

# Visualising Data Resources

The Potential of a Wikimedia Platform for the  
Digital Humanities

Workshop  
27 – 28 April 2012  
Forschungszentrum Gotha

Bericht



Funded by Wikimedia Germany out of the  
Community-Project Budget

# Visualising Data

## Resources

### Steps towards a Wikimedia Platform for the Digital Humanities

Gefördert mit einem Etat von 7.500 Euro im Rahmen des Community Projekt Budgets veranstaltete das Forschungszentrum Gotha in Kooperation am 27. und 28. April 2011 gemeinsam mit Wikimedia Deutschland einen Workshop, der Chancen einer neuen für Forschung interessanten Wikimedia-Datenbank-Ressource ausloten sollte.

#### Vorgeschichte

Mit der Zusammenführung großer Datenbanken und der fortschreitenden Digitalisierung wuchs in den letzten Jahren (auch) unter Historikern das Interesse an Visualisierungen von größeren Datenmengen. Mit ihnen lassen sich Trends im unüberschaubaren Terrain auszumachen, Erklärungsdefizite aufzuzeigen, Thesen großer Entwicklungen, die man bislang eher als Lebenssumme formulierte genauer erfassen.

Ein immenses Problem ist das Fehlen einer zentralen Plattform. Projekte der Digital Humanities arbeiten mit ähnlichen Datenlagen – etwa Meta-Daten von Katalogisaten. Standardisierbare Korrelationen werden dabei hergestellt: Beziehungsnetze, Zeit-Verlaufskurven, Verteilungen von Befunden auf Landkarten. Weder benutzt man dazu eine gemeinsame Software, noch gelingt es Daten auf Plattformen unkompliziert zusammenzuführen. Andernorts bereits funktionierende Softwarelösungen werden in neuen Projekten neu finanziert und generiert. Forschungsergebnisse bleiben – im prekären Fall vorgelegt als CDs oder auf individuellen Websites – ohne Anschlussforschung.

Eine Trendumkehr lässt sich in den letzten Jahren beobachten: Förderungsinstitutionen drängen zunehmend auf Open-Source Lösungen und Commons-Lizensierungen, um sicherstellen, dass geförderte Arbeit von Nachfolgeprojekten genutzt wird und damit in der Forschung Impulse gibt.

Unklar ist bislang, auf welche Standards und Plattformen man sich einigen kann. Einzelne miteinander rivalisierende individuell ausgehandelte Kooperationen sind das Ergebnis. Neue Projekte erhalten dabei keinen einfachen Experimentraum.

#### Wikimedia und Forscher zusammenbringen: Planung und Idee

Die Idee, Wikimedia für den Aufbau einer größeren Plattform zu gewinnen, geht zurück auf einen *Cultures of Knowledge* Workshop, den ich im April 2010 in Oxford besuchte. Bei den anwesenden Fachleuten stieß der Vorschlag, an Wikimedia heranzutreten, auf gemischte Gefühle. Einerseits ist man in Fachkreisen fasziniert von der Dynamik bisheriger Wikimedia-Projekte und angetan von den Rahmenbedingungen:

- Arbeit wird in Wikimedia Projekten attraktiv lizenziert,

- Zugriff wird Einzelnen wie Projekten unbürokratisch und als Selbstverständlichkeit gestattet
- eine große Community nimmt an hier geschehender Arbeit teil
- Softwarelösungen stehen auf einer solchen Plattform automatisch allen Beteiligten zur Verfügung

Als Problemzone erscheint dagegen der hierarchielose Zugriff. Bisher gestatten wissenschaftliche Projekte Mitarbeitern streng hierarchisierte Zugriffe auf die jeweilige Datenbasis. Über die internen Korrekturmechanismen von Wikimedia-Projekten bestand unter den Zuhörern in Oxford wenig Klarheit, obwohl gerade sie die fruchtbare Ausgangssituation für rapides Wachstum von Projekten sind.

### Antragstellung im Rahmen des Community Projekt Budgets

Der als Ideenbörse anvisierte Workshop wurde im Rahmen des Community Projekt Budgets geplant und realisiert und fällt in die erste Projektphase dieses Finanztopfes.

Die Antragstellung verlief unkompliziert. Die Betreuung durch Sebastian Sooth gestaltete sich im Verlauf menschlich äußerst angenehm und in allen Arbeitsschritten konstruktiv.

Bewilligt wurden 7.500 Euro. Mein Ziel war es, mit diesem Etat Wikipedianer, vor allem aber Verantwortliche von Wikimedia Deutschland und Wissenschaftler vorerst allein aus dem Bereich der Frühe-Neuzeit-Forschung zusammenzubringen.

Angestrebt war ein Erfahrungsaustausch: Was machen Wissenschaftler gegenwärtig mit Daten? Welche Softwarelösung bietet das Wikimedia-Spektrum? Welche Chancen, eine Plattform probeweise aufzubauen, bestehen?

Das Projekt wurde von mir auf einer Plattform in der deutschen wie der englischen Wikipedia, den internen Verlautbarungsorganen und Mailinglisten beworben.<sup>1</sup>

### Interferenz mit dem Open Data Projekt

In der Planungsphase schien anfangs glücklich, dass Wikimedia Deutschland für den Aufbau der neuen Wikidata-Ressource zuständig wurde.

Vorgespräche, die ich mit Mathias Schindler und Daniel Kinzler führte, sowie ein Mail-Austausch mit Denny Vrandečić ließen absehen, dass Wikidata als Plattform geeignet wäre, größere Datenmengen direkt aus Forschungsprojekten aufzunehmen und zu verarbeiten. Die gesuchten Visualisierungen wären eher einfache Standards.

Klar wurde zweitens auf der WikiCon in Nürnberg, im September 2011, auf der ich den Workshop in seiner Planung der Diskussion aussetzte, dass ein Semantic Media-Wiki sich als Zwischenlösung und Experimentierraum anbieten würde.

Im Spätherbst 2011 wurde der ursprünglich für den März geplante Termin nach Beratung mit der Wikimedia-Geschäftsstelle verlegt – es schien desiderabel, den Workshop nach dem Termin auszurichten, an dem die Wikidata-Arbeitsgruppe konstituiert würde.

<sup>1</sup> [http://de.wikipedia.org/wiki/Wikipedia:Bibliographisch-archivalische\\_Datenbank](http://de.wikipedia.org/wiki/Wikipedia:Bibliographisch-archivalische_Datenbank)  
[http://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Bibliographical\\_Database](http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Bibliographical_Database)

Die anlaufende Planung von Wikidata erwies sich wenig später als für den Workshop eher verheerende Ausgangslage. Ab dem Januar 2012 zeichnete sich ab, dass die Wikidata-Projektverantwortlichen als interessierte Gesprächspartner ausfielen.

Ich versuchte auf der 10. Wikimedia-Mitgliederversammlung am 21. April noch persönlich und eine Woche vor dem Workshop hier ein Umdenken zu erreichen und dafür zu plädieren, dass man sich, nachdem die Finanzierung dafür vorhanden ist, dazu entschließt, den bevorstehenden Workshop zumindest als Inspiration zu nutzen. Im kleinen Kreis Wikidata-Projektgruppe, war die Antwort, dass man das eigene Projekt im ersten Arbeitsschritt als Medium zwischen den Wikipedien handhaben werde. Die Softwarelösung, die man anstrebe, sei für eine Kooperation mit Wissenschaftlern ungeeignet, eine Kooperation mit Wissenschaftlern uninteressant, da diese „original research“ betrieben. Indes sei man bereit, das eigene Projekt in Gotha von zwei Wikimedia-Repräsentanten vorzustellen zu lassen: Daniel Kinzler und Mathias Schindler würden die Abgesandten sein.

Mit den Verweisen auf die eigenen Projektziele und die eigene Software-Entwicklung war effektiv ab der Jahreswende klar, dass von offizieller Vereinsseite aus betrachtet, die geplante Veranstaltung ein eher störendes Kooperations-Anliegen aufbrachte.

### Für Wikimedia eine eher prekäre Veranstaltung

Der am 27. und 28. April 2012 in den Räumen des Forschungszentrums Gotha veranstaltete Workshop fand auf Seiten der angesprochenen Wissenschaftler größtes Interesse. Nachdem die auf dem Gebiet führenden Arbeitsgruppen aus dem Bereich der Frühe-Neuzeit Forschung – in Stanford, Oxford und Den Haag – zugesagt hatten, war es unproblematisch, die Einladungen nach Interesse am breiten Spektrum historisch versierter Digital Humanities auszuweiten. Projektpräsentationen deckten am Ende den größeren Zeitraum von der Antike bis in die Gegenwart ab und griffen global aus – Forschungsvorhaben zum Dritten Reich waren ebenso vertreten wie Forschung zum islamischen Raum in Mittelalter und früher Neuzeit.

Martin Rulsch bewillkommnete die Gäste als Wikimedia-Vorstandsmitglied. Mathias Schindler und Daniel Kinzler hielten die Fachvorträge von Wikimedia aus – über Kooperationen generell und das Wikidata-Projekt im Besonderen. Beide Referenten verabschiedeten sich nach ihren Vorträgen aus persönlichen Gründen, was irritierte, wie die Vorträge selbst: Wikimedia stellte sich als erfolgreiches Unternehmen dar, das nicht länger als Bittsteller um Kooperationen warb. Heute könne man den Kooperationswünschen gar nicht mehr nachkommen.

Auf Seiten der eingeladenen Wissenschaftler bestand am anderen Tag Ernüchterung darüber, dass man die eigenen Vorträge nicht mehr hielt, um mit Wikimedia in einen Erfahrungs- und Gedankenaustausch zu treten. Drei Wikipedianer waren zwar als Gäste gekommen. Die Ansprechpartner für Kooperationen hatten sich jedoch verabschiedet.

Spannend blieb im weiteren forschungsinternen Austausch die Wahrnehmung, dass man gegenwärtig fachübergreifend an sehr ähnlichen Problemen mit sehr ähnlichen Lösungsvorschlägen arbeitet. Als fruchtbar erwies sich die Mischung der Teilnehmer aus Forschung und Datendesign-Praxis. Man lotete Möglichkeiten für zukünftige Know-How-Transfers aus.

## Fazit

Als ein Projekt, das Wikimedia und Forschung zusammenbringen sollte, war der Workshop geplant und intern wie extern beworben. Es gelang mir nicht, die Veranstaltung zu mehr denn einer Einbahnstraße werden zu lassen, bei der Wikimedia eine Chance suchte, sich selbst darzustellen und unter den Anwesenden einen eher prekären Eindruck zurückließ. Der Etat – wir werden mit diesem Bericht nicht ausgeschöpfte 926,92 € zurücküberweisen – hätte es erlaubt, die gesamte Belegschaft des Wikidata-Projektes allein für inspirierende Darbietungen einzuladen.

Konsens bestand unter den Teilnehmern des wissenschaftlichen Feldes – der vertanen Chance ungeachtet – darüber, dass eine gemeinsame Plattform gerade auf dem Gelände einer Wikimedia-Ressource prinzipiell attraktiv wäre.

- Benötigt würde ein Semantic-Mediawiki, dem Wikimedia institutionelle Sicherheit bieten würde.
- Spannend wäre die Unterbringung im Wikimedia-Konsortium unter den freien Lizenzbedingungen – sie erlauben es Projekten, die eigene Arbeit kommerziell zu verwerten und sichern gleichzeitig die Weiterentwicklung nach dem jeweiligen Förderungsende.
- Interessant wäre eine Wikimedia-Plattform, da auf ihr Software-Lösungen jeweils nur einmal entwickelt würden und danach der gesamten Community zur Verfügung stünden. Die zentralen Visualisierungen sind mittlerweile Standard im kostenlosen Google-Angebot. Man benötigte eine Semantic-Mediawiki das diese Standards integrierte:
  - die Quantitative Darstellungen von Entwicklungen auf Zeitachsen,
  - die Distributionen von Funden und Beziehungen zwischen Befunden auf Landkarten,
  - Netzwerkanalysen,
  - Darstellungen von Genealogien.
- Aussichtsreich wäre eine Wikimedia Plattform, da sie es Forschungsprojekten gestatten würde, kollektiv an derselben Datenlandschaft zu arbeiten.
- Nötig wäre es, eine solche Plattform in der Anfangsphase von betreut anzubieten – wichtig ist, dass von Experten für die Software erfasst wird, welcher Art Daten hier mit welchem Auswertungsinteresse eingefügt werden.

Für Wikimedia müsste die Kooperation mit Forschungsprojekten interessant sein, da hier

- stabile Projektpartner gewonnen werden, die mit eigenen finanziellen Ressourcen Entwicklungen tragen,
- Wissenschaftler enger an Wikipedia gebunden würden,
- an einer Ressource gearbeitet würde, die im Verlauf Wikidata Daten zur Verfügung stellen würde,
- „Original Research“ erstmals klarer definiert in ein Wikimedia Projekt einzöge.

Theoretisch müsste hier eine Situation beiderseitigen Gewinns bestehen.

## Finanzielle Rechenschaft Kurzüberblick

Zu den folgenden Punkten bieten das anliegende Veranstaltungsprogramm und die der Geschäftsstelle bereits zugegangene Abrechnung mit allen Belegen den ausführlichen Hintergrund. Aus den Anlagen geht hervor, wer worüber sprach. Hier nur die Summen und die Namen zu den Belegen:

20 Vortragende:	James Brown Mark Curran Neil Jefferies Karsten Krummrück Martin Mulsow Judith Pfeiffer Leif Scheuermann Olaf Simons Charles van den Heuvel Scott Weingart	Nicole Coleman Wiebke Hemmerling Daniel Kinzler Thea Lindquist Guido Naschert Martin Rulsch Mathias Schindle Wolfgang Spickerman Jens Weber Andreas Wolter
6 Gäste:	Steffen Prößdorf Anne Simone Knöfel A. McKenzie-McHarg	Jan Eissfeldt Mathias Rekow Riccarda Suitner
13 Reisekosten-Abrechnungen	3903,56 € James Brown Mark Curran Wiebke Hemmerling Karsten Krummrück Judith Pfeiffer Leif Scheuermann Andreas Wolter	Nicole Coleman Jan Eissfeldt Neil Jefferies Thea Lindquist Martin Rulsch Charles van den Heuvel
Zusätzliche Aufwendungen	2669,52 € Für Verpflegung, Unterbringung und Organisation – siehe die Detailaufschlüsselung der Abrechnung	
Budget	7500 €	
Rückzuerstattender Überschuss:	926,92 €	

## Anlage

Projektvorstellung im Antragstadium, Vorstellung der einzelnen Workshop-Referate und Teilnehmer, am Ende realisiertes Veranstaltungsprogramm

Die detaillierte Aufschlüsselung und einzelne Rechnungsbelege gingen der Geschäftsstelle bereits zu.

Olaf Simons  
Gotha 3. November 2012

# Projekt Strukturierte Daten

## Detailprojekt Datenbank bibliographischer Objekte

### Strukturierte Daten

Momentan befindet sich ein Projekt erheblicher Dimension im Wikipedia Konsortium in Planung: Eine Ressource „strukturierter Daten“.

Es handelt sich dabei um das Projekt einer zentralen Datenbank, die beliebige Objektdaten sammelt und diese *strukturiert* verwaltet – ein Projekt gewaltiger Dimension und weitreichender Implikation. „Objekte“ können sein: Der Mond, das Kraftfahrzeug einer bestimmten Baureihe, ein Buch, das über Bibliothekskataloge recherchierbar ist.

Objektdaten sind theoretisch beliebig definiert: Gewicht des Mondes, Daten seiner Erdumlaufbahn, Zahl und Spezifikation der Einzelteile im genannten Kraftfahrzeug, Leistungsdaten seines Motors, Bibliotheksstandorte des recherchierten Buches, Angaben zu Seitenzahl, Format, Inhalt, Genre, Autor, Übersetzer, Vorlage und so fort.

Die Daten wären – als freies Wissen – beliebig nutzbar. Wikipedia, die Enzyklopädie, speiste etwa alle einzelsprachlichen Artikel zum Mond aus der Ressource. Verändert man in der zentralen Ressource ein Datum, wird dieses in allen Artikeln, die dieses Datum nutzen aktualisiert.

Der Nutzen außerhalb der Online-Enzyklopädie ist im Moment vollkommen unabsehbar. Momentane Suchmaschinen und Übersetzungsprogramme arbeiten ohne „Weltwissen“, das heißt sie übersetzen von einer Sprache in die andere, ohne zu wissen, warum es geht, suchen Äquivalenten für Sprachartikel. Zukünftige Programme sollten erkennen, wovon gesprochen wird, Verständnis-Rückfragen stellen können und dann Antworten aus der zentralen Wissensressource generieren können.

## Generelles Anliegen: Das Projekt sollte frühzeitig wissenschaftlich betreut werden

Die deutsche Wikimedia Community sollte versuchen, Fachleute in Kontakt miteinander zu bringen, um Interessen an einer solchen Ressource zu bündeln und frühzeitig an die Entwicklung zu binden. Wichtig wäre es, Vorstellungen davon zu entwickeln, wo die Ressource Nutzung finden könnte und welchen Anforderungen sie dabei genügen sollte, um Entwicklungsspielräume nicht zu verschenken. Mitunter sind Nutzungen einfach, wurden aber schlicht nicht vorgesehen, von denen die an die technische Realisation dachten, und die keine Vorstellung vom späteren Interesse an ihrer Arbeit hatten.

Im Moment sehe ich hier einen mehrseitigen Informationsbedarf – vom Aufbau der neuen Ressource hört man ein Raunen, während es spannend wäre, sich abzeichnende Entwicklungen frühzeitig zu sehen und konstruktiv von den späteren Interessenten aus zu begleiten.

## Zweites und konkretes Projekt: Der Aufbau einer zentralen Ressource bibliographisch-archivalischem Wissens

Das nachfolgend näher skizzierte Projekt würde man am effizientesten im Rahmen des skizzierten großen Projektes realisieren. Es gilt dem Detailbereich der momentan weltweit disparat primär von nationalen Ressourcen verwalteten bibliographischen und archivalischen Objektdaten.

Mit dem Aufbau der Bibliotheks-Online-Kataloge wurden Bücher weltweit recherchierbar. „Metakataloge“ führen dieses Wissen derzeit unterschiedlich zusammen:

- Ressourcen wie der KVK – der Karlsruher Virtuelle Katalog – greifen auf verschiedene Onlinekataloge zu und stellen Suchanfragen an sie. Der Benutzer erhält als Ergebnis die Informationen der befragten Ressourcen nach den Anbietern geordnet.
- Dem stehen – weitaus interessanter – Ressourcen gegenüber wie die gegenwärtig im Aufbau befindlichen Nationalkataloge. Bei ihnen werden Kataloge unter erheblichem Arbeitsaufwand zusammengeführt: Man erstellt von einem mehrfach verfügbaren Objekt (Büchern einer Auflage, die heute in vielen Bibliotheken liegen) einen einzigen Datensatz, der unter anderem notiert, welche Bibliotheken den Titel vorrätig haben.



Die zweite Form der Kataloge ist zukunftsweisend, da in ihr Informationen sinnvoll zusammengefügt werden. Deutlich wird das, wenn man auf die Fragen sieht, die sich erst an Kataloge des zweiten Typs stellen lassen: Man kann den ESTC (den English Short Title Catalogue) für alle auf Englisch respektive im englischen Sprachraum zwischen 1473 und 1800 gedruckten Titel statistisch auswerten. Wie viele Titel wurden im Jahre 1700 gedruckt? Wie viele davon lassen sich als Romane einstufen? Wieviele Bücher wurden in Edinburgh gedruckt? Wie viele in London? Jede einzelne Auflage eines Titels ist im ESTC nur einmal notiert. Eine Ressource wie der KVK listet dagegen den einen gesuchten Titel so oft, wie sie positive Rückmeldungen von konsultierten Katalogen erhielt – die Treffermenge entscheidet sich über die Zahl ausgewählter Kataloge. Präzise Fragen lassen sich an den menschlich erstellten Katalog stellen: Wie viele Verleger arbeiteten an einem Ort zum beliebigen Zeitpunkten simultan? Wie viele Bücher druckte ein Verleger durchschnittlich etc.

- Kataloge wie der englische ESTC, der niederländische STCN, die deutschen Jahrhundertkataloge VD16, VD17 und, soeben in Konstruktion, VD18 werden soeben in nationalen Interessen erstellt. Man will mit ihnen mehr über das Schrifttum der eigenen Nation und Sprache erfahren.
- Die Produktion der Gesamtkataloge geht derzeit von Bibliothekskooperationen aus – entsprechend bibliothekarisch sind die Anfragemöglichkeiten formuliert.
- Benutzer sind bislang durchweg von Mitarbeit an diesen Ressourcen ausgeschlossen.<sup>1</sup>

Der Ausschluss der Benutzer hat gravierende Konsequenzen. Um Korrekturen (man hielt etwa den Titel in der Hand, sah, dass das Titelblatt falsch abgeschrieben wurde, kann den Titel präziser datieren, weiß, wer der Autor ist) kann man per e-mail einreichen, sein Wissen vergibt man dabei ohne weitere Spuren zu hinterlassen. Fachwissen kann in die Ressourcen nicht einfließen: Sie sehen nicht vor, dass Informationen begründet werden – man erhält etwa verschiedenen Angaben zu möglichen Autoren, nicht aber in Fußnoten die Gründe für die divergierenden Mutmaßungen. Fachliteratur muss man in anderen Ressourcen recherchieren etc.

Komplexes Wissen bleibt in diesen Katalogen aus: man würde sich bei einer Recherche wie „Robinson Crusoe“ eine strukturierte Publikationsgeschich-

---

<sup>1</sup> Ich sprach jüngst mit dem Leiter der Bibliothek Wolfenbüttel, und erfuhr dass man hier derzeit an eine Öffnung denkt, an eine Kooperation mit Wikipedia dachte man bislang dort nicht.

te wünschen: Was ist die Erstausgabe, wie liegt die exakte Chronologie ermittelter Folgeauflagen, welche Raubdrucke und Übersetzungen folgten welchen Vorlagen – die Kataloge geben allenfalls einen maschinell chronologische geordneten Überblick, den sie aus den internen Daten generieren, nicht die tatsächliche Publikationsgeschichte des Titels.

Prekär sind die nationalen Grenzziehungen: Der ESTC nennt die englischen *Robinson Crusoe* Ausgaben grob geordnet, er erlaubt keinen Blick auf die europäische Verbreitung. Sie muss in den anderen Nationalkatalogen recherchiert werden.

Speziell bei statistischen Auswertungen lassen die aktuellen Kataloge Wünsche offen – die Daten sind vorhanden, dass jedoch jemand wissen will, wie sich der Buchdruck an einem Ort entwickelte, sah man nicht vor – die komplexeren Suchanfragen lassen sich oft nicht stellen, obwohl alle Detaildaten im System liegen.

#### Desiderat: Der von Benutzern betreute und fortentwickelte Metakatalog

Interessant wäre aus Sicht der Forschung ein bibliographisches Rechercheinstrument, das von den Benutzern mitgestaltet würde und die Informationen der nationalen Projekte konstruktiv zusammenführte. Es sollte möglich sein, in dieser Ressource

- Daten manuell zu verknüpfen und zu strukturieren (man erhält etwa bei der Erstausgabe die Publikationsgeschichte samt Links in die Folgeauflagen, Übersetzungen und Digitalisate).
- Daten zu korrigieren.
- Daten zu kommentieren und der kritischen kollektiven Überprüfung auszusetzen.
- Wissen zu Objekten einzuspeisen (Inhaltsangaben, Exzerpte, Verweise auf Fachliteratur sind hier wertvoll).
- Forschung an der Datenbank in einem eigenen Projektnamensraum an sie anzubinden

Der letzte Punkt mag mit zwei Beispielen illustriert sein. Promovenden, die mit Büchern arbeiten geben über ihre Projekte bislang auf Universitätsseiten und Seiten der Stipendienggeber Auskunft. Sinnvoll wäre es, sie vernetzen ihre Arbeit mit Materialien in der Ressource selbst mit ihrem Projekt. Spannend wäre die Interaktion mit den Forschungsprojekten: Momentan versuchen verschiedene Projekte Perspektiven auf die Briefzirkulation unter Gelehrten der frühen Neuzeit zu geben. Die Briefe liegen in verschiedenen Bibliotheken und Archiven, man katalogisiert sie, wertet sie inhaltlich

aus, vernetzt sie und versucht Einblick in die Netzwerkstrukturen und deren dynamische Entwicklung zu geben.<sup>2</sup> Die verschiedenen Forschergruppen bauen sich dazu jeweils eigene Datenbanken auf, Insellösungen, die rasch veralten und deren Informationen aufgrund der insolierten technischen Standards nicht mehr in Nachfolgeprojekte fließen. Spannend wäre die Ressource die zentral die Datenlage produziert und die Arbeit *mit* der Datenlage *an* sie anbindet.

### Kommunikationsangebote, die von uns ausgehen müssten

Nötig wäre es, zügig auf Kooperationspartner zuzugehen, die Daten zur Verfügung stellten. In Gesprächen die ich im Vorfeld des VD18 führte, erhielt ich Signale einer Bereitschaft, die Daten mit einem Projekt wie Wikipedia zu teilen, es ist unklar, wer heute mit uns kooperieren würde.

Interessant wäre es, das bibliographische Projekt gemeinsam mit Google Books zu realisieren. Google Books ging breite Kooperationen mit Institutionen wie der Bayerischen Staatsbibliothek ein. Der Zugriff auf die Digitalisate geschieht derzeit über eine Datenbank, die mehr an Youtube und Amazon erinnert, denn an einen sinnvollen Katalog. Man sucht Surfer, die sich vielleicht auch für das nächste ähnlichen Link interessieren. Man hofft wie bei Amazon auf Benutzer, die eine „Review“ schreiben, eine Kurzbeurteilung durch den Kunden – ein Kuriosum in einem Medium, das Wissenschaftlern Bücher aus der frühen Neuzeit zur Verfügung stellt. Interessant wäre es, Google-Books anzubieten, den Zugriff auf die Digitalisate durch eine interaktive bibliographische Community-gestützte Ressource zu optimieren.

### Nächstliegendes Antragsziel

wäre nach dem gesagten eine Konferenz auf der Fachkreise über die laufende Planung informiert würden und in einem Workshop Ideen generierten, in welchen Richtungen das Projekt Entwicklungsspielräume gewinnen sollte.

Kostenfaktoren wären hier die Organisation eines Konferenzortes (etwa Bibliothek Göttingen), die Kommunikation mit Teilnehmern im Vorfeld (Recherche, Kontaktaufnahmen, Vorabgespräche), die Finanzierung von Reisen und Übernachtungen.

[#akuter Beratungsbedarf meinerseits o.s.]

---

<sup>2</sup> Siehe etwa das Stanford-Projekt: <http://shc.stanford.edu/intellectual-life/video-podcasts/detail/tracking-18th-century-social-network-through-letters>



UNIVERSITÄT  
ERFURT



## Visualising Data Resources

### The Potential of a Wikimedia Platform for the Digital Humanities Workshop

Forschungszentrum Gotha der Universität Erfurt

Seminarraum, 27 – 28 April 2012

Funded by Wikimedia Germany out of the Community-Project Budget

## Participants and Work

### **Martin Mulsow (Forschungszentrum Gotha)**

...is the head of the *The Research Center for Social and Cultural Studies in Gotha* affiliated to Erfurt's university. His research focuses on the correspondence and exchange of ideas in early modern Europe and is informed by methods derived from research into constellations, that is small groups engaging in mutual intense communication, where the interest has most often been directed towards underground communication occurring within milieus characterized by clandestine activities, and peopled by free thinkers, spies, diplomats, alchemists and religious separatists.

<http://www.uni-erfurt.de/forschungszentrum-gotha/mitar/mulsow/>

[martin.mulsow@uni-erfurt.de](mailto:martin.mulsow@uni-erfurt.de)

### **Olaf Simons (Forschungszentrum Gotha)**

Book historian on a research grant at the *Gotha Research centre* is presently working on the concept of literature as a historical category organizing the book market – and is also a Wikipedian since 2004 with a history in different web projects.

I have been involved in the technical planning of Martin Mulsow's project on constellations of academic exchange and will (in the course of this workshop) try to bring different communities together: the community of Wikimedia developers and Wikipedians and the community of scholars and technicians working in the Digital Humanities on the early modern period.

[http://de.wikipedia.org/wiki/Benutzer:Olaf\\_Simons](http://de.wikipedia.org/wiki/Benutzer:Olaf_Simons)

<http://www.uni-erfurt.de/forschungszentrum-gotha/grad/dr-olaf-simons/>

**Martin Rulsch (Wikimedia Germany, Berlin)**

organized a conference [Wikipedia:Wikipedia trifft Altertum](#) and is presently a board member of Wikimedia Germany. Martin Rulsch is otherwise an expert on the ancient world and Greek and Roman literatures.

<http://de.wikipedia.org/wiki/Benutzer:DerHexer>

[martin.rulsch@wikimedia.de](mailto:martin.rulsch@wikimedia.de)

**Mathias Schindler (Wikimedia Germany, Berlin)**

project manager at Wikimedia has supervised major joint ventures between Wikimedia-projects and public institutions and will speak about the chances and risk of projects that involve Wikipedia/Wikimedia.

[http://de.wikipedia.org/wiki/Benutzer:Mathias\\_Schindler](http://de.wikipedia.org/wiki/Benutzer:Mathias_Schindler)

[mathias.schindler@wikimedia.de](mailto:mathias.schindler@wikimedia.de)

**Daniel Kinzler (Wikimedia Germany, Berlin)**

has been a software developer at Wikimedia since 2008, he is presently part of the team that develops the new Wikidata platform

<http://de.wikipedia.org/wiki/Benutzer:Duesentrieb>

[daniel.kinzler@wikimedia.de](mailto:daniel.kinzler@wikimedia.de)

**James Brown (University of Oxford), Neil Jeffries (Bodleian Library), Nicole Coleman (Stanford University): Cultures of Knowledge**

A presentation on Cultures of Knowledge but the underlying Object model is a generic one with much wider applicability (for example Judith Pfeiffers' IMPACT). Projects of Neil Jeffries include Cultures of Knowledge, IMPACT, Bodleian Ballads, DataFlow (Data Deposit and Repository), DAMARO (Data Management Rollout at Oxford), DMSTech/Shared Canvas and IIF.

**Contact Details**

Neil Jeffries

R&D Project Manager for the Bodleian Libraries

[neil.jeffries@bodleian.ox.ac.uk](mailto:neil.jeffries@bodleian.ox.ac.uk)

James Brown

[james.brown@history.ox.ac.uk](mailto:james.brown@history.ox.ac.uk)

Nicole Coleman

[cncoleman@stanford.edu](mailto:cncoleman@stanford.edu)

**Charles van den Heuvel (Huygens Institute, Den Haag), Scott Weingart (Indiana University):  
Desiderata for Visualizing Uncertainty and Contextuality in the Digital Humanities. Experiences  
with the Digital Republic of Letters**

In preparation for a recent conference organized by The Cultures of Knowledge project of the University of Oxford, Nicole Coleman, Johanna Drucker, Paolo Ciuccarelli and Charles van den Heuvel explored the visualization of uncertainty and complexity in the digital humanities. This exploration was inspired by the article “Humanities Approaches to Graphical Display” of Johanna Drucker with the intention to apply expressions of uncertainty and complexity to the study of the correspondences of letters in the research projects The Republic of Letters (Stanford Humanities Research Center) and Circulation of Knowledge (HuygensING/Descartes Centre UU) as input for the design of a new graphical language by and with the Density Design studio. Here, we try in addition to this expressed need for a humanities graphical language based on qualitative and subjective information to formulate some desiderata for visualizations that contextualize quantitative information and that might serve as a bridge between the various approaches. Humanities data often lie balanced between wholly subjective interpretations and fairly clear self-evidence, but even the most agreed upon data points have degrees of uncertainty. Data uncertainty comes in many flavors; it may have an untrustworthy provenance, it may be incomplete or, perhaps, there may be competing expert opinions. Any visualization or analysis that uses imperfect data without taking that uncertainty into account is at risk of fundamentally misrepresenting our knowledge of the world. We argue that the solution is two-fold: modeling uncertainty into the data themselves, and contextualizing the data against encompassing or related datasets.

Here, we formulate desiderata for a humanities graphical language which embraces qualitative and subjective information while simultaneously allowing for the comparison and contextualization of quantitative data. To this end, we will discuss experiences with visualizations made within the context of the Circulation of Knowledge project, as well as in comparison with the Cultures of Knowledge project, along with expectations of some new (initiatives for) e-history of science projects within Huygens ING, Digital Grotius and The Transition from Letters to the Periodical.

Charles van den Heuvel is Head Research of History of Science at the Huygens Institute for the History of the Netherlands of the Royal Netherlands Academy of Arts and Sciences where he is involved in various research projects on annotation and visualization in the history of science and on concept extraction from 17th century correspondences of Dutch scholars. Furthermore, he makes part of the e-humanities research group of the Royal Netherlands Academy of Arts and Science. He studied Art History and Archaeology at Groningen University, The Netherlands (PhD 1991). He publishes regularly on history of architecture, fortification and town planning, history of cartography, history of science and history of information science. Themes of interest in the latter discipline are the history of classification, history of the WWW and the history of visualizations of knowledge. Currently, he is writing a book entitled “Imagining Interfaces to Universal Knowledge. The visualizations of Paul Otlet (1868-1944)”.

Scott Weingart is an NSF Graduate Research Fellow and Ph.D. student at Indiana University, where he studies Information Science and the History of Science. His research focuses on the intersection of historiographic and quantitative methodologies, particularly as they pertain to the study of scholarly communication in the past and present. As a member of Indiana University’s Information Visualization Lab and the Cyberinfrastructure for Network Science Center, Scott has published on modern publication and citation analyses. He has also collaborated closely with various international organizations on visually representing and quantitatively analyzing early modern correspondence networks. In his spare time, Scott writes a blog called the scottbot irregular, aiming to make computational tools and big data analytics accessible to a wider, humanities-oriented audience.

**Contact Details**

Dr. C.M.J.M. van den Heuvel  
Head Research Group History of Science and Scholarship  
Huygens Institute for the History of the Netherlands  
Royal Netherlands Academy of Arts and Sciences

[www.huygens.knaw.nl](http://www.huygens.knaw.nl)

[Charles.van.den.heuvel@huygens.knaw.nl](mailto:Charles.van.den.heuvel@huygens.knaw.nl)

+31-70-3315823 (direct)

+31-70-3315800 (general)

In office: Mon, Tue, Thu, Fri

Out of office Wed (contact number)

+31-43-3101455

Scott B. Weingart

Ph.D. Student, School of Library and Information Science

History of Science Liaison, Cyberinfrastructure for Network Science Center

Indiana University

[scbweing@indiana.edu](mailto:scbweing@indiana.edu)

[www.scottbot.net](http://www.scottbot.net)

**Mark Curran (University of Cambridge): Mapping the Trade of the Société Typographique de Neuchâtel**

Mark Curran has spent the past five years working on The French Book Trade in Enlightenment Europe, 1769-1794, an initiative which utilises cutting edge database technology to map the trade of the Swiss publishers the Societé de Neuchâtel. The resulting database provides a remarkable insight into the francophone book trade of the late-eighteenth century. During his tenure as Munby Fellow, Dr Curran is piloting adding other kinds of sources to the database - from bookseller's catalogues to newspaper reviews - pioneering a wider system for the study of the dissemination and reception of books. His work aims to link realised book-history projects by embedding common and core aspects of their datasets into library catalogues and digital content providers and aggregators.

**Contact Details**

Dr. Mark Curran

Munby Fellow in Bibliography, University of Cambridge

[www.c18booktrade.com](http://www.c18booktrade.com)

[mark@c18booktrade.com](mailto:mark@c18booktrade.com)

**Thea Lindquist (University of Colorado): Primary Source Corps and Linked Data Demonstration Project – World War I Case**

Primary Source Corps is a gamification project meant to engage students and other users with online primary source material. The concept is based upon the Army Corps of Engineers – players work together to complete missions and, through these actions, build resources that both increase the findability of and create context for primary sources in digital library collections. At the same time, they are contributing to the common good by helping to build content and connections in Wikisource and Wikipedia.

Linked Data Demonstration Project – World War I Case, undertaken with the Semantic Computing Research Group (Aalto University, Finland), is meant to showcase how deep linking historical data can create context for and enhance access to online primary sources. This approach allows the association of related concepts in the sources using a specialized vocabulary, enriching them with outside resources, and enabling semantically rich services that empower users.

Both projects use the University of Colorado Boulder's World War I Collection Online as a test bed: <http://libcudl.colorado.edu/wwi/index.asp>. The collection comprises nearly 1,200 titles (over 55,000 pages) published from 1914 to 1920, originating from a variety of geopolitical regions and covering a wide range of topics, from ethnic and religious conflict to propaganda and nationalism.

#### **Contact Details**

Thea Lindquist  
History Librarian / Associate Professor  
University of Colorado Boulder  
[thea.lindquist@colorado.edu](mailto:thea.lindquist@colorado.edu)

#### **Leif Scheuermann, Wolfgang Spickermann (Universität Erfurt): AIDA – Adaptive, Interactive, Dynamic Atlas on History**

The main aim of the AIDA is the development of a framework for a database providing a, freely extensible and inter=ADactive atlas for the history of Europe for research and education. The dynamic maps will enable the visualization of spatial and temporal changes and give a better insight in historical processes. Main sources of this atlas will be various databases provided by the project as well as already existing external databases with are connected via an interface to the project. Basis for this data exchange will a meta data model defined by ontologies that are already standardized and already widely used. The end-product will be a web-portal, including a detailed user management, an upload function for new maps, a web dynamic mapping system and communication tools to dis=ADcuss and comment the maps. To guarantee a high quality level front end editing will be explicitly not be supported and the upload function will be restricted to registered and trained persons.

#### **Contact details**

Leif Scheuermann  
[l.scheuermann@ike.uni-stuttgart.de](mailto:l.scheuermann@ike.uni-stuttgart.de)

#### **Wiebke Hemmerling (Akademie der Wissenschaften zu Göttingen): Scholarly journals as Networks of Knowledge in the Age of Enlightenment**

In 2011 the Academy of Sciences and Humanities in Göttingen started a long-term project that will tackle the task of indexing a representative scope of 18th century scholarly journals. The project is carried out in cooperation with the State and University Library in Göttingen, the Leipzig University Library, and the Bavarian State Library Munich. A database already presenting a collection of 250 periodicals containing more than 80.000 book reviews and about 100.000 scholarly papers, announcements, reports of scientific discoveries and projects, news of learned institutions and persons, provides access to a specific part of the periodical press of the 18th century, the 'key medium' of the Enlightenment period.

#### **Contact Details**

Wiebke Hemmerling



Rubenow 25  
17392 Boldekow  
039722-20311  
[hemmerling@freenet.de](mailto:hemmerling@freenet.de)

**Judith Pfeiffer (University of Oxford): From Late Medieval to Early Modern: A Wishlist for Visualizing the Transition in Islamic Intellectual History**

A presentation of the ERC-Projekt IMPACT (Islamic Medieval Philosophy And Theology: <http://impact.orient.ox.ac.uk/>), a database which is growing structurally and with regard to content. Its visualizations have not yet been realized, but "futuristic" ideas will be presented.

**Contact Details**

Judith Pfeiffer  
Faculty of Oriental Studies  
Pusey Lane  
Oxford, OX1 4LE  
UK  
[judith.pfeiffer@orinst.ox.ac.uk](mailto:judith.pfeiffer@orinst.ox.ac.uk)  
+44 (0)1865 240 782  
+44 (0)7775 434 836

**Jens Weber, Andreas Wolter (mediaarchitecture, Bauhaus Universität Weimar): The ImpulsBauhaus Project as an Example for Visualizing Social Networks**

The ImpulsBauhaus project investigates the social network of the Bauhaus and its global influence. Extensive biographical information was collected and stored on a specially developed research platform. With the aid of computer-generated information graphics and an interactive table, the initial results of the project were presented at the Exhibition NÂ°1 in Weimar 2009.  
<http://impulsbauhaus.de>

"mediaarchitecture.de" is an award-winning, Weimar-based studio operating at the intersection of art, design and technology. Some of the studio's projects resemble classic commercial work, but venture regularly into art, design or technology research. "mediaarchitecture.de" was founded in 2006 by Jens Weber and Andreas Wolter in Weimar.

**Contact details**

Jens Weber  
[weber@mediaarchitecture.de](mailto:weber@mediaarchitecture.de)  
<http://mediaarchitecture.de>  
[ma@mediaarchitecture.de](mailto:ma@mediaarchitecture.de)

# Visualising Data Resources

The Potential of a Wikimedia Platform for the  
Digital Humanities

Workshop  
27 – 28 April 2012  
Forschungszentrum Gotha

Preliminary Programme



Funded by Wikimedia Germany out of the  
Community-Project Budget

# Visualising Data

## Resources

Steps towards a Wikimedia Platform for the Digital Humanities

Friday, 27<sup>th</sup> April 2012

15:00-15:15	Martin Mulsow (Forschungszentrum Gotha) Welcome Address
15:15-15:30	Welcome Address: Wikimedia Martin Rulsch
15:30-16:30	Introduction: Why we are here Olaf Simons
16:30-17:00	Coffee break
17:00-17:30	Mathias Schindler (Wikimedia, Berlin): Wikimedia Cooperations
17:30-18:00	Daniel Kinzler (Wikimedia, Berlin): Wikidata
19:00-21:00	Dinner

Saturday, 28<sup>th</sup> April 2012

9:30-10:00	James Brown (University of Oxford), Neil Jefferies (Bodleian Library, Oxford), Nicole Cole (Stanford University)
10:00-10:30	Charles van den Heuvel (Huygens Institute, Den Haag), Scott Weingart (Indiana University): Desiderata for Visualizing Uncertainty and Contextuality in the Digital Humanities
10:30-11:00	Mark Curran (University of Cambridge): Mapping the Trade of the Société Typographique de Neuchâtel
11:00-11:30	Discussion

12:00-13:30	Lunch break
13:40-14:00	Thea Lindquist (University of Colorado): Primary Source Corps and Linked Data Demonstration Project – World War I Case
14:00-14:30	Leif Scheuermann, Wolfgang Spickermann (Universität Erfurt): AIDA – Adaptive, Interactive, Dynamic Atlas on History
14:30-15:00	Discussion
15:00-15:15	Coffee break
15:15-15:45	Judith Pfeiffer (University of Oxford): From Late Medieval to Early Modern: A Wishlist for Visualizing the Transition in Islamic Intellectual History
15:45-16:15	Wiebke Hemmerling (Systematischer Index zu deutschsprachigen Rezensionen-Zeitschriften des 18. Jahrhunderts, Akademie der Wissenschaften zu Göttingen): Scholarly journals as networks of knowledge in the Age of Enlightenment
16:15-16:45	Guido Naschert: Guido Naschert (Forschungszentrum Gotha): An open database for European visitors books
16:45-17:15	Jens Weber, Andreas Wolter (mediaarchitecture, Bauhaus Universität Weimar): The ImpulsBauhaus Project as an Example for Visualizing Social Networks
17:15-17:45	Karsten Krumrück (Wikipedia/ PSIPENTA Software Systems GmbH), Semantic Mediawiki – the next Level of Wikipedia
17:45-19:00	Discussion
19:30-22:30	Dinner