an Tagungen, Workshops an Workshops, an den Universitäten werden Einführungskurse angeboten, Datenzentren werden aufgebaut, die Forderungen nach Open Access und Open Data werden dringlicher – untrügliche Zeichen, dass die Beschäftigung mit den digitalen Geisteswissenschaften en vogue ist.*

Das Bild der hereinbrechenden Flut von Informationen, die es zu bewältigen gilt, wurde verschiedentlich evociert.<sup>1</sup> Man spricht auch von einzelnen Wellen der Digital Humanities: Die erste Welle in den 1990er-Jahren zeichnet sich durch Infrastrukturprojekte und grosse Digitalisierungsvorhaben aus, war demnach eher technisch geprägt. Seit den 2005er-Jahren ist eine zweite Welle in den DH zu konstatieren, die sich durch eine Beschäftigung mit den Inhalten der digitalen Daten, deren Visualisierung, Nutzung, Vernetzung und den dafür nötigen Tools und Methoden auszeichnet. Die dritte Welle soll uns noch bevorstehen, wobei sie eher an ein langsames Ansteigen der Pegel erinnert, da der Medienwandel die geisteswissenschaftlichen Episteme, deren «harter Kern» (Imre Lakatos), schleichend und zugleich grundlegend verändernd soll, so die Thesen von Todd S. Pressner und David M. Berry.<sup>2</sup>

### Riding the wave

Die SAGW sprang schon verhältnismässig früh auf die digitale Welle auf: Die Beschäftigung der Akademie mit dem Digitalen reicht mittlerweile über 25 Jahre zurück. Anfang der 1990er-Jahre lancierte die SAGW die «Datenbank Schweizerischer Kulturgüter», ein Repository für museale Objekte *avant la lettre*, das – wie man rückblickend feststellen muss – mit den damals vorhandenen Ressourcen noch nicht realisierbar und auf Dauer zu halten war. Wenige Jahre später konnte der Schweizerische Informations- und Datenarchivdienst für die Sozialwissenschaften SIDOS die ersten Datensätze sozialwissenschaftlicher Forschung für die weitere Nutzung anbieten, aus dem der Dienstleister «Data and research information services DARIS» bei FORS in Lausanne entstanden ist. Im April 2005 veranstaltete die SAGW eine Tagung zum Thema «Digitalisierung und Urheberrecht», 2007 folgten zwei weitere Veranstaltungen, diesmal zu Open Access (März 2007) und zum digitalen Gedächtnis der Schweiz (November 2007). Im gleichen Jahr wurden erste Stimmen aus der geisteswissenschaftlichen Fachgemeinschaft laut, die einen dringenden Bedarf für die nachhaltige Vorhaltung von Forschungsdaten anmeldeten. Daraufhin begannen die Vorbereitungsarbeiten für ein Daten- und Dienstleistungszentrum für geisteswissenschaftliche Forschungsdaten, das heute unter dem Namen «Data and Service Center for the Humanities DaSCH» im Pilotbetrieb seine Dienste anbietet und voraussichtlich 2017 institutionalisiert werden kann.

Mit den genannten Projekten und Initiativen werden die Themen genannt, welche die Akademie heute und aller Voraussicht nach auch in der näheren Zukunft im Bereich der Digital Humanities beschäftigen werden:

- die Erzeugung von digitalen Forschungsdaten,
- deren nachhaltige Sicherung,
- die Gewährleistung des offenen Zugangs sowie
- Voraussetzungen für die Nachnutzung von Forschungsdaten schaffen.

1 Vgl. etwa *Riding the Wave: How Europe can gain from the rising tide of scientific data* (2010), Download: <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>

2 Todd Samuel Pressner: *Digital Humanities 2.0: A Report on Knowledge*, 2010; Download: [cnx.org/content/m34246/1.6/?format=pdf](http://cnx.org/content/m34246/1.6/?format=pdf); David M. Berry: *Introduction: Understanding the Digital Humanities*, in: ders. (Hrsg.): *Understanding Digital Humanities*. London 2012, S. 1–20.

### Vieles ist schon erreicht – vieles bleibt noch zu tun

Mit Blick auf die genannten Themen, mit denen sich die SAGW seit längerem beschäftigt, können wir feststellen, dass vieles schon erreicht worden ist. Wir haben ein funktionierendes Daten- und Dienstleistungszentrum im Pilotbetrieb, der Vernetzungsdienst Metagrid ist in Betrieb, viele Periodika der SAGW können online und frei zugänglich konsultiert werden, und die Daten der durch die Akademie betriebenen Forschungsinfrastrukturen liegen ebenfalls zum grössten Teil frei zugänglich vor.

Doch damit sind noch längst nicht alle Ziele erreicht, vieles bleibt zu tun:

Damit Digital Humanities überhaupt betrieben werden können, müssen *digitale Daten* vorliegen. In der Mehrjahresplanung 2017 bis 2020 wird als strategisches Ziel der Akademie festgehalten, «dass bis 2020 die durch sie [i.e. SAGW] betreuten Einrichtungen korpusfähige, maschinenlesbare Daten zur Verfügung stellen können».<sup>3</sup> Die Akademieunternehmen der Nationalen Wörterbücher, *Année Politique Suisse* und das Inventar der Fundmünzen der Schweiz befinden sich auf diesem Weg, andere wie die Diplomatischen Dokumente der Schweiz und das Historische Lexikon der Schweiz oder *infoclio.ch* sind bereits digitalisiert. Es stehen jedoch noch weitere Herausforderungen an wie die Zertifizierungen der Datenbestände (z.B. durch das *Data Seal of Approval*: [www.datasealofapproval.org](http://www.datasealofapproval.org)) und die Aufnahme in Verzeichnisse von Datenrepositorien wie die *Registry of Research Data Repositories* ([www.re3data.org](http://www.re3data.org)). Auf der Ebene der einzelnen Daten muss eine persistente Identifikation angeboten werden, damit die Daten zitierfähig werden, wobei das *Digital Object Identifier System* ([www.doi.org](http://www.doi.org)) am häufigsten verwendet wird. Schliesslich sind auch die rechtlichen Aspekte eindeutig zu klären, etwa mit Hilfe des flexiblen Systems von Zugriffsstufen von *Creative Commons* ([www.creativecommons.ch](http://www.creativecommons.ch)), wobei der offene Zugang zu den Daten und die Möglichkeit zum uneingeschränkten Download die Zielgrösse bleibt. Diese Massnahmen dienen letztlich dazu, Vorausset-

zungen für die Nachnutzung von Forschungsdaten zu schaffen.<sup>4</sup>

### Nachhaltige Sicherung des Zugangs

Die nachhaltige Sicherung des Zugangs zu den Forschungsdaten beschäftigt gleich mehrere Initiativen in der Schweiz, so etwa das *SUK-P-2-Förderprogramm* (siehe auch S. 40/48). Das SAGW-Projekt des Daten- und Dienstleistungszentrums *DaSCH* befindet sich auf gutem Wege zur stabilen Einrichtung ab 2017, so die Verlautbarungen des Staatssekretariats für Bildung, Forschung und Innovation. Zu entscheiden ist noch über die Anschubfinanzierung von lokalen Satelliten an den geisteswissenschaftlichen Fakultäten, die *First-Level-Support* anbieten können. Mit diesem Gesamtpaket – eine zentrale Koordinationsstelle an der Uni Basel mit lokalen, schweizweit verteilten Satelliten – bietet sich die einmalige Chance, eine zeitgemässe Infrastruktur für geisteswissenschaftliche Forschungsdaten aufbauen zu können, welche die genannten Erfordernisse wie Zertifizierung, Identifizierung und auch *Linked Open Data* zur Vernetzung der Daten entscheidend vorantreiben könnte.

### Baustelle Open Access

Der *Open Access* zu wissenschaftlichen Publikationen ist nach wie vor eine Baustelle in der Schweiz, obwohl sich viele Akteure, darunter auch die SAGW, um deren Behebung bemühen. Wie gegenüber den grossen Wissenschaftsverlagen *Open-Access-Forderungen* erfolgreich durchgesetzt werden können, demonstrierten kürzlich die Niederlande. Kurt de Belder (Universitätsbibliothek Leiden) legte an den *Open-Access-Tagen* in Zürich dar (8. September 2015), wie ein konzertierte Aktion des Bildungsministeriums, der UniversitätsrektorInnen und der Universitätsbibliotheken erfolgreich bessere *Open-Access-Konditionen* für die Forschenden mit den meisten internationalen Wissenschaftsverlagen aushandeln konnte.<sup>5</sup> Auch hierzulande dürfte ein Zusammengehen zwischen Politik, Hochschulen und Wissenschaftsförderung die besten Aussichten bieten, Erfolge zu erzielen.

3 MJP 2017–2020 der SAGW, S. 39; Download: [http://www.sagw.ch/dms/sagw/info\\_sagw/mjp/Mehrjahresplanung\\_2017-2020\\_SAGW\\_web](http://www.sagw.ch/dms/sagw/info_sagw/mjp/Mehrjahresplanung_2017-2020_SAGW_web)

4 Siehe dazu auch die einschlägigen Kapitel im ALLEA E-Humanities working group report «Going Digital: Creating Change in the Humanities» 2015, Download: [http://www.allea.org/Content/ALLEA/WG%20E%20Humanities/Going%20Digital\\_digital%20version.pdf](http://www.allea.org/Content/ALLEA/WG%20E%20Humanities/Going%20Digital_digital%20version.pdf)

5 Siehe [https://www.open-access.net/fileadmin/oat/oat15/slides/08-De-Belder-Kurt-The\\_Dutch\\_OA\\_proposition.pdf](https://www.open-access.net/fileadmin/oat/oat15/slides/08-De-Belder-Kurt-The_Dutch_OA_proposition.pdf)

## Eine nationale «Service-Drehscheibe» für die Wissenschaften

40

(bk) Das Programm «Wissenschaftliche Information: Zugang, Verarbeitung und Speicherung» (SUK 2013–2016 P-2) von swissuniversities will die Entwicklung national verfügbarer Dienste zur Datenbereitstellung und -verarbeitung fördern. Mit einem nachhaltigen Grundangebot von digitalen wissenschaftlich relevanten Daten und optimalen Werkzeugen für deren Verarbeitung soll die Schweizer Wissenschaft im internationalen Wettbewerb gestärkt werden.

**BEATRICE KÜBLI** Frau Schneider, Sie betreuen das Programm «Wissenschaftliche Information: Zugang, Verarbeitung und Speicherung». Zunächst ganz kurz: Was ist das Ziel dieses Programms?

**GABI SCHNEIDER** Das Programm soll dazu beitragen, dass elektronische Publikationen und Forschungsdaten möglichst ohne Zutrittsbarrieren gefunden und genutzt werden können. Dazu sollen die Hochschulen vermehrt zusammenarbeiten und eine nationale Service-Drehscheibe aufbauen.

**BK** Woran liegt es, dass die Forschungsdaten bisher nicht ohne Weiteres zugänglich sind?

**GS** Nun, es gibt viele Hindernisse. Zunächst einmal sind viele Quellen nur analog vorhanden. Bei Forschungsdaten ist der Zugriff oft nicht möglich, weil sie nur lokal vorhanden sind. Bei digitalen Publikationen wiederum ist der Zugang häufig an Lizenzen geknüpft, die sich nicht alle Institutionen leisten können. Bei älteren Publikationen oder Bildern wiederum müssen die Nutzungsbedingungen oft aufwändig abgeklärt werden. Wollen Forschende digitale Objekte aus verschiedenen Systemen verarbeiten, stossen sie unter Umständen auf Formatprobleme. Und dies sind nur einige der Schwierigkeiten. Hinzu kommt eine Vielzahl von Umgebungen und Tools, die nicht immer miteinander kompatibel sind. Das Programm geht davon aus, dass eine vermehrte Zusammenarbeit der Hochschulen das verbessern

könnte. Es ist möglich, dass wir diesen Nutzen für die Hochschulen noch zu wenig deutlich machen konnten. Auch die föderalistische Zuständigkeit bei Bildung und Kultur stellt eine Herausforderung dar.

**BK** Was verstehen Sie unter einem «Grundangebot» digitaler Daten und was sind für Sie «optimale Werkzeuge»?

**GS** Unter einem Grundangebot stellen wir uns vor, dass möglichst viele Publikationen, Zeitschriften, Bücher, Archivbestände und Forschungsdaten digital verfügbar und breit zugänglich werden. Optimale Werkzeuge zu haben, würde heissen, dass es zwischen den verschiedenen Anwendungen Schnittstellen gibt und dass die Hochschulen sich zusammenschliessen, z.B. um gemeinsame IT-Tools anzubieten. Das würde wissenschaftliches Arbeiten erleichtern.

**BK** Ein freier Zugang zu Zeitschriften? Ist das realistisch?

**GS** Open-Access-Zeitschriften sind schon heute frei zugänglich. Die SAGW unterstützt einige davon. Schwieriger ist das bei lizenzpflichtigen Titeln. Im Projekt «Nationallizenzen» will das Konsortium der Schweizer Hochschulbibliotheken für verschiedene digitale Zeitschriften und Datenbanken Lizenzen aushandeln, die national gültig sind. Bei diesen Produkten werden alle Hochschulen zumindest in Bezug auf die

*«Grosse Wissenschaftsverlage haben heute ein Quasi-Monopol, was die Auseinandersetzung um faire Preise deutlich erschwert.»*

Archive gleichgestellt sein. In diesem Projekt wollen wir die Verhandlungsposition der Bibliotheken gegenüber den Verlagen stärken und insgesamt bessere Konditionen erreichen. Grosse Wissenschaftsverlage haben heute ein Quasi-Monopol, was die Auseinandersetzung um faire Preise deutlich erschwert. Für die Budgets der Bibliotheken ist das eine grosse Belastung.

**BK** Das tönt interessant, aber weshalb sollten die Verlage dem zustimmen?

**GS** Weil Bibliotheken wichtige Kunden sind und die Wissenschaftsverlage ihrerseits auf die Bibliotheken angewiesen sind. Es gibt einen Verhandlungsspielraum. Im aktuellen Projekt steht der Einkauf von

digitalen Archiven im Zentrum. Diese sollen an die Lizenzen für die laufend erscheinenden Publikationen angebunden werden, so dass die Archive laufend mit

*«Zukünftig wird man eher für die Publikation eines Artikels bezahlen, nicht, um ihn lesen zu dürfen.»*

den neuen Artikeln ergänzt werden. Das heisst, die neuen Artikel sind zwar nicht sofort für alle Hochschulen zugänglich, aber doch immerhin nach einer gewissen Zeit. Im Zeitalter von Open Access müssen Verlage und Bibliotheken ohnehin umdenken. Zukünftig wird man eher für die Publikation eines Artikels bezahlen, nicht, um ihn lesen zu dürfen. Die Universitäten und Hochschulen werden weniger in die Subskription investieren und dafür mehr in die Artikel.

**BK** Braucht es denn überhaupt noch Verlage?

**GS** Ja, sicher. Die Kernkompetenz eines guten Verlags ist die Qualitätssicherung durch Selektion und durch das Lektorat sowie der Vertrieb. Das braucht es weiterhin. Zur Diskussion stehen die Geschäftsmodelle. Es gibt inzwischen z.B. reine Open-Access-Verlage.

**BK** Wie können Forschungsdaten zuverlässig gefunden werden? Über eine gemeinsame Einstiegsseite?

**GS** Sie meinen, wie Swissbib ([www.swissbib.ch](http://www.swissbib.ch)) für Bibliotheksbestände? Es gibt internationale Projekte wie beispielsweise DataCite (<https://www.datacite.org>), und es gibt unser Pilotprojekt Open Research Data ([www.openresearchdata.ch](http://www.openresearchdata.ch)). Die Herausforderung bei Forschungsdaten ist aber immer noch, sie überhaupt zu sichern und verfügbar zu machen. Nicht alle Forschungsdaten können zudem offen sein, und in den Geisteswissenschaften wird diskutiert,

*«Heute will man Daten so aufbereiten, dass sie über verschiedene Wege auffindbar sind.»*

was sie für ihre Disziplinen als Forschungsdaten verstehen wollen. Aber die Idee des allumfassenden, zentralen Sucheinstiegs ist ein überholter Ansatz. Heute will man Daten so aufbereiten, dass sie über verschiedene Wege auffindbar sind. Publikationen z.B. über die grossen Suchmaschinen. Es ist zudem wichtig,

mit offenen Standards zu arbeiten, damit Quellen auch für das Data Mining genutzt werden können.

**BK** Was ist Data Mining?

**GS** Beim Data Mining geht es darum, mit Hilfe der Informatik Strukturen und Trends in grossen Datenmengen zu suchen. Das geht in der Regel allerdings nur, wenn die Daten in interoperablen Formaten und Standards gespeichert sind. Die Repositories beispielsweise können recht gut miteinander vernetzt werden.

**BK** Wie können einheitliche Standards erreicht werden?

**GS** Standards entwickeln sich laufend und leben von möglichst breiten Interessengemeinschaften. Es macht keinen Sinn, Standards allein für die Schweiz zu definieren, da müssen wir uns an der internationalen Zusammenarbeit orientieren. Die Position der Schweiz dafür zu verbessern, war einer der Gründe, das Programm zu lancieren.

**BK** Wer ist in der Schweiz für solche Standards zuständig?

**GS** In der Schweiz gibt es verschiedene Brückenköpfe für Standards. Z.B. Archive haben bei der Langzeitarchivierung und im Datenmanagement Vorarbeit geleistet, die grossen Bibliotheken bei der Retrodigitalisierung oder bei der Vergabe von Standardnummern wie dem DOI (Digital Object Identifier, Anm. d. Red.). SWITCH ist die Anlaufstelle für Identity Management und kann mit den Erfahrungen bei den IT-Strukturen beitragen.

**BK** «Dienste» ist eine der vier Hauptstossrichtungen, die Anfang 2014 in der «Nationalen Strategie» für die «Bündelung der Kräfte in der wissenschaftlichen Information» genannt werden, neben Publikationen, eScience und Basis. Was erhoffen Sie sich von diesem Bereich?

**GS** Der Umsetzungsbereich «Dienste» ist ursprünglich für den eLearning-Bereich entstanden. Er sollte die Öffnung hochschuleigener Lösungen für weitere Teilnehmer unterstützen. Diese Idee wurde dann ausgeweitet: Das Programm würde z.B. das Teilen von Repositories oder von Publishing-Plattformen unterstützen, es würde sowohl die Öffnung bestehender Lösungen wie auch die Zustiegskosten neuer Teilnehmer

fördern. Bisher wurden allerdings nur wenige Projekte bewilligt. Hier sind wir etwas ratlos. Die E-Learning-Community konnte mit dem Programm bisher nicht abgeholt werden. Weshalb, wissen wir nicht. Vielleicht wollen die Hochschulen hier gar keine gemeinsamen Lösungen. Das wäre dann auch eine Erkenntnis.

**BK** Was unternehmen Sie, um diese Community doch noch an Bord zu holen?

**GS** Wir werden für die nächste Finanzierungsperiode prüfen, welche Bereiche nicht so gut funktionierten und weshalb. Es kann auch sein, dass wir die Rahmenbedingungen in einigen Bereichen anpassen müssen. Eine solche Überprüfung werden wir auch für den Umsetzungsbereich Open Access machen, wo bisher kaum überzeugende Projekte eingereicht wurden.

**BK** Welches sind denn Ihre Leuchtturmprojekte?

**GS** Von Leuchttürmen zu sprechen, ist vielleicht etwas gefährlich, aber wir erhoffen uns viel von den Projekten Data Life-Cycle Management (DLCM, siehe Beitrag auf Seite 48), Swiss Library Service Platform (SLSP) und Swiss edu-ID.

**BK** Gibt es Unterschiede im Infrastrukturbedarf zwischen Geistes- und Sozialwissenschaften und anderen Disziplinen? Wenn ja, welche? Wenn nein, was ist allen Disziplinen gemeinsam?

**GS** Natürlich gibt es Unterschiede. In der Bioinformatik oder der Hochenergiephysik drehen sich Infrastrukturen häufig um grosse Rechenleistungen, um Big Data und Cloud Computing, während in den Geisteswissenschaften eher die Auseinandersetzung mit Quellen aus Bibliotheken und Archiven im Mittelpunkt steht. Es gibt aber auch Gemeinsamkeiten. Die Informatik eröffnet zwar neue Forschungsfelder

*«Die Informatik eröffnet zwar neue Forschungsfelder und -methoden, ist in der Regel jedoch nicht die Kernkompetenz der Forschenden.»*

und -methoden, ist in der Regel jedoch nicht die Kernkompetenz der Forschenden. Also braucht es den Dialog zwischen Forschung, Lehre und den Dienstleistern. Es braucht Kompetenzzentren wie die SAGW, um Dienstleistungen aufzubauen, die zwischen den heutigen Rechen-

zentren, den Gedächtnisinstitutionen und den Universitäten angesiedelt sind.

**BK** Berücksichtigt das Programm «Wissenschaftliche Information» die unterschiedlichen Bedürfnisse der Disziplinen?

**GS** Das Programm will eine nationale «Service-Drehscheibe» mit Diensten aufbauen, die allen Disziplinen zugutekommen. Mit der nationalen Strategie haben wir versucht, einen möglichst überlappenden Bedarf aller Disziplinen «festzunageln». Die 35 Umsetzungsmassnahmen in unserem «White Paper» sind ein Anforderungskatalog, der im Dialog mit den Stakehol-

*«Das Programm will eine nationale «Service-Drehscheibe» mit Diensten aufbauen, die allen Disziplinen zugutekommen.»*

dern erstellt wurde. Für rein fachspezifische Projekte ist das Programm nicht zuständig. Wobei wir schon unterschiedliche Angebote haben; so bedient die Hauptstossrichtung «Publikationen» eher die Geisteswissenschaften. Die Grenze ist nicht immer einfach festzulegen und ist manchmal auch verhandelbar: Wir möchten vor allem gute, wegweisende Projekte fördern.

**BK** Das Programm vergibt Beiträge als Anschubfinanzierung. Wie kann die nachhaltige Verankerung der unterstützten Projekte gewährleistet werden?

**GS** Das ist die zentrale Frage. Am aussichtsreichsten ist die Verankerung, wenn unsere Fördermittel eine Institution dazu ermutigen, ein Vorhaben, das sie sowieso gerne umgesetzt hätte, etwas grösser, ambitionierter und partizipativer vorzulegen. Wenn also der Projekthalt aus einer Kernkompetenz der Antragsteller wächst. Das

*«Projekte, die allein aus Programmmitteln finanziert werden, haben tendenziell ein Verankerungsproblem.»*

ist z.B. bei der Swiss edu-ID von SWITCH oder beim Projekt «Swiss Library Service Platform» der Schweizer Bibliotheksverbände der Fall. Die Entwicklungen in diesem Jahr haben gezeigt, dass es auch bei der SAGW ein grosses Potenzial gibt.

Projekte, die allein aus Programmmitteln finanziert werden, haben tendenziell ein Verankerungsproblem. Die Frage ist, ob die Hochschulen überhaupt bereit sind, einander als Dienstleister unter die Arme zu grei-

fen. Das ist ein klares Ziel des Programms. Aber da müssen wir noch Überzeugungsarbeit leisten

**BK** Welches Potenzial haben die Aktivitäten der SAGW im Bereich Digital Humanities?

**GS** Es gibt verschiedene Bereiche, bei denen sich die Pläne der SAGW mit jenen des Programms decken. Open Access, DaSCH, Retrodigitalisierung, Editionen und Infrastrukturen sind Themen, die auch für unser Programm von Interesse sind. Je stärker die Parteien sind, die etwas umsetzen wollen, desto grösser ist die Chance, gemeinsam Projekte zu entwickeln.

---

### Weitere Informationen

SUK-Programm 2013-2016 P-2 «Wissenschaftliche Information: Zugang, Verarbeitung und Speicherung»:  
<http://www.swissuniversities.ch/isci>

---

### Zur Interviewpartnerin

#### Gabi Schneider



Gabi Schneider, M.A., Geisteswissenschaftlerin und wissenschaftliche Bibliothekarin, ist stellvertretende Leiterin des Programms «Wissenschaftliche Information: Zugang, Verarbeitung und Speicherung» (SUK P-2). Ihr Arbeitsort ist die Universitätsbibliothek Basel.

## Forschungsdaten in den Geisteswissenschaften

*Patrick Sahle, Cologne Centers for eHumanities*

*Der Begriff der Forschungsdaten ist schwierig. Er wird nirgendwo explizit und vollständig definiert. Alle aktiven und passiven Wortverwender haben eine implizite Vorstellung, die sehr unterschiedlich sein kann. In der Diskussion der letzten Jahre hat sich das Verständnis von «Forschungsdaten» weiterentwickelt, wobei hier auch ein Emanzipationsprozess der Geisteswissenschaften zu beobachten ist.*

43

### Wissenschaftsfelder

In einer differenzierenden Annäherung fällt zunächst ein Unterschied zwischen diesem Wissenschaftsfeld und den übrigen Bereichen der Forschung auf. In den Naturwissenschaften, aber z.B. auch in den Sozialwissenschaften, scheinen Forschungsdaten relativ klar fassbar zu sein, weil es sich hier fast immer um Mess- und Erhebungsdaten handelt, die in einem definierten Rahmen (Experiment, Messung, Befragung) nach einem festen Modell für eine bestimmte Fragestellung gewonnen werden. In den Geisteswissenschaften ist die Lage hingegen komplexer, weil sich hier die meisten Fragestellungen auf die Untersuchung von Äusserungen und Artefakten des menschlichen Denkens und der menschlichen Kultur richten. Auch aus diesen können zwar planvoll und im Rahmen eines Beobachtungsmodells Daten extrahiert werden. Typischer ist aber eine Herangehensweise, bei der die Objekte des Kulturerbes als Ganzes in eine neue Repräsentationsform überführt und dabei zunehmend tiefer erschlossen und durchdrungen werden, wobei sie zugleich Gegenstand der hermeneutischen Deutung und Analyse sind.

Während die meisten Wissenschaftsfelder also mit vorgängigen, präskriptiven Modellen als Grundlage der Gewinnung von Forschungsdaten arbeiten, sind die Geisteswissenschaften vor allem von einer nachgängigen, deskriptiven Modellierung geprägt – wenn sie überhaupt Modellierung als Explizierung von Wahrnehmungsstrukturen betreiben. Dieser häufig inkrementelle, ungleichmässige und ungleich komplexere Prozess hat Konsequenzen sowohl für das Begriffsver-

ständnis «Forschungsdaten» als auch für den Umgang mit ihnen im Forschungszyklus und ihre Verwaltung und dauerhafte Bereitstellung.

### **Forschungszyklus**

Die «Research Data Life Cycles» sehen nur auf den ersten Blick und bei starker Reduktion in allen Fachbereichen ähnlich aus: Forschungsfrage – Forschungsdaten – Analyse – Ergebnispublikation. In den Naturwissenschaften lassen sich die einzelnen Schritte noch relativ klar voneinander unterscheiden. Insbesondere sind die Forschungsdaten sauber von den Publikationen getrennt. Dies erklärt, warum es recht erfolgreiche Repositorien gibt, die für zahlreiche wissenschaftliche Zeitschriften konsequent die zu einem Aufsatz zugehörigen Forschungsdaten archivieren. Ein Beispiel hierfür ist «Dryad» ([datadryad.org](http://datadryad.org)), das die Daten zu Beiträgen in über 400 Journalen aufnimmt – darunter aber keines aus den Geisteswissenschaften.

In den weniger empirisch und mehr hermeneutisch arbeitenden Fächern scheint eine scharfe Trennung der Schritte im Forschungszyklus schwieriger, ihre klare Abfolge fraglicher und die Bestimmung von Forschungsdaten unklar zu sein. Wenn man hier in einer Strategie der Naivität zunächst auf eine naheliegende, weite Grunddefinition von Forschungsdaten zurückgeht, dann lautet diese: «Forschungsdaten sind alle Daten, die im Forschungsprozess eine Rolle spielen». Das umfasst dann alle Arten von «Quellen», die materiellen Objekte der Kultur ebenso, wie ihre digitalen Repräsentationsformen, verschiedene Arten zusätzlichen Wissens dazu, intermediäre Verarbeitungsformen und schliesslich Veröffentlichungen, die ja selbst auch Grundlage der Forschung sind. Eine klare Trennung von Primärdaten, Sekundärdaten und Publikationen erscheint schwierig, wenn die erschlossenen Quellen z.B. als Kataloge, Register oder Editionen selbst das Ergebnis der Forschung und ihre Publikation sind. Die Unterschiede im Begriffsverständnis werden aber auch durch unterschiedliche Zwecke von Forschungsdaten und ihrer Bewahrung deutlich: Während es in der Medizin, den Naturwissenschaften oder den Sozialwissenschaften vor allem um die Sicherung der Transparenz und Nachprüfbarkeit von behaupteten Forschungsergebnissen geht, sollen die Forschungsdaten in den Geisteswissenschaften als digitale Repräsentanten kultureller Objekte eine permanente Weiternutzung für alle möglichen Fragestellungen ermöglichen und für die Anlagerung weiteren erschliessenden oder deutenden Wissens offen sein.

### **Forschungsdaten**

Wegen dieser Unterschiede, wegen der Schwierigkeiten mit einem einfachen Datenbegriff und der Gefahr, dass aus einem sachfremden Begriffsverständnis die falschen Lösungen entwickelt werden, wird seit einiger Zeit diskutiert, ob nicht insbesondere für den Bereich der Geisteswissenschaften eher von digitalen Objekten oder von digitalen Ressourcen gesprochen werden sollte. Mit dem Begriff der «Daten» ist zu sehr ihre Abtrennbarkeit von umgebenden Systemen, ihre Unverbundenheit und eine Gleichmässigkeit in Datensätzen konnotiert. Dagegen werden bei «digitalen Objekten» der Bezug zu möglichen materiellen Vorlagen, die Zusammengehörigkeit komplexer Daten und ein intellektueller Rahmen mitgedacht: Daten bilden konzeptionelle Einheiten ab, die in sich komplex und mit ihrer Aussenwelt verbunden sind und die kontinuierlich verändert und erweitert werden können. Wenn von «digitalen Ressourcen» die Rede ist, dann ist der Rahmen noch weiter gefasst. Hier geht es um die Idee, dass Daten und Objekte häufig nicht von ihren technischen Umgebungen der Erstellung, Bearbeitung und öffentlichen Bereitstellung getrennt werden können, ohne sinnlos oder unbenutzbar zu werden. Spätestens mit diesem Konzept betreffen Forschungsdaten nicht mehr nur einen einzelnen Schritt im Forschungsprozess, sondern wird der Forschungszyklus insgesamt zum Rahmen der Beschäftigung mit ihnen (siehe den Beitrag von Pierre-Yves Burgi, S. 48).

### **Forschungsdatenarten**

In den Geisteswissenschaften kommen durchaus auch Mess- und Erhebungsdaten vor, die scharfen, präskriptiven Modellen folgen. Zu denken ist hier z.B. an die experimentellen Forschungen der Phonetik. Über die Fächer hinweg werden manchmal gezielt Informationen aus Quellen extrahiert, die dann in strukturierten Datenbanken verwaltet werden. Dies ist aber eher die Ausnahme als der Regelfall. Typischer sind Korpora z.B. der Sprachwissenschaften, die gleichmässige Textdaten, im einfachsten Fall «plain text» enthalten. Etwas komplexer sind dann schon die Grabungsdaten der Archäologie, die Bilddatenbanken z.B. der Kunstgeschichte, die audiovisuellen Aufzeichnungen der Sprachdokumentation oder der Ethnologie oder die strukturierten Texte der Literaturwissenschaften. Geradezu paradigmatisch ist dann aber das Modell der fortschreitenden und umfassenden Digitalisierung, die auf eine Repräsentation von Objekten zielt, die von der mechanischen Reproduktion (Faksimilierung) ausgeht, eine vertiefende Erschliessung nahelegt und mindestens potenziell eine Edition als Endpunkt hat (siehe die Beiträge von Hundt, S. 51, und Hanke, S. 53

in diesem Heft). Solche digitalen Objekte sind als Forschungsdaten schwer fassbar, weil sie aus verschiedenen Schichten und miteinander verbundenen Komponenten bestehen und damit äusserst komplex werden. Tatsächlich fehlen ihnen eine eindeutige Grenze und ein eindeutiger Zusammenhalt. Perspektivisch, insbesondere unter dem Paradigma des «Semantic Web», zerfallen sie eher in «semantische Partikel». In einzelne Aussagen, die zwar durch die eindeutige Identifikation und Adressierung von Objekten und durch eine Bindung der Aussagen über sie an explizite Ontologien sehr konkret sein können. Wenn aber das Wissen als freie und an keinen einzelnen Erkenntniszweck gebundene «Annotation» verstanden wird, dann ist unklar, was einen bestimmten Satz an Forschungsdaten ausmachen sollte und wie er zu fassen wäre (siehe die Beiträge von Bickel und Graf, S. 57, und Āurĉo, S. 61, in diesem Heft). «Big Data» ist auch in den Geisteswissenschaften schon lange angekommen. Verschiedene Arten von Korpora, die für die Forschung erstellt werden, haben einen Umfang, der nur noch algorithmische Formen der Analyse erlauben. Typisch für die Geisteswissenschaften sind daneben aber vergleichsweise kleine, heterogene, komplexe digitale Objekte und Ressourcen. Die Zukunft wird zeigen, ob sich diese beiden Paradigmen der «big data analytics» und der tiefen semantischen Annotation komplexer Ressourcen in einer Zukunft methodisch verbinden lassen, in der Forschungskorpora erst durch die übergreifende Nutzung verteilter Informationen entstehen.

### Konsequenzen

Für die Bereitstellung und anhaltende Benutzung von Forschungsdaten führen die verschiedenen Datenarten zu verschiedenen Lösungsmöglichkeiten, die nicht nur technisch, sondern auch konzeptionell und institutionell zu betrachten sind. Auf der technischen Seite gilt die einfache Faustformel: Je homogener, je einfacher, je standardisierter Forschungsdaten sind, umso leichter können sie mit generischen technischen Lösungen verwaltet werden. Allgemeine oder gegenstandsspezifische Repositorien können Daten, Objekte und Korpora verwalten, solange diese von den Systemen ihrer Erstellung oder Bereitstellung abtrennbar sind. Häufig führen Projekte in den Geisteswissenschaften aber zu komplexen, idiosynkratischen Ergebnissen, die ohne ihre eigene Softwareumgebung weder verständlich noch mit vertretbarem Aufwand benutzbar wären. Hier bleibt unter Umständen manchmal nur die dauerhafte Pflege der «lebenden» Systeme. Konzeptionell und politisch stellt sich die Frage, welche Forschungsdaten eigentlich welche «Pflegestufe» erfordern und welche man ihnen zubilli-

gen will. Hier kann unterschieden werden zwischen der reinen Archivierung von Daten, ihrer direkten Bereitstellung, ihrer Vernetzung und Einspeisung in globalere Informationssysteme, ihre weitergehende Kuratierung, die Konservierung von Anwendungen und schliesslich die Aufrechterhaltung von Arbeitsumgebungen zur weiteren Veränderung, Bearbeitung und Anreicherung mit weiteren Informationen. Politisch muss auch die institutionelle Frage geklärt werden, wer eigentlich für die Forschungsdaten und ihre Bereitstellung dauerhaft zuständig sein soll. Hier liegt zwar der BAM-Sektor als Ausgangspunkt für digitale Repräsentationsformen kultureller Objekte nahe, Fragen der technischen Komplexität und der fachlichen Durchdringung legen aber weitere Akteure nahe. Hier ist an Rechenzentren zu denken, die aber normalerweise weder die inhaltliche Fachkompetenz für komplexe geisteswissenschaftliche Datenbestände vorhalten, noch institutionell einen «Ewigkeitsauftrag» wie die Archive haben. Möglicherweise liegt die Lösung deshalb in dedizierten Datenzentren, die als neue institutionelle Rahmenorganisationen die vielfältigen Aufgaben dadurch meistern, dass sie die Kompetenzen der Bestände bewahrenden, der technischen Infrastruktur bereitstellenden und der methodischen und fachwissenschaftlichen Kompetenz produzierenden Akteure zusammenbringen (siehe den Beitrag von Lukas Rosenthaler, S. 46, in diesem Heft).

### Weitere Informationen

Weiterführende Hinweise befinden sich auf der Website: <http://www.sagw.ch/bulletin>

### Zum Autor



#### Patrick Sahle

Dr. Patrick Sahle ist Geschäftsführer des Cologne Centers for eHumanities (CCeH) an der Universität zu Köln. Er ist dort u.a. mit der Koordinierungsstelle Digital Humanities für die Betreuung der digitalen Transformation der Langzeitvorhaben der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste betraut. Ausserdem koordiniert

er den Aufbau des Data Center for the Humanities (DCH), eines Datenzentrums für die Geisteswissenschaften, das ebenfalls an der Philosophischen Fakultät der Universität zu Köln angesiedelt ist und sich um die dauerhafte Bereitstellung von Forschungsdaten und den andauernden Betrieb von Forschungsanwendungen kümmern soll.

## Geisteswissenschaftliche Daten sichern und verfügbar halten

Lukas Rosenthaler, Digital Humanities Lab,  
Universität Basel

46

*Der Begriff der «Digital Humanities», der in der jüngeren Vergangenheit schon fast inflationär gebraucht wird, zeigt eines mit grosser Deutlichkeit: Auch in der geisteswissenschaftlichen Forschung entstehen immer mehr digitale Daten, welche in sich selbst einen grossen Wert darstellen.*

Datenbanken, welche im Rahmen von Forschungsprojekten primär als – privates oder lokales – Hilfsmittel entstehen; Digitalisate, welche kommentiert und beschrieben werden; Transkriptionen für kritische Editionen, welche mit digitalen Hilfsmitteln kodiert werden (z.B. in TEI/XML) etc. All dies sind digitale Forschungsdaten, welche auch nach Ende der Finanzierung eines Forschungsprojektes der Forschungsgemeinschaft zur Verfügung stehen sollten (auch im Rahmen der «Open Access»-Politik). Die Forderung, dass die nachhaltige Verfügbarkeit solcher Forschungsdaten garantiert werden soll, spiegelt sich sowohl auf nationaler wie auch internationaler Ebene in den Diskussionen und dem «Kleingedruckten» der Institutionen ab, welche die Forschung finanzieren.

### Ein Datenzentrum mit Beratung

Die SAGW hat zusammen mit dem SBFI vor rund zweieinhalb Jahren die Initiative ergriffen und ein Pilotprojekt für ein «Daten- und Dienstleistungszentrum für geisteswissenschaftliche Forschungsdaten» (engl. Data and Service Center for the Humanities, DaSCH) ins Leben gerufen, das nun als nationale Forschungsinfrastruktur verstetigt werden soll. Mit im internationalen Vergleich sehr bescheidenen Mitteln ist es gelungen, exemplarisch die Funktionsweise einer solchen Dienstleistungsinstitution erfolgreich aufzubauen und zu testen. Von Beginn weg war das Konzept, nicht nur ein Repository für die Daten zur Verfügung zu stellen (das dann allzu oft zu einem Datenfriedhof verkommt), sondern auch Beratungsfunktionen in den Bereichen Ablieferung, Datenmanagement und nachhaltige Nutzung anzubieten.

### Kriterien und Anforderungen

Die «Daten» dürfen nicht isoliert von den Werkzeugen zu deren Nutzung betrachtet werden. Die Werkzeuge sind oft essenziell, um die Daten später revalorisieren zu können. Dieses Problem teilen die Geisteswissenschaften übrigens mit anderen Wissenschaften wie z.B. der Hochenergiephysik, wo die Auswertungsprogramme selbst Teil der zu bewahrenden Daten darstellen.

Ein weiteres Kriterium, das entscheidend für den Erfolg oder das Scheitern eines Datenzentrums für die langfristige Vorhaltung von Forschungsdaten beiträgt, ist die Nähe zu den Forscherinnen und Forschern. Deshalb wurde für das DaSCH ein Organisationsmodell gewählt, das einerseits eine zentrale Funktion (Koordination, Softwareentwicklung und «second level support») vereint, während für die direkte Interaktion an jeder teilnehmenden Institution lokale «Satelliten» installiert werden, welche den «first level support» für Forscherinnen und Forscher übernehmen und in der Lage sind, die Datenübernahme zu unterstützen und kleine Eigenentwicklungen und projektspezifische Anpassungen vorzunehmen

### Unterstützung von Forschungsprojekten

Grundsätzlich sind drei Varianten der Unterstützung von Forschungsprojekten möglich:

#### *ab ovo*

Ein Forschungsprojekt wird von der Entstehung her begleitet. Der erste Kontakt findet vor der Projekteingabe statt, das DaSCH unterstützt das Projekt in der Erstellung des «data management plan», der in Zukunft jede Projekteingabe begleiten sollte, und der Planung von «Open Access» bzw. «Open Data» (Creative-Commons-Lizenzen etc.). Falls das DaSCH das Projekt aktiv im Bereich IT unterstützen soll (z.B. durch die Entwicklung spezieller Software in grösserem Rahmen), so müssen die entsprechenden Mittel mit beantragt

werden. Während der gesamten Projektdauer unterstützt und berät das DaSCH das Projektteam in Fragen des Datenmanagements und der Nachhaltigkeit. Bei Projektende stehen die Forschungsdaten in der Infrastruktur des DaSCH zur Nachnutzung bereit.

#### *in vivo*

Erst im Verlaufe eines Forschungsprojektes wird klar, dass Forschungsdaten angefallen sind und die nachhaltige Nutzung gefordert wird. In diesem Falle unterstützt das DaSCH die Transformation der Daten in Formate und Datenmodelle, welche die spätere Überführung zur langfristigen Nutzung ermöglichen. Dabei wird nicht nur auf die aktuellen Bedürfnisse der laufenden Forschung Rücksicht genommen, sondern auch die aktuellen Nutzungsmöglichkeiten verbessern sich dadurch erheblich. Bei Projektende stehen die Forschungsdaten in der Infrastruktur des DaSCH zur Nachnutzung bereit.

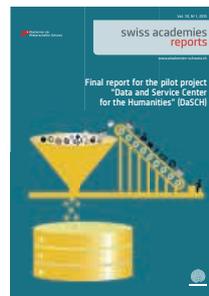
#### *post mortem*

Sind ein Forschungsprojekt und seine Finanzierung abgeschlossen, aber es liegen digitale Forschungsdaten vor, die für weitere Forschungsprojekte von hohem Wert sind, so versucht das DaSCH, diese Daten zu «retten» und in die zentrale Infrastruktur so zu überführen. Diese Variante ist offensichtlich nicht optimal, aber wird noch einige Zeit (in abnehmender Frequenz) eine wichtige Aufgabe des DaSCH darstellen. Diese Aufgabe kann unter Umständen äusserst komplex sein und stellt hohe Anforderungen an die IT-Kenntnisse, aber auch an das Verständnis für das jeweilige Fachgebiet. Oft ist – soweit dies machbar ist – die Rücksprache mit den ehemaligen Projektteilnehmern unerlässlich, um zum Erfolg zu kommen.

Es ist offensichtlich, dass diese Variante die meisten Probleme aufwirft und wenn immer möglich vermieden werden soll.

Die Pilotphase hat exemplarisch gezeigt, dass das DaSCH in der Lage ist, alle drei Varianten erfolgreich umzusetzen. Mit der geplanten Verstetigung kann diese Dienstleistung für alle Forscherinnen und Forscher im nationalen Umfeld zur Verfügung gestellt werden und wird einen wesentlichen Beitrag leisten, «Datenfriedhöfe» weitgehend zu vermeiden. Selbstverständlich verfügt das DaSCH über ausgezeichnete internationale Kontakte und interagiert sehr eng mit entsprechenden Initiativen im Ausland.

### Weitere Informationen



Projekt DaSCH und Schlussbericht Pilotprojekt «Daten- und Dienstleistungszentrum»: <http://www.sagw.ch/sagw/laufende-projekte/ddz.html>

### Zum Autor

#### Lukas Rosenthaler



PD Dr. Lukas Rosenthaler doktorierte 1987 in Physik an der Universität Basel. Er habilitierte 2008 in der Philosophisch-Historischen Fakultät zum Thema «Digitale Langzeitarchive» und leitet heute das «Digital Humanities Lab». Sein Forschungsgebiet sind «Virtuelle Forschungsumgebungen für die Geisteswissenschaften». Zusätzlich ist er auch im Auftrag der SAGW und des SBFI für den Aufbau eines «Daten- und Dienstleistungszentrums für geisteswissenschaftliche Forschungsdaten» verantwortlich.

## Data Life-Cycle Management: The Swiss Way

Pierre-Yves Burgi, directeur SI adjoint, Université de Genève

48

*L'information contenue dans les données de recherche constitue un bien précieux pour le chercheur. En revanche, le chercheur ignore trop souvent l'importance du cycle de vie de ces données, un concept qui englobe une multitude de facettes, dont certaines dépendent de la discipline considérée.*

D'une manière générale, ce cycle débute par l'acquisition de données dites «brutes»<sup>1</sup>. Par la suite ces données sont analysées pour les confronter à des hypothèses et modèles. Les résultats de ces études conduisent généralement à des publications, un jalon important dans la vie des données, tout comme la fin d'un projet (ou de son financement). En effet, ces données sont alors souvent délaissées, typiquement sur l'espace de stockage du chercheur, et/ou d'un serveur institutionnel, pour être rapidement oubliées, en omettant de les valoriser dans d'autres contextes.

### Collecter, sélectionner et sécuriser

Le phénomène de «numérisation» de la recherche s'accéléralant, il devient urgent d'aborder le cycle de vie complet des données. A titre d'exemple, le radiotélescope SKA (Square Kilometer Array<sup>2</sup>) devrait, dès 2024, collecter chaque jour 14 exabytes<sup>3</sup> de données dont 1 petabyte sera sauvegardé quotidiennement. Cet exemple illustre les besoins gigantesques en stockage que certains consortia internationaux devront couvrir

dans un futur proche. La plupart des laboratoires de recherche en Suisse ont certes des besoins plus modestes, couramment décrits comme «the long tail»<sup>4</sup>. Reste que la gestion de cette information, aussi minime soit-elle, est sujette au second principe de la thermodynamique selon lequel le désordre (ou entropie) a tendance à irrémédiablement augmenter, avec pour conséquence que toute donnée devient inévitablement corrompue avec le temps. Cela implique que pour conserver l'information il faut contrecarrer ce désordre naturel. En l'occurrence, l'ordre peut être rétabli par la capacité de la matière à effectuer des calculs pour rétablir une information corrompue<sup>5</sup>. Cependant ce mécanisme a des limites, ne serait-ce que par l'énergie requise pour son fonctionnement. Ainsi une sélection de l'information devient une nécessité, tout comme la durée de l'archivage dans le temps.

### Origine du projet et ses intervenants

Le projet Data Life-Cycle Management (DLCM) s'est concrétisé dans le contexte du programme national suisse CUS P-2 (2013–2016) «Information scientifique: accès, traitement et sauvegarde». En novembre 2013, l'auteur de cet article a initié des contacts avec des experts du domaine DLCM dans les universités suisses dans le but de former un partenariat avec l'objectif de déposer une proposition de projet CUS P-2 en 2014. Au terme de cette démarche, un partenari-

---

1 Une dénomination ambiguë par le fait que les instruments scientifiques prétraitent souvent les données à différents niveaux. La NASA a par exemple défini 6 niveaux de prétraitement ([science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products](https://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products)).

2 [www.skatelescope.org](http://www.skatelescope.org)

3 1 Exabyte (EB) correspond à 1000 Petabytes (PB), soit 1 000 000 Terabytes (TB). Aujourd'hui, un disque de 1 TB coûte environ CHF 100.

4 «The long tail» fait référence à la loi de Pareto selon laquelle 80% des effets sont le produit de 20% des causes (e.g., [fr.wikipedia](https://fr.wikipedia.org/wiki/Principe_de_Pareto)).

---

[org/wiki/Principe\\_de\\_Pareto](https://fr.wikipedia.org/wiki/Principe_de_Pareto)). Pour les données de la recherche, cela se traduit par le postulat que 20% des producteurs de données comptent pour 80% du volume produit; «the long tail» rappelle qu'il ne faut pas oublier les 80% des chercheurs qui produisent plus modestement des données, mais constituent une population importante ayant des besoins avérés en DLCM.

5 Sachant que cette même matière est aussi sujette à l'augmentation de l'entropie et nécessite aussi un traitement pour être maintenue opérationnelle.

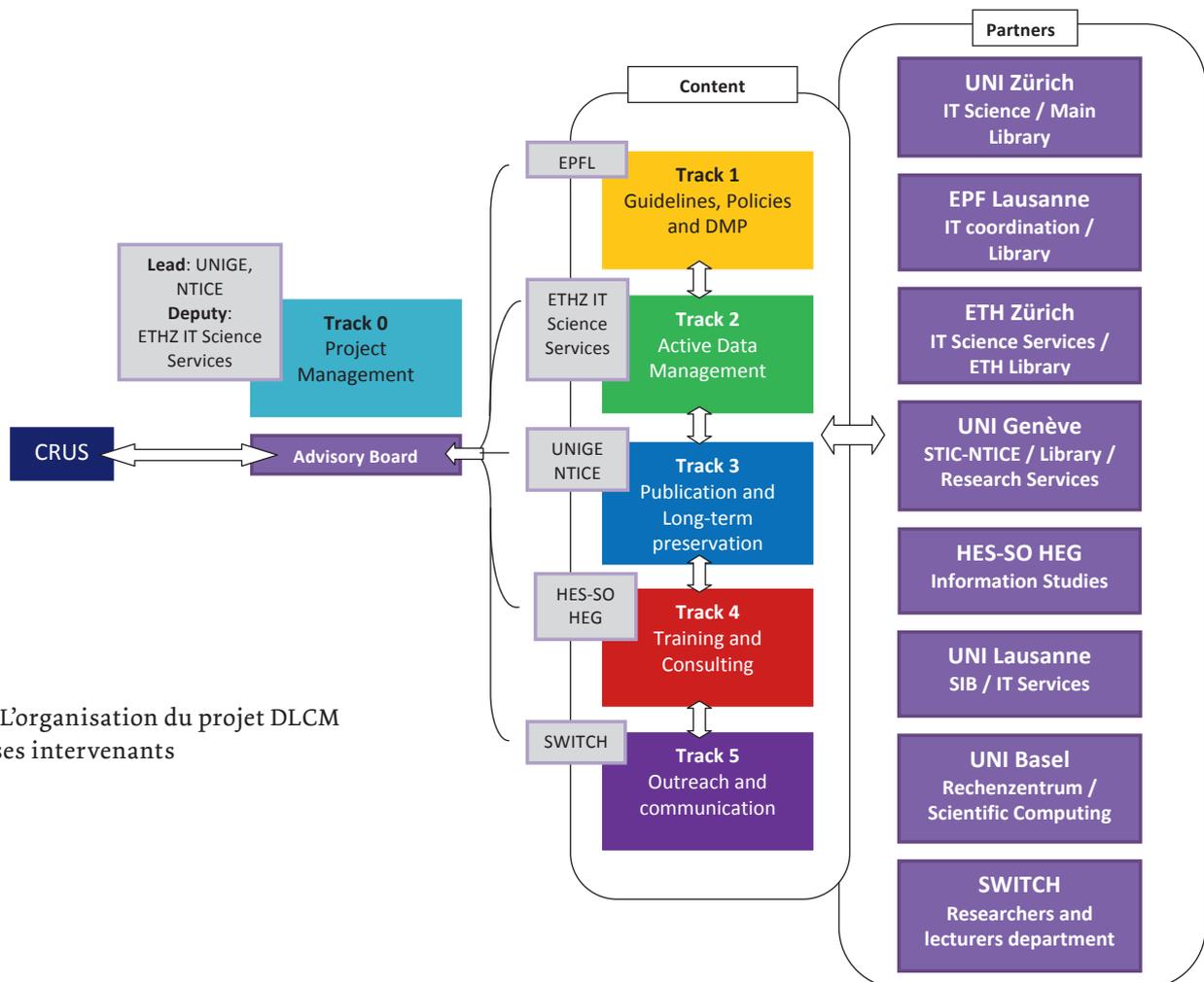


Fig. 1 L'organisation du projet DLCCM avec ses intervenants

at entre les deux écoles polytechniques fédérales (ETH-Z et EPFL), les universités de Zurich, Bâle, Lausanne, et Genève, la HEG de la HES-SO, et SWITCH s'est finalement établi (Fig. 1), avec à la clé une proposition aboutie et soumise en février 2015. Suite à l'acceptation de cette proposition en juillet 2015, le démarrage officiel du projet a eu lieu le 1<sup>er</sup> septembre 2015.

### Objectifs du projet

Concrètement, le projet DLCCM aborde tous les aspects du cycle de vie des données, à savoir l'acquisition (e.g., LIMS<sup>6</sup>, ELN<sup>7</sup>), la normalisation et description des données en vue d'un référencement pérenne, leur traitement, et conservation sur le long terme (Fig. 2). Ces différentes étapes du cycle de vie sont couvertes dans cinq parties distinctes (Fig. 1 et 2):

Le «track 1» traite de la dimension politique du DLCCM («guidelines, politiques»), de plus en plus présente dans les instances qui financent la recherche (e.g., Horizon 2020, Fonds national suisse, National Institutes of Health, etc.), et qui nécessite l'écriture de «data management plans» (DMP);

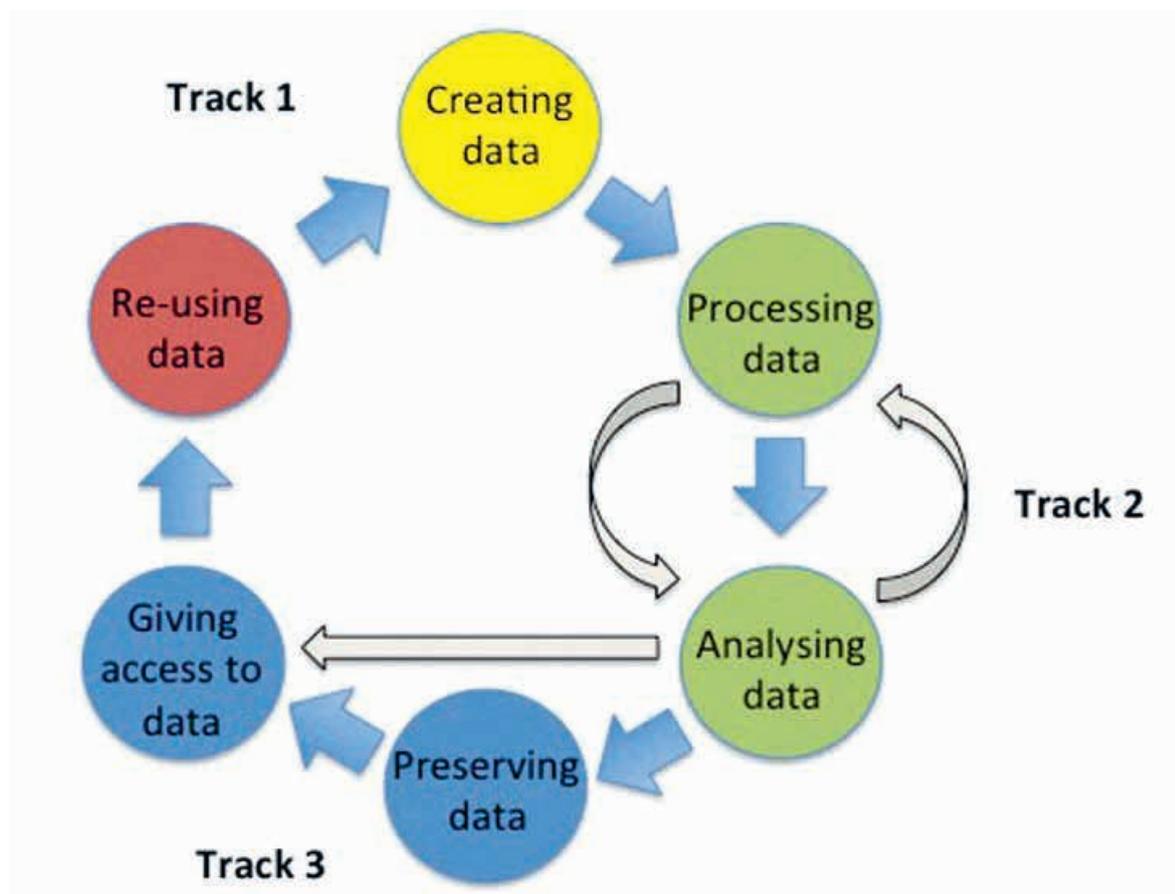
Le «track 2» porte sur le traitement «actif» des données; Le «track 3» considère la publication et conservation (long terme) des données;

Le «track 4» concerne la formation des utilisateurs sur l'usage des outils du DLCCM et l'élaboration de DMP;

Le «track 5» traite de la dissémination au niveau national des services nouvellement établis.

6 Laboratory Information Management System (LIMS).

7 Electronic Laboratory Notebook (ELN).



50

Fig. 2 Les éléments clés du cycle de vie des données du projet DLCM

### Les axes forts du projet

Un point vital du programme CUS P-2 est qu'au terme des différents projets de nouveaux services nationaux puissent être proposés à la communauté universitaire. Le projet DLCM, défini sur 3 ans, a pour ambition de traiter chaque étape du DLCM, avec à la clé la mise en place d'une pluralité de services construits sur la base d'une collaboration entre des informaticiens, des bibliothécaires, et des services à la recherche. Pour éviter cependant de trop se disperser dans les objectifs, deux disciplines «pilotes» sont considérées: les sciences de la vie et les humanités numériques. Ces disciplines, de prime abord antinomiques, vont au contraire forcer une conceptualisation de plus haut niveau de l'information, pour conduire à des solutions plus génériques. Par exemple, des réseaux sémantiques basés sur le modèle RDF sont agnostiques au domaine considéré et s'appliquent aussi bien en génomique qu'en sciences humaines. Reste que pour développer des services nationaux «génériques» couvrant le domaine du DLCM à large échelle, des modèles économiques robustes sont encore à concevoir, ce représente un défi supplémentaire.

### L'auteur

Pierre-Yves Burgi



Dr Pierre-Yves Burgi est directeur SI adjoint, responsable depuis juin 2003 du service des Nouvelles Technologies de l'Information, Communication, et Enseignement (NTICE) dans la division du Système de l'Information de l'Université de Genève. Il a reçu son diplôme d'ingénieur en informatique de l'École Polytechnique Fédérale de Lausanne en 1986, et le titre de docteur ès sciences de l'Université de Genève en 1992. Ses études doctorales ont été suivies par un post-doctorat d'une durée de 5 ans en Neurosciences de la vision, d'une part au Smith-Kettlewell Eye Research Institute, San Francisco, CA, et d'autre part au CNRS à l'Université Paul Sabatier, Toulouse, France.

# Standards bei digitalen Editionen

Marianne Hundt, Universität Zürich

*Digitale Editionen sind weit mehr als nur elektronische Ausgaben von Bucheditionen im herkömmlichen Sinne. Sie unterscheiden sich gleichzeitig von digitalen Archiven, die lediglich Originalmanuskripte als Bilddateien zur Verfügung stellen.*

Eine digitale Edition verbindet digitale Faksimiles von Originalmanuskripten bzw. verschiedene Versionen ein und desselben Textes aus Autorenhand mit einer diplomatischen Transliteration des handschriftlichen Textes bzw. der Textversionen. Informationen, die in einer analogen Printedition verloren gehen, wie z.B. Veränderungen im Schriftbild, lassen sich in einer digitalen Edition somit problemlos integrieren. Gleichzeitig ermöglichen digitale Editionen eine systematische Suche nach Textstellen mithilfe elektronischer Suchmaschinen. Anstelle des bei analogen Editionen üblichen kritischen Apparates können digitale Editionen den edierten Text mit Datenbanken verknüpfen, die umfangreiches zusätzliches Material (z.B. biographischen oder historischen Inhalts) zur Verfügung stellen.

## Internationale Standards

Um eine möglichst langfristige und nachhaltige Nutzung digitaler Editionen von Manuskripten zu gewährleisten, ist es unerlässlich, dass die diplomatischen Transliterationen allgemein anerkannten, internationalen Standards genügen. Diese Standards werden seit den 1990er-Jahren von einem internationalen Konsortium, der Text Encoding Initiative (TEI)<sup>1</sup>, entwickelt und in Form von Richtlinien und frei zugänglichen Tools der Akademischen Community zur

Verfügung gestellt. Die TEI verbindet zudem die Entwicklung und Anpassung geeigneter Standards und Tools mit der Diskussion inhaltlicher und auch theoretischer Aspekte (wie zum Beispiel dem Verhältnis von annotiertem Text und dessen «Textualität» [s. Buzzetti & McGann]).

## Umfangreichere Nutzungsmöglichkeiten

Analoge Printeditionen sind aufgrund ihrer Bindung an das Buchformat relativ starre, lineare Texte, die auch mit einem umfangreichen kritischen Apparat immer eine Interpretation eines Originaltextes darstellen und es dem Nutzer nicht unbedingt ermöglichen, den Interpretationsprozess in jedem einzelnen Fall nachzuvollziehen. Im Gegensatz dazu zeichnen sich digitale Editionen dadurch aus, dass sie vielfältige Annotations- und Verknüpfungsmöglichkeiten bereitstellen. Diese erlauben es beispielsweise, Bilddateien von Originalmanuskripten mit deren Transliterationen und normalisierten Textversionen zu verbinden, die es dem Benutzer ermöglichen, den edierten Text zu durchsuchen und bei Bedarf die edierte Fassung mit dem Originalmanuskript abzugleichen. Mit anderen Worten: Die Nutzer sind nicht auf die Interpretation des Textes durch die Herausgeber angewiesen. Da digitale Editionen zudem mehrere Versionen eines Textes (z.B. eine dem Original sehr nah verwandte und eine normalisierte Version, die orthographische Variation ausblendet, oder auch mehrere Fassungen eines Textes aus Autorenhand) miteinander verknüpfen, stellen sie eine gegenüber analogen Texteditionen deutlich verbesserte Zugriffsmöglichkeit auf ursprünglich handschriftliche Texte dar.

---

<sup>1</sup> [www.tei-c.org](http://www.tei-c.org)

### Arbeit mit Metadaten

Auch die Verknüpfung mit Datenbanken, die Metadaten enthalten, macht digitale Editionen anders und vielfältiger nutzbar als herkömmliche Printeditionen. Als Beispiel sei ein Briefwechsel zwischen einer historischen Person A und ihrem sozialen Netzwerk genannt. In einer digitalen Edition lassen sich biographische Daten über jede einzelne Person in einer Datenbank speichern und in die Suche integrieren, so dass eine Suche nach Sprachgebrauch in Abhängigkeit des Adressaten (hinsichtlich von Merkmalen wie Geschlecht, Herkunft etc.) möglich wird. Da im Transliterations- und Annotationsprozess aber beispielsweise auch Durchstreichungen, Textkorrekturen etc. kodiert werden, wird es möglich, auch den Textentstehungsprozess systematischer digital zu untersuchen, indem Varianten ein und desselben Wortes jeweils in ihrem Kontext suchbar gemacht werden.

### Herausforderung für die Konzeption

Die gegenüber einer linearen Printversion deutlich umfangreicheren Nutzungsmöglichkeiten stellen gleichzeitig eine Herausforderung für die Konzeption von digitalen Editionen dar. Während konventionelle Editionen in der Regel für ein recht klar definiertes Zielpublikum erstellt werden, sind digitale Texteditionen potenziell für ein breites Nutzerpublikum interessant. Ein Beispiel wären die unveröffentlichten Manuskripte von Jane Austen<sup>2</sup>, die den Ansprüchen an digitale Editionen hinsichtlich Enkodierung und Verknüpfung von Manuskript und Transliteration

vollumfänglich entsprechen. Allerdings wurden die Texte offensichtlich in erster Linie für literaturwissenschaftliche Nutzung editiert, so dass die Suchmaske für benachbarte Disziplinen in ihren Optionen einige Wünsche offenlässt: Aus sprachwissenschaftlicher Perspektive wären Suchen über variante Wortformen, Wortarten, komplexere Konstruktionen und die Möglichkeit zur nachträglichen Annotation von Konkordanzen beispielsweise wünschenswert (auf der Basis des vorhandenen Materials aber sicher auch implementierbar).

---

### Zur Autorin

#### Marianne Hundt



Prof. Dr. Marianne Hundt leitet das Englische Seminar der UZH und das Zürcher Kompetenzzentrum für Linguistik. Ihre Forschungsschwerpunkte liegen im Bereich der korpusbasierten Beschreibung von diachroner und synchroner Variation der englischen Sprache und ihrer regionalen Ausprägungen. Sie hat eine Reihe elektronischer Korpora des Gegenwartsenglischen und der englischen Sprache der späten Neuzeit kompiliert und zu Fragen der Annotation (besonders PoS-Annotation) gearbeitet. Gemeinsam mit Daniel Schreier gibt sie seit 2013 die Zeitschrift «English World-Wide» heraus.

---

<sup>2</sup> [www.janeausten.ac.uk](http://www.janeausten.ac.uk)

## Digitale Sammlungen in Bibliotheken

Eva Martina Hanke, Zentralbibliothek Zürich

*Digitale Inhalte sind neben den physisch vorhandenen ein fester Bestandteil der Bibliotheksangebote geworden. Ein möglichst umfangreiches Spektrum bereitzustellen und dieses kontinuierlich zu erweitern, ist in der Strategie der Zentralbibliothek (ZB) Zürich verankert. Den grössten Teil machen E-Zeitschriften, E-Books, elektronische Nachschlagewerke und Datenbanken aus, welche die Bibliothek lizenziert. Hinzu kommen elektronische Hochschulschriften, neuerdings Forschungsdaten, Webseiten aus dem Kanton oder auch digitale Dokumente aus Nachlässen.*

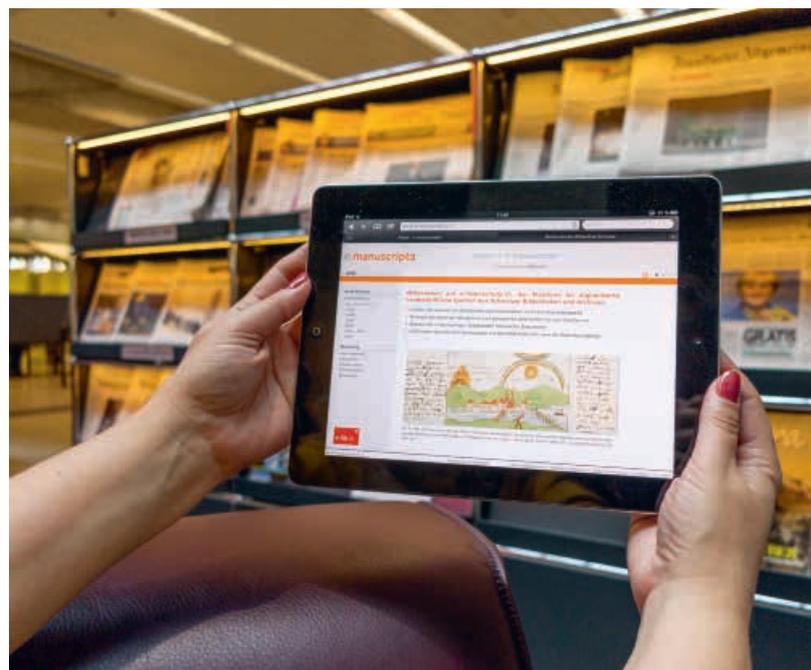
Neben diesen «born digital»-Inhalten, die zum Sammelauftrag gehören, gewinnt die sog. Retrodigitalisierung an Bedeutung. Mittlerweile sind zu den digitalen Reproduktionen aus Benutzeraufträgen in der ZB Digitalisierungsprojekte hinzugekommen, in denen systematisch urheberrechtsfreie Drucke, Handschriften, Musikalien, Bilder, Karten, Zeitungen oder Zeitschriften eingescannt, mit Metadaten erschlossen und online gestellt werden. Diese digitalen Sammlungen dienen dem Schutz der historischen Originale und erleichtern zugleich durch die zeitlich, örtlich und nicht auf jeweils eine Person beschränkte Benutzung den Zugang. Da die arbeits- und kostenintensive Dienstleistung der Retrodigitalisierung in der ZB noch nicht in Leistungsauftrag und Budget vorgesehen ist, ist sie wie andere Bibliotheken auf Kooperationen und Drittmittel angewiesen.

### Vom Original zum Digitalisat

«Highlights», vom Zerfall bedrohte Dokumente oder eine systematische Digitalisierung von grösseren Sammlungen nach Provenienz, Thema, Bedeutung für die Bibliothek oder aufgrund besonderer Nachfrage aus Forschung oder Öffentlichkeit? Inhaltlich lassen sich beim Aufbau digitaler Sammlungen verschiedene Strategien verfolgen. Wichtig ist jedoch, dass die Bibliothek auch bei der Retrodigitalisierung an ihren Standards von qualitativ hochwertigen und nachhaltigen Dienstleistungen festhält.

So läuft die Digitalisierung als Prozess ab, der von der Auswahl der Originale und deren Erschliessung mit konsistenten Metadaten über Massnahmen der Bestandserhaltung bis hin zur Datenlogistik und -ablage weit mehr als das eigentliche Einscannen umfasst. Dieses erfolgt in grossen Bibliotheken wie der ZB im eigenen Digitalisierungszentrum mit einheitlichen Vorgaben (TIFF-Dateien mit mindestens 300 dpi und Farbkeil) und unter maximaler Schonung der Originale. Um Qualität und Nachhaltigkeit sicherzustellen, sind neben dem Aufbau von entsprechenden Abläufen und Kompetenzen bei den Mitarbeitenden grössere Investitionen in die Infrastruktur wie leistungsfähige Scanner oder ausreichende Netzwerk-, Server- und Speicherkapazitäten nötig. Dies gewährleistet nicht nur die Produktion hochwertiger Digitalisate, sondern auch einen schnellen und möglichst langfristigen Zugriff auf die digitalen Sammlungen.

53



### Digitale Sammlungsbestände für Forschung und breitere Öffentlichkeit

Recherchierbar sind die retrodigitalisierten Bestände über den Bibliothekskatalog, über den auch die Digitalisate zugänglich sind. Schweizer Bibliotheken und Archive nutzen für deren Präsentation Plattformen wie e-rara.ch, e-manuscripta.ch, e-codices, retroseals.ch oder Schweizer Presse Online, die zwar jeweils nur eine Materialart, jedoch Inhalte aus mehreren Institutionen enthalten. Die Dokumente sind kostenlos zu nutzen und durch permanente Links dauerhaft zitierbar. Bei gedruckten Texten ist eine Volltexterkennung mit OCR, für e-manuscripta.ch und e-codices die Einbindung von Transkriptionen bzw. Annotationen im Einsatz bzw. in Vorbereitung. Auf diesen Wegen zum Volltext werden die Bibliotheken eng mit der Forschung zusammenarbeiten.

Zwar bleibt am Ende jedes noch so hochwertige Digitalisat nur die Kopie eines Originals, dessen Konsultation bei gewissen Forschungsfragen unerlässlich bleibt. Dennoch profitieren vom Auf- und Ausbau solcher digitaler Sammlungen nicht nur Forschung und Öffentlichkeit, sondern auch die vielfältigen historischen Bibliotheksbestände, die durch ihre digitalen Stellvertreter entgegen dem Trend «was nicht online verfügbar ist, existiert nicht» sichtbar bleiben und zur Nutzung einladen.

---

### Zur Autorin

#### Eva Martina Hanke



Dr. Eva Martina Hanke promovierte nach dem Studium der Musikwissenschaft, Anglistik und Rechtswissenschaften in Marburg/Lahn (D) und York (GB) an der Universität Zürich im SNF-Projekt «Musik in Zürich». Seit 2005 arbeitet sie in der Musikabteilung der ZB Zürich. Nach dem MAS in Bibliotheks- und Informationswissenschaften an der Universität Zürich übernahm sie zudem 2012 die Projektleitung/Geschäftsstelle von e-manuscripta.ch sowie 2013 die Projektleitung des Digitalisierungsprojekts DigiTUR, in dem die ZB Zürich mit Unterstützung des Lotteriefonds des Kantons Zürich systematisch historische Zürcher Bestände digitalisiert und online präsentiert.

---

### Weitere Informationen

Digitale Sammlungen und Digitalisierungsprojekte der Zentralbibliothek Zürich:

E-Ressourcen: <http://www.zb.uzh.ch/recherche/e-ressourcen>

Projekt DigiTUR: <http://www.zb.uzh.ch/spezialsammlungen/digitur>

## Bilddatenbanken: Geisteswissenschaften im Medienwandel

Tristan Weddigen, Kunsthistorisches Institut der Universität Zürich

*Die digitale Revolution hat unsere Welt tiefgreifend verändert. Damit die Geisteswissenschaften an dieser Realität Anteil haben und sie mitgestalten, müssen sie sich in einem Feld positionieren, das vor allem von einer globalisierten Marktwirtschaft definiert wird. Das ist kein leichtes Unterfangen, aber aufgeben ist erst recht keine Option.*

Weil die Geisteswissenschaften sich daran beteiligen, Kultur zu bewahren, zu vermitteln und zu erneuern, ist die Frage des Medienwandels von allgemeinem Interesse. In Europa, wo Kultur, Bildung, Wissenschaft demokratisch und staatlich gestützt werden, sind zahlreiche, allerdings nur wenig koordinierte Initiativen entstanden. Der Beitrag der SAGW zur Diskussion und Koordination der digitalen Geisteswissenschaften in der Schweiz ist daher sehr zu begrüßen. Allerdings stehen die Geisteswissenschaften angesichts des technischen Fortschritts und der zunehmenden Komplexität der Aufgaben immer noch, ja immer wieder am Anfang. Von der Illusion einmaliger, zentralisierter und universeller digitaler Lösungen muss man sich verabschieden. Die Digitalisierung auch des kulturellen und geisteswissenschaftlichen Bereichs ist ein Prozess ständiger Weiterentwicklung.

### Aufbau und Vernetzung von Bilddatenbanken

Die geisteswissenschaftlichen Disziplinen sind teilweise durch ihre Gegenstände und Medien definiert. Dies gilt insbesondere für die Kunstgeschichte, deren Beschäftigung mit visuellen Artefakten und deren Reproduktionen seit jeher zu ihren Grundlagen gehört und deren Methoden sich mit dem digitalen Medienwandel verändert haben. Daher stellen der Aufbau und die Vernetzung von Bilddatenbanken für die Kunstgeschichte, aber auch für andere bildinteressierte Geisteswissenschaften ein eigenes Aufgaben- und Problemfeld dar. Die Forschung macht Dinge sichtbar, sieht aber selbst auch nur das, was aufgrund anderer Faktoren sichtbar wird.

### Herausforderung Digitalisierung

Die digitale Revolution hat den Horizont der Geisteswissenschaften, auch der Kunstgeschichte, global erweitert. Allerdings stehen ihnen noch nicht die entsprechenden wissenschaftlichen Ressourcen zur Verfügung. Nur ein schwindender Bruchteil der für eine globale Kunstgeschichte relevanten Quellenbestände liegt heute in digitaler Form vor. Allein wirtschaftlich hoch entwickelte Länder verfügen über flächendeckende fotografische Dokumentationen ihres künstlerischen Erbes, die zudem in der Regel nur in analoger Form in Fototheken und Denkmalpflegeämtern vorliegen. Nur wenige Museen haben damit begonnen, ihre Bestände systematisch der Weltöffentlichkeit auf dem Netz frei zur Verfügung zu stellen. Die Digitalisierung künstlerischen Erbes, die heute zunehmend als Dienstleistung für die Öffentlichkeit erwartet wird,

55



Der Medienwandel in der Kunstgeschichte ist nie abgeschlossen: Alte Medien werden zum Kulturgut und Studienobjekt.

stellt für alle betroffenen Institutionen eine grosse technische, finanzielle und auch wissenschaftliche Herausforderung dar, für die es in den seltensten Fällen nationale oder gar internationale Lösungsansätze gibt. Das trifft gerade auch für die föderale und mehrsprachige Schweiz zu. Allerdings bietet der sehr gute Wissenschaftsstandort die Chance, Modelllösungen mit internationalem Potenzial zu entwickeln.

#### **Verlässlichkeit von Metadaten verbessern**

Geisteswissenschaftliche Digitalisierungsprojekte dienen der Sicherung, Vermittlung und Erneuerung kulturellen Wissens. Sie gehören daher wissenschaftlich gesehen zur Grundlagenforschung, ja zu ihrem existenziellen Rahmenwerk. Forschungsgelder müssten daher kontinuierlich und koordiniert der infrastrukturellen Weiterentwicklung zufließen. Andernfalls werden sich die Geisteswissenschaften im Mainstream globaler Konzerne treiben lassen, die uns vorspiegeln, Wissen entstehe gratis und von selbst. Bereits jetzt zeigt sich, dass die wissenschaftliche Qualität in der Verwendung von Bildmaterial in der kunsthistorischen Forschung wegen der grossen Suchmaschinen eher abgenommen hat. Wenn in Forschung und Lehre digitale Bilder die Gegenstände und Suchmaschinen die Sammlungen zunehmend ersetzen, sind ein kritisches Instrumentar und eine erneuerte wissenschaftliche Bildkompetenz nötig. Es ist deswegen wichtig, dass die Kunstgeschichte die Verlässlichkeit von Metadaten, die ein Objekt und seine Reproduktionen beschreiben, aufgrund international gültiger Normierung verbessert. Dann ist die dezentrale Vernetzung der zahlreichen existierenden Bilddatenbanken von Hochschulen, Museen und Archiven technisch und sprachübergreifend herzustellen und der Zugang zu ihnen zu erleichtern. Gerade darin kann die mehrsprachige, an Kultur- und Forschungsinstitutionen reiche Schweiz Modelllösungen entwickeln und ein grosses kulturelles und wissenschaftliches Potenzial ausschöpfen.

#### **Nationales Kompetenzzentrum für digitale Bildwissenschaft**

Für die Koordination der dezentralen Ressourcen wäre ein nationales Kompetenzzentrum für digitale Bildwissenschaft einzurichten. Ein Netzwerk der Bilddatenbanken würde einerseits neue Wissensquellen zugänglich machen und andererseits ihre Verlässlichkeit und Nachhaltigkeit sichern. Sowohl lokal wie auch global ausgerichtete Forschung würde dadurch neue Impulse erhalten. Langfristig müsste ein solches Netzwerk zu einer frei zugänglichen, globalen Bilddatenbank ausgebaut werden. Diese würde nicht nur die eurozentrischen ästhetischen Kanones aufbrechen, sondern vor allem auch jene grundlegenden Mittel zur Dokumentation und Erforschung kulturellen Erbes weltweit zur Verfügung stellen, welche bisher das Privileg weniger Industrienationen sind.

Ein solches Unterfangen ist so dringend wie herausfordernd. An wissenschaftlichen, technischen, sprachlichen, rechtlichen, finanziellen, interdisziplinären Schwierigkeiten mangelt es nicht. Eines der grössten Risiken ist jedoch, es nicht zu versuchen und untätig zu bleiben. Mit digitalen Diensten an der Gemeinschaft sind keine akademischen Lorbeeren zu pflücken, und darauf auszuruhen, dazu wird es nie kommen – doch ein Anfang muss gemacht werden.

---

#### **Zum Autor**

##### **Tristan Weddigen**



Prof. Dr. Tristan Weddigen hat den Lehrstuhl für Kunstgeschichte der Neuzeit an der Universität Zürich inne. Zu seinen Forschungsfeldern gehören die frühneuzeitliche Kunst und Kunsttheorie, das textile Medium, die globale Kunstgeschichte und die Wissenschaftsgeschichte. Zuletzt hat er Heinrich Wölfflins «Kunstgeschichtliche Grundbegriffe» neu und kritisch auf Englisch herausgegeben.

# ortsnamen.ch – Portal der schweizerischen Ortsnamenforschung

Hans Bickel, Martin Graf, Schweizerisches Idiotikon

*ortsnamen.ch ist das Portal der wissenschaftlichen Ortsnamenforschung in der Schweiz. Kernstück ist die georeferenzierende Visualisierung der Forschungsdaten über verschiedene aktuelle und historische Online-Kartensysteme.*

Das Portal stellt die Forschungsarbeit der unabhängigen Namenforschungsprojekte über eine Internet-Plattform der Öffentlichkeit kostenlos zur Verfügung, stellt die technische Aktualisierung und Speicherung der digitalen Daten aus diesen Projekten sicher, sammelt und publiziert alle erhältlichen Informationen zur Schweizer Ortsnamenforschung und hilft, die unterschiedlichen Projekte zu vernetzen.

## Die Akteure

ortsnamen.ch wurde 1997 von Eugen Nyffenegger, Kreuzlingen, unter dem Namen Datenbank der Schweizer Namenbücher gestartet. Finanziert wurde die zunächst noch als lediglich offline funktionierende Datenbank vom Schweizerischen Nationalfonds. Seit 2006 liegt die Projektleitung bei Hans Bickel. Als die aufwändigsten Erfassungen von nicht digitalisierten Daten aus Namenbüchern abgeschlossen werden konnten und die Datenbank in eine Internet-Plattform überführt wurde, wurde das Projekt 2010 dem Schweizerischen Idiotikon angegliedert und wird seither von der SAGW als geisteswissenschaftliche Infrastruktur finanziert. Gleichzeitig wurde die Projektleitung um Martin Graf erweitert.

## Der Zweck

Ausgangspunkt des Projektes war die Erkenntnis, dass in der Schweiz die Ortsnamenforschung weitgehend kantonal organisiert ist und dass die elektronischen Daten, die für die Publikation eines Namenbuchs erarbeitet werden, nach erfolgter Publikation häufig nicht mehr zur Verfügung standen oder gar gelöscht wurden. Darum sollte eine Datenbank, die die Daten aus abgeschlossenen, archivierten oder ab-

gebrochenen Projekten aufarbeitet und erschliesst, eine solide EDV-Grundlage bieten und es den noch laufenden Namenforschungsprojekten ermöglichen, die Daten der abgeschlossenen Projekte einzusehen und zu nutzen. Zudem sollten die erschlossenen Daten auch für die Langzeiterhaltung archiviert und technisch auf dem neuesten Stand gehalten werden. Eine Aufgabe der Ortsnamenforschung ist es, die erhobenen Ortsnamen nicht nur sprachwissenschaftlich zu analysieren und historisch einzubetten, sondern sie auch zu lokalisieren und ihre Lage und geographische Beschaffenheit zu beschreiben. Nach Aufkommen der im Internet verfügbaren Karten, zuerst GoogleMaps, später swisstopo und OpenStreetMap, konnten unsere Informatiker Felix Nyffenegger und Tobias Roth ein System entwickeln, das die Visualisierung der Forschungsdaten auf dynamischen Karten erlaubt. Im Sinn des Open Access konnten schliesslich die Forschungsdaten auch für die breite Öffentlichkeit zugänglich gemacht werden.

## Die Erfahrungen

ortsnamen.ch geniesst in der Community einen ausgezeichneten Ruf. Nicht zuletzt dank zusätzlicher Features wie einer stets aktuell gehaltenen, gut strukturierten und umfassenden Bibliographie zur Schweizer Ortsnamenforschung oder auch einer Rubrik zu aktuellen Veranstaltungen und Publikationen ist ortsnamen.ch zur wichtigsten Anlaufstelle für onomastische Fragen in der Schweiz geworden. Mit mehr als 200 Besuchern täglich wird die Seite auch gut genutzt. Der Aufwand für die Aufrechterhaltung eines derartigen Online-Angebots, das auch externe Daten miteinbezieht, darf allerdings nicht unterschätzt werden.

## Die Erwartungen

Von Nutzerseite werden in erster Linie eine intuitive Bedienung des Internetangebots sowie möglichst

58

vollständige Informationen zu den Ortsnamen erwartet. Von wissenschaftlicher Seite wird zudem gewünscht, dass Vernetzungsmöglichkeiten angeboten werden.

### Die Herausforderungen

Für die Betreiber stehen grosse Herausforderungen an. Zum einen müssen sehr heterogene Daten, die in verschiedenen unabhängigen Projekten mit meist einer proprietären Datenbanklösung erfasst wurden, zu einem möglichst einheitlichen, konsistenten System vereinigt werden. Mit dem zunehmenden Bedürfnis nach kontrollierten Normdaten steigt die Erwartung, dass die Namensdaten nach einheitlichen, vorgegebenen Kriterien kategorisiert und Namendoubletten (z.B. kantonsübergreifende Gewässer- oder Hügelnamen) eliminiert werden. Zudem sind Kooperationen oder zumindest Abstimmungen mit anderen Projekten nötig, die ebenfalls georeferenzierte Namensdaten anbieten, insbesondere mit der Landestopografie (swisstopo), mit einem Projekt der Staatsarchive der Schweiz (topoterm) sowie der internationalen Namensdatenbank Geonames. Unter dem Namen Histhub.net ist zudem ein Projekt in Planung, das nach Möglichkeit sogenannte Linked Open Data zu historischen und geographischen Entitäten aller Epochen und Regionen der Schweiz bereitstellen soll. ortsnamen.ch wird darin verantwortlich für georeferenzierte Daten sein.

---

### Zu den Autoren

#### Martin Graf



Dr. Martin Graf studierte in Zürich Germanistik, allgemeine Geschichte und Mittel-lateinische Philologie. Er ist Redaktor am Schweizerdeutschen Wörterbuch (Idiotikon) und Co-Projektleiter des Infrastrukturprojekts ortsnamen.ch.

#### Hans Bickel



Prof. Dr. Hans Bickel studierte in Basel Germanistik, Ethnologie und Volkskunde. Er ist Redaktor am Schweizerdeutschen Wörterbuch (Idiotikon), Titularprofessor für Deutsche Philologie an der Universität Basel und Co-Projektleiter des Infrastrukturprojekts ortsnamen.ch.

---

### Weitere Informationen

[www.ortsnamen.ch](http://www.ortsnamen.ch)

# Das Digitale Wörterbuch der deutschen Sprache

Wolfgang Klein, Max-Planck-Institut für Psycholinguistik

*Den Wortschatz einer Sprache in seinen unterschiedlichen Eigenschaften zu erforschen und die dabei gewonnenen Einsichten der Öffentlichkeit in umfassender, klarer und verständlicher Form zu vermitteln, ist eine Aufgabe von nicht nur wissenschaftlichem, sondern auch von beträchtlichem gesellschaftlichem Interesse.*

Verlässliche Informationen über Schreibweise, Aussprache, Bedeutungen, grammatische Eigenschaften, regionale wie stilistische Variation und Herkunft der Wörter sind für Wissenschaftler vieler Fächer, aber auch für Journalisten, Übersetzer, Lehrer und nicht zuletzt für all jene, die die Sprache lernen wollen, wichtig. Der traditionelle Weg, diese Aufgabe zu lösen, sind gedruckte Wörterbücher. So bewährt dieser Weg ist, er hat eine Reihe von Nachteilen, was die Abdeckung des Wortschatzes, die Erweiterbarkeit und Korrigierbarkeit der Informationen, aber auch die Darstellungsweise der verschiedenen Worteigenschaften angeht.

## Neue digitale Methoden

Das Aufkommen digitaler Methoden hat es möglich gemacht, gedruckte Wörterbücher durch umfassende digitale lexikalische Systeme (DLS) abzulösen, die

- auf Computer stehen und für jedermann über das Internet zugänglich sind,
- lexikalische Analyse der Worteigenschaften und Quellen in Form grosser, gut annotierter Korpora miteinander verbinden, so dass jederzeit von dem einen ins andere gewechselt werden kann,
- aus einzelnen, unabhängig zu bearbeitenden und auch finanzierbaren Modulen bestehen,
- sich jederzeit leicht erweitern und korrigieren lassen und
- es erlauben, das bisherige lexikalische Wissen, wie es sich in den bisherigen guten Wörterbüchern niederschlagen hat, flexibel zu integrieren.

## Aufbau des DWDS

Das Digitale Wörterbuch der deutschen Sprache (DWDS) ist ein solches digitales lexikalisches System, das seit etwa fünfzehn Jahren an der Berlin-Brandenburgischen Akademie der Wissenschaften schrittweise entwickelt wird. Es besteht im Wesentlichen aus drei Komponenten: Korpora, Wörterbücher und verschiedene Werkzeuge. Die wichtigsten Teilkorpora sind

- (a) das «Kernkorpus», das sorgfältig ausgewählte Texte von 1900 bis in die Gegenwart im Umfang von etwa 100 Millionen Wörtern umfasst,
- (b) das Deutsche Textarchiv ([www.deutsches-textarchiv.de](http://www.deutsches-textarchiv.de)) mit etwa 1500 wichtigen Werken, zumeist aufgrund der Erstausgaben, von 1600–1900, und
- (c) das Ergänzungskorpus mit zumeist neueren Zeitungstexten (Umfang etwa 1,8 Milliarden).

## Integrierte Wörterbücher

Es wurden bislang drei bedeutende ältere Wörterbücher integriert; das von den Brüdern Grimm begonnene «Deutsche Wörterbuch» in der Erstbearbeitung, das «Etymologische Wörterbuch» von Wolfgang Pfeifer (das von diesem stetig korrigiert und erweitert wird) sowie das eigentliche DWDS-Wörterbuch. Letzteres beruht auf dem von Wolfgang Steinitz initiierten «Wörterbuch der deutschen Gegenwartssprache» (1964–1977), das jedoch inzwischen erheblich korrigiert und erweitert wurde; auch wurde es durch eine Aussprachekomponente ergänzt.

## Werkzeuge des DWDS

Zu den Werkzeugen zählen nicht nur die üblichen Suchmöglichkeiten, sondern auch statistische Verfahren, die beispielsweise das Vorkommen eines Wortes über die Zeit oder in verschiedenen Textsorten zu vergleichen erlauben. Dem Nutzer stehen verschiedene

«Ansichten» zu Verfügung, die er weitgehend selbst zusammenstellen kann. Hier ist ein Beispiel für das Wort Sex mit sechs verschiedenen «Fenstern».

### Zugang

Das DWDS ist über das Internet ([www.dwds.de](http://www.dwds.de)) frei zugänglich; für einige wenige neuere Texte ist aus rechtlichen Gründen eine – gleichfalls kostenfreie – Registrierung erforderlich. Derzeit gibt es etwa 35 000 registrierte Benutzer aus 140 Ländern mit täglich etwa 150 000 Aufrufen.

### Weiterentwicklung

Das DWDS wird nach Massgabe der Geldmittel in allen drei Komponenten systematisch ausgebaut. Wir versuchen zunehmend, auch andere Forschungsstätten zur Mitarbeit zu gewinnen, für die sich der modulare Aufbau des Unternehmens in besonderem Mass eignet.

### Zum Autor

#### Wolfgang Klein



Prof. Dr. Wolfgang Klein (1946) hat Germanistik, Romanistik und Philosophie studiert. Er wurde 1970 in Saarbrücken promoviert und hat sich 1972 in Heidelberg habilitiert. Von 1972 bis 1976 war er in Heidelberg, danach bis 1980 in Frankfurt/M. Professor für Germanistik. Seither ist er Direktor am Max-Planck-Institut für Psycholinguistik. Er interessiert sich für alles Mögliche.

## LOD entzaubert: Was semantische Technologien wirklich können

Matej Ďurčo, Austrian Centre for Digital Humanities

*Obwohl die berühmte LOD-Cloud<sup>1</sup> unaufhaltsam wächst und die semantischen Technologien auch schon im kommerziellen Bereich<sup>2</sup> längst angekommen sind, scheint in der (geisteswissenschaftlichen) Forschung der Durchbruch noch auszustehen. Aber lohnt es sich überhaupt, sich mit diesem Hype zu beschäftigen? Und ist es noch ein Hype?*

### Linked Open Data (LOD) löst alle unsere Probleme

Wer das Semantic Web und Linked Open Data (LOD) als die (Er)lösung herbeisehnt, wird enttäuscht sein. LOD ist zwar ein genauso einfacher wie genialer Ansatz, aber nüchtern betrachtet ist das zugrunde liegende RDF<sup>3</sup> ein Datenaustauschformat und ein radikal einfaches obendrein. RDF – Resource Description Framework – dient der Beschreibung von Ressourcen mit Aussagen bestehend aus genau 3 Worten, dem «Triple»: Subjekt Prädikat Objekt. Eine Ressource kann ein digitales Dokument sein, aber auch eine Stadt, oder eine Körperzelle. Es kann eine eindeutige Entität sein, aber auch eine Klasse von Entitäten. Das wohl entscheidende Merkmal der semantischen Technologien ist das globale Referenzierungssystem. Dies ist

auch die erste Regel des LOD-Paradigmas: «Use URIs as names for things»<sup>4</sup>, bzw. noch schärfer formuliert: «If it doesn't use the universal URI set of symbols, we don't call it Semantic Web.»

### Was ist dann eigentlich das Beschreibungsvokabular?

Obwohl man also im Prinzip absolute Freiheit bzgl. der «Modellierung» und Strukturierung der eigenen Daten hat, ist es ratsam sich an existierenden Vokabularen zu orientieren. Einige Verzeichnisse helfen dabei, eine Übersicht in dem Dschungel zu gewinnen oder zu bewahren, wie beispielsweise das Portal LOV – «Linked Open Vocabularies»<sup>5</sup>, das 520 solcher Vokabularen auflistet, zumindest vier seien hervorgehoben: «Simple Knowledge Organisation Systems», oder SKOS<sup>6</sup>, für Taxonomien und Thesauri; «Friend of a Friend», FOAF<sup>7</sup>, für Beschreibungen und Beziehungen von Personen; «dcterms»<sup>8</sup> als der gemeinsame Nenner vieler Metadaten-Ansätze; und «schema.org»<sup>9</sup> als eine gemeinsame Initiative von kommerziellen Big Playern, die das Ziel verfolgen durch eine einheitliche semantische Auszeichnung bessere Strukturierung und dadurch Auffindbarkeit von Web-Inhalten zu gewährleisten.

### Womit verlinken?

Auch bei den existierenden Datensätzen, den Kandidaten fürs Verlinken, herrscht eine unüberschaubare Vielfalt. Die LOD-Cloud repräsentiert zwar nur einen Ausschnitt aus dem riesigen Fundus, aber immerhin bietet sie eine anschauliche erste Landkarte des «Web of Data». Die Hauptstadt ist von Anbeginn dbpedia<sup>10</sup>, die (kontinuierlich) in RDF übersetzte Version von Wikipedia. Inzwischen gibt es einige weitere riesige semantische Ressourcen, kompiliert und konsolidiert aus verschiedenen Quellen, wie z.B. YAGO mit 120 Mio. «Fakten» über mehr als 10 Mio. Entitäten. Eine besonders aktive Community im Rahmen der Geisteswissenschaften ist die Linguistik mit der Website «Linguistic Linked Open Data» (LLOD)<sup>11</sup>, die mit eigener LLOD-Cloud und eigenem LingHub<sup>12</sup> eine feiner aufgelöste Sicht auf die Daten in dieser Domäne bietet.

Besonders relevant fürs Verlinken sind die grossen Referenzressourcen. Zunehmend werden die grossen traditionellen Klassifikationsschemata wie die «Gemeinsame Normdatei» der Deutschen Nationalbibliothek, die «Library of Congress Subject Headings» (LCSH), die «Getty Thesauri» etc. als LOD bereitgestellt. Eine auch linguistisch sehr relevante Ressource ist «EuroVoc»<sup>13</sup>, der multilinguale Thesaurus der EU.

1 <http://lod-cloud.net/>

2 <http://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-things-not.html>

3 <http://www.w3.org/RDF/>

4 <http://www.w3.org/DesignIssues/LinkedData.html>

5 <http://lov.okfn.org/dataset/lov/>

6 <http://www.w3.org/2009/08/skos-reference/skos.html>

7 <http://xmlns.com/foaf/spec/>

8 [dublincore.org/2012/06/14/dcterms](http://dublincore.org/2012/06/14/dcterms)

9 <http://schema.org/>

10 <http://wiki.dbpedia.org/>

11 <http://linguistic-lod.org/lod-cloud>

12 <http://linghub.lider-project.eu/>

13 <http://eurovoc.europa.eu/drupal/>

### Wozu ist es gut?

Das Leitthema der DHd<sup>14</sup>-2016-Tagung in Leipzig<sup>15</sup> bringt es auf den Punkt: «Modellierung – Vernetzung – Visualisierung: die Digital Humanities als fächerübergreifendes Forschungsparadigma»<sup>16</sup>. Interessanterweise trifft der Dreifachschwerpunkt genauso gut auf semantische Technologien wie auf DH zu.

Ausgehend von den Forschungsobjekten muss also zuerst die Entscheidung getroffen werden, wie sie digital erfasst, modelliert werden. Eine Reihe von Applikationen ermöglicht komfortable Eingabe von RDF-basierten Daten<sup>17,18,19,20,21</sup>. Zum Speichern und Abfragen der RDF-Daten werden sogenannte Triple-Stores eingesetzt, auch hier gibt es eine breite Auswahl an Lösungen<sup>22,23</sup>. Es ist aber auch möglich, die Daten in traditionellen Datenbank-Applikationen einzugeben und nur eine Export-Möglichkeit für RDF zu schaffen.

Der spannendste Teil der Arbeit ist wohl die Annotation, das Anreichern der eigentlichen Forschungsdaten mit Links zu den Referenzressourcen, am häufigsten wohl die Auflösung von Bezeichnungen in Texten zu Entitäten. Hierfür steht auch eine Reihe von (tlw. web-basierten) Diensten zur Verfügung: «Reconciliation Services»<sup>24,25,26,27</sup>. Obwohl diese Dienste mit riesigen Datensammlungen im Hintergrund und mit beachtlichem Durchsatz arbeiten, variiert die Trefferquote beträchtlich. Das automatische Verarbeiten sollte daher durch manuelle Nachkontrolle verifiziert werden.

Ein Aspekt, der oft unbeachtet bleibt, ist das «Reasoning», das logische Schliessen bzw. Inferieren neuer Fakten. Während es traditionell in der Domäne der wissensbasierten Systeme eine zentrale Rolle spielt, begnügt man sich beim Semantic Web oft mit reiner Beschreibung. Ein Grund dafür könnte sein, dass für diesen Zweck viel strikter definierte Ontologien mit einer Reihe logischer Aussagen über die modellierte Domäne erforderlich ist.

### Darstellung der Daten

Zu guter Letzt die Darstellung der aufbereiteten und vernetzten Daten. Eine versatile Möglichkeit, die RDF-Daten zu erkunden, bietet die Abfragesprache SPARQL.

Wesentlich leichter zu nutzen ist die semantische Suche, die durch die Kombination von Volltext-Suche und weiterführenden Links in den Resultaten die Daten als einen navigierbaren Graphen vor den Benutzern ausbreitet. Dazu kommen zunehmend integrierte Visualisierungskomponenten wie Kartensichten, Zeitleisten, dynamische interaktive Netzwerke.

All das gibt und gab es auch ohne LOD, aber die semantischen Technologien und insbesondere das LOD-Paradigma liefern zusammengenommen einen Werkzeugkasten, der der (geisteswissenschaftlichen) Forschung völlig neue methodische Horizonte eröffnet, eine konzeptuelle und technische Grundlage für ein globales Wissensnetzwerk. Was uns trotz allem nicht erspart bleibt, sind die interpretative wissenschaftliche Arbeit und die Neugier, dieses Netzwerk zu erkunden.

### Zum Autor

#### Matej Ďurčo



Matej Ďurčo war maßgeblich am Aufbau des 2015 eröffneten Austrian Centre for Digital Humanities beteiligt, bei welchem er heute Leiter der technischen Abteilung ist. Nach seinem Informatik-Studium an der Technischen Universität Wien war er an diversen DH-Projekten der Österreichischen Akademie der Wissenschaften und der Universität Wien

beteiligt. Er nahm insbesondere an sprachorientierten Projekten teil, wo er sich auf die Erstellung und Erschliessung digitaler Korpora konzentrierte. Speziell zu erwähnen ist sein Engagement für CLARIN (Digital Research Infrastructures for the Arts and Humanities) in zahlreichen Aktivitäten sowohl auf österreichischer wie auf europäischer Ebene. Für seinen Einsatz bei CLARIN erhielt er im Oktober 2014 den ersten «Major Achievements Young Scientist Award».

### Weitere Informationen

Eine ausführliche Version des Artikels befindet sich auf der Website: [www.sagw.ch/bulletin](http://www.sagw.ch/bulletin)

14 <http://dig-hum.de/>

15 <http://dhd2016.de/>

16 <http://www.dhd2016.de/node/9>

17 <http://protege.stanford.edu/>

18 <http://aksw.org/Projects/OntoWiki>

19 <http://www.w3.org/2001/sw/wiki/Category:Editor>

20 <http://www.vocabularyserver.com/>

21 <http://openskos.org/>

22 <http://www.w3.org/wiki/LargeTripleStores>

23 [http://www.w3.org/2001/sw/wiki/Category:Triple\\_Store](http://www.w3.org/2001/sw/wiki/Category:Triple_Store)

24 <http://babelfy.org/>

25 <http://stanbol.apache.org/>

26 <http://spotlight.dbpedia.org/>

27 <https://openlp.apache.org/>

# Geisteswissenschaftliche Forschung und ihre Daten: Stand, Herausforderungen und Perspektiven im digitalen Zeitalter

Freitag, 20. November 2015, Universität Bern (UniS)

09:00–09:30 Eintreffen der Teilnehmenden, Begrüssungskaffee

09:30–09:45 **Begrüssung und Einleitung**  
Walter Leimgruber, Uni Basel

## Teil 1: Forschungsdaten: Was geschieht damit?

### Moderation Teil 1: Enrico Natale

09:45–10:05 **Geisteswissenschaftliche Daten im digitalen Zeitalter: definitorische Zugänge**  
Franziska Frey, Harvard Library

10:05–10:25 **Créer des conditions pour un management des données de recherche réussi: le projet «Research Data Lifecycle Management» dans le cadre du CUS P-2**  
Pierre-Yves Burgi, Uni Genève

10:25–10:45 **Sicherung und Vorhaltung von geisteswissenschaftlichen Forschungsdaten: das «Data and Service Center for the Humanities» und seine Dienstleistungen**  
Lukas Rosenthaler, Uni Basel

10:45–11:05 **Kaffeepause**

11:05–11:35 **Diskussion Teil 1**

## Teil 2: Forschungsprojekte: Welche Auswirkungen hat der digitale Wandel auf die geisteswissenschaftliche Forschung? Überlegungen am Beispiel digitaler Editionen und Sammlungen

### Moderation Teil 2: Markus Zürcher

11:40–12:00 **Standards et best practices pour les projets d'éditions digitales**  
Laurent Pugin, RISM Bern

12:00–12:20 **Standards und best practices für digitale Sammlungen**  
Eveline Wandl-Vogt, Wien (ÖAW)

12:20–12:40 **Empfehlungen für das Aufbereiten von Daten in den Geisteswissenschaften im Hinblick auf Vernetzung und Nachnutzung**  
Martin Volk, Uni Zürich

12:40–13:40 **Mittagspause**

13:45–14:15 **Diskussion Teil 2**

## Teil 3: Forschungsvernetzung: Wie können Forschungsdaten vernetzt werden?

### Moderation Teil 3: Lukas Rosenthaler

14:15–14:35 **Vernetzte Daten und die Praktiken des Historikers – die Suche nach Anreizen und Lösungen**  
Tobias Hodel, Uni Zürich

14:35–14:55 **Qualitative Vernetzung: Die Dienstleistungen von metagrid**  
Sacha Zala, DDS Bern

14:55–15:15 **Vernetzung durch Linked Open Data und standardisierte Vokabularien**  
Thomas Hänslı, Uni Zürich / ETHZ

15:15–15:35 **Kaffeepause**

15:40–16:10 **Diskussion Teil 3**

16:15–17:15 **Podiumsdiskussion: Sicherung und Verfügbarhaltung von Forschungsdaten – welche Bedürfnisse hat die Fachgemeinschaft – welche Daten sind verfügbar zu halten – wer ist dafür zuständig – welche Standards sind zu beachten?**  
Moderation: Madeleine Herren  
Teilnehmende: Gabi Schneider (SUK P-2), Ingrid Kissling (SNF), Nicole Schaad (SBFI), Lukas Rosenthaler (DaSCH), Marianne Hundt (Vertretung Wissenschaft)

17:15 **Schlusswort**  
Markus Zürcher

ab 17:15 **Apéro**