



UNIVERSITÄTSBIBLIOTHEK



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

DATAVERSE IN GERMANY UND DIE UNIVERSITÄTSBIBLIOTHEK HEIDELBERG IN DEN VEREINIGTEN STAATEN

*Bericht über das Dataverse Community Meeting, 11. Bis 13. 7. 2016, Harvard Medical School
Jochen Apel, Universitätsbibliothek Heidelberg, apel@ub.uni-heidelberg.de*

Vom 11. Bis 13 Juli 2016 fand zum zweiten Mal das Dataverse Community Meeting an der Harvard University statt. Dataverse ist eine am Institute for Quantitative Social Science (IQSS) Harvard entwickelte Open-Source-Software für Forschungsdatenrepositorien.¹ Die Dataverse-Community aus Softwareentwicklern, Repositorienbetreibern, Datenkuratoren und Wissenschaftlern ist eine stetig wachsende und mittlerweile beachtlich große Gemeinschaft, die auch über spezifische Fragen der Dataverse-Software hinaus in den nationalen und internationalen Diskussionen zum Forschungsdatenmanagement aktiv ist und das Thema vorantreibt.

Insgesamt gibt es derzeit weltweit 19 Installationen der Dataverse-Software, weitere an verschiedenen Standorten sind in Planung. Während das Dataverse der Harvard University allen Wissenschaftlern weltweit kostenfrei für die Veröffentlichung ihrer Forschungsdaten zur Verfügung steht,² bieten andere Betreiber ihr System nur für spezifische Nutzergruppen als Publikationsplattform an. So ist das DataverseNL eine nationale Lösung für verschiedene niederländische Hochschulen³ oder unser heiDATA ein institutionelles Forschungsdatenrepositorium für die Wissenschaftler Universität Heidelberg.⁴

Knapp 200 Personen waren der Einladung des Dataverse-Teams zum zweiten Community Meeting nach Boston gefolgt, überwiegend natürlich Teilnehmer aus den USA, aber auch zahlreiche internationale Konferenzteilnehmer fanden den Weg in die Vereinigten Staaten: aus China, Singapur, Brasilien, Kanada, den Niederlanden, und Norwegen waren Teilneh-

¹ <http://dataverse.org/>

² <https://dataverse.harvard.edu/>

³ <https://dataverse.nl/dvn/>

⁴ <https://heidata.uni-heidelberg.de/dvn/>

mer vertreten und berichten in Vorträgen vom Stand des Forschungsdatenmanagements in ihren Einrichtungen und Herkunftsländern. Aus Deutschland durfte ich mich dank eines Reisekostenstipendiums von BI International auf den Weg machen, um die Heidelberger Forschungsaktivitäten in einem Vortrag und auf einem Poster zu präsentieren. Erfreulich war neben der internationalen Ausrichtung der Tagung zudem die bunte Mischung des Teilnehmerfeldes aus Informatikern, Bibliothekaren und vor allem Wissenschaftlern aus unterschiedlichen Disziplinen, die ihre jeweiligen Perspektiven auf das Thema einbrachten. Eine Übersicht über die unterschiedlichen Vorträge, in vielen Fällen inklusive der Präsentationsfolien, findet sich auf der Webseite der Veranstaltung⁵, ein Twitter-Storify gibt einen lebendigen Einblick in die Diskussionen während der Veranstaltung.⁶

Eröffnet wurde die Tagung mit einer Keynote Lecture von Sarah Thomas, Vice President for the Harvard Library und University Librarian sowie Roy E. Larsen Librarian for the Faculty of Arts and Sciences. Thomas skizzierte in Ihrem Vortrag zunächst wie sich die Harvard Library – ein gigantisches Bibliothekssystem mit einem Jahresbudget von 175 Mio \$, 750 Vollzeit-äquivalenten, einem Bestand von 20 Mio Medieneinheiten und einem Zugang von 250.000 Einheiten pro Jahr, den Herausforderungen stellt, die das neue Aufgabenfeld des Forschungsdatenmanagements für die Bibliotheken bringt und vor welche Schwierigkeiten dabei selbst eine so gut ausgestattete Einrichtung wie Harvard gestellt wird. Die Harvard Library hat eine digitale Strategie formuliert, in der sie u.a. als Ziel ausgibt, die Bibliotheksunterstützung im Bereich des Forschungsdatenmanagement auszubauen. Dazu gehört neben der Mitarbeit im Bereich der Datenkuration für das Harvard'sche Dataverse die Verabschiedung von Guidelines und Policies, die Unterstützung von Wissenschaftlern bei der Erstellung von Datenmanagementplänen und die Veranstaltung eines Data Symposiums. Dies sind alles wichtige Aspekte einer bibliothekarischen Forschungsdateninfrastruktur, die man auch in Deutschland mittlerweile an vielen Universitäten vorfindet oder zumindest im Aufbau hat. Selbst in Harvard ist aber die Etablierung entsprechender Infrastruktur und Fachkompetenz nicht leicht umzusetzen, insbesondere falle es schwer entsprechende Personalressourcen zu generieren. Im Anschluss gab Thomas, wie sie unumwunden zugab, ein wenig neidisch, einen Überblick über die Aktivitäten an anderen Standorten im Land, von den Forschungsdatenservices der Universitäten von Minnesota und Virginia bis zu den Berufsbildern des GIS Librarian und des Reproducibility Librarian im Datenmanagementteam der NYU.

⁵ <http://projects.iq.harvard.edu/dcm2016>

⁶ <http://projects.iq.harvard.edu/dcm2016/storify>

Im weiteren Verlauf löste sich die Tagung dann von der rein bibliothekarischen Perspektive auf das Thema und nahm zu einen allgemeine Entwicklungen im Forschungsdatenmanagement wie die FAIR Data Principles⁷ oder die Aktivitäten der RDA in den Blick.⁸ Zum anderen wurde die Nutzung und Weiterentwicklung der Dataverse-Software und Infrastruktur aus verschiedenen Blickwinkeln beleuchtet. Aus Harvard berichteten insbesondere eine Reihe von Forschern über ihre Erfahrungen mit Datenpublikationen. Vorgestellt wurden Use Cases, die auch die Kombination von Dataverse mit anderen Tools beleuchteten, z.B. Geovisualisierungen von publizierten Daten via Worldmap⁹ oder statistische Analysen sowie optische Aufbereitung entsprechender Modelle mit TwoRavens.¹⁰ Aber auch empathische Plädoyers für Open Science und Data Sharing wurden vorgetragen. Die Texas Digital Library berichtete, für deutsche Ohren zunächst etwas befremdlich, dass ihr Datenrepositorium sowie zahlreiche andere digitale Dienste in der Amazon Cloud gehostet werden. Das Odum Institute skizzierte Aufwand und nutzen für eine DSA-Zertifizierung ihrer Dataverse Instanz.¹¹ Aus Syracuse gab es einen Bericht über das dort im Aufbau befindliche Qualitative Data Repository.¹²

Der erste Konferenztage wurde abgeschlossen bei einem Empfang im IQSS, in dessen Rahmen auch eine Reihe von Postern präsentiert wurden, u.a. eines auf dem unsere Heidelberger Forschungsdatenservices und unser Einsatz von Dataverse skizziert werden.¹³

Der zweite Konferenztage wurde wieder von einer Keynote Lecture eröffnet. Diese lieferte Patrica Cruse, Executive Director von DataCite. Nach einem allgemeinen Überblick über die DataCite-Services und –Aktivitäten und einer Vorstellung des EU-geförderten Projects THOR¹⁴ schlug Cruse den Bogen von persistenten Identifiern für Forschungsdaten und Publikationen (DOIs) zur Identifiern für die eindeutige Autorenidentifikation (ORCID) und die Ihres Erachtens bestehende Leerstelle im Hinblick auf eindeutige Identifier für Institutionen. Dies sei, so Cruse, ein künftiges Aktivitätsfeld für DataCite.

⁷ <https://www.force11.org/group/fairgroup/fairprinciples>

⁸ <https://rd-alliance.org/>

⁹ <http://guides.dataverse.org/en/latest/user/data-exploration/worldmap.html>

¹⁰ <http://guides.dataverse.org/en/latest/user/data-exploration/tworavens.html>

¹¹ <http://www.datasealofapproval.org/en/>

¹² <https://qdr.syr.edu/>

¹³ <http://doi.org/10.11588/heidok.00021660>

¹⁴ <https://project-thor.eu/>

Im Anschluss gaben die Dataverse-Entwickler einen Einblick in die aktuelle Fortentwicklung des Systems und geplante Weiterentwicklungen der Software.¹⁵ Aus Heidelberger Perspektive erfreulich war dabei, dass u.a. auch unsere spezifische Anforderungen auf der Agenda stehen, so soll z.B. die Möglichkeit der DOI-Registrierung via da|ra von den Dataverse-Entwicklern in Kooperation mit GESIS angegangen werden.

Unter dem Motto „Dataverse Repositories around the World“ präsentierten anschließend internationale Repositorienbetreiber ihren Einsatz von Dataverse. Von Kanada über die USA, China und Brasilien ging die Reise nach Deutschland und schließlich nach Norwegen. Mein Vortrag in dieser Session „Dataverse in Germany: Research Data Management at Heidelberg University“ kann online eingesehen werden.¹⁶

Im Anschluss wurden die oben bereits erwähnten, mit Dataverse kombinierbaren Tools noch einmal genauer beleuchtet und zudem die kollaborative Schreibumgebung Authorea vorgestellt.¹⁷ Seinen Abschluss fand der Tag dann in drei Breakout Session zu Fragen der Data Provenance, operativen Fragen des Betriebs eines Datenrepositoriums und einer Train-the-Trainer-Session, mit dem Ziel die Nutzung eines Datenrepositoriums effizient und überzeugend an Wissenschaftler vermitteln zu können.

Der dritte Konferenztag schließlich bestand aus zwei getrennten Veranstaltungen: Einmal ein Arbeitstreffen der aktiven Dataverse-Entwickler und zum anderen einem Workshop unter dem Titel „Managing Privacy in Research Data Repositories“, der in Kooperation mit den Fachbereichen für Informatik und Rechtswissenschaft veranstaltet wurde. Hier wurde sowohl ein allgemeiner Überblick über die Schwierigkeiten und Problemfelder beim Umgang mit personenbezogenen Daten gegeben, insbesondere natürlich im Hinblick auf das Teilen und Publizieren dieser Daten, als auch technische Systeme zur Unterstützung mit diesen Herausforderungen vorgestellt. Dabei ging es um Klassifizierungssysteme für die Vertraulichkeit von Daten und eine damit einhergehende Implementierung eines Rechtemanagementsystems,¹⁸ „Robot Lawyer“ genannte Systeme, die helfen sollen, automatisiert Nutzungsvereinbarungen für bestimmte Datenbestände, abgestimmt sowohl auf Nutzeranforderungen als auch auf relevante rechtliche Grundlagen, zu generieren, Tools für die Herstellung sog. Differential Privacy bei statistischen Daten und einiges mehr. In

¹⁵ Eine Übersicht bekommt man ggf. auf folgenden Seiten <http://dataverse.org/releases-roadmap>, <https://github.com/IQSS/dataverse/issues>, <https://groups.google.com/forum/#!forum/dataverse-dev>, <https://groups.google.com/forum/#!forum/dataverse-community>

¹⁶ <http://doi.org/10.11588/heidok.00021660>

¹⁷ <https://www.authorea.com/>

¹⁸ <http://datatags.org/>

einer abschließenden Podiumsdiskussion wurde das Thema noch einmal zusammengefasst und zudem der Fokus von der amerikanischen auch auf die europäische Rechtsordnungen ausgedehnt.

Zusammenfassend lässt sich festhalten, dass die Teilnahme am Dataverse Community Meeting in verschiedenen Hinsichten eine lohnend und bereichernde Erfahrung war. Zunächst sind natürlich die vielen persönlichen Kontakte zu nennen, sei es mit dem Dataverse-Team, mit dem schon im Vorfeld enger Austausch per Mail und in den Entwicklerforen bestand, oder sei es mit den anderen Konferenzteilnehmer aus unterschiedlichen Ländern und mit unterschiedlichen professionellen Hintergründen. Hinzu kommen die unvergesslichen Eindrücke, die ein Besuch in Boston und Cambridge sowie die Teilnahme und der Vortrag bei einer Veranstaltung der Harvard University hinterlassen. Aber auch auf der konkreten Arbeitsebene gibt es viele Aspekte, die sich direkt in die Arbeit an den Heidelberger Forschungsdatenservices integrieren lassen. Diese resultieren sowohl aus dem Austausch mit amerikanischen Wissenschaftlern, bei denen das geplante Management von Forschungsdaten durch die Vielzahl von diesbezüglichen Anforderungen amerikanischer Drittmittelgeber, bereits fest etabliert ist, als auch aus spezifischen Anregungen hinsichtlich des Einsatzes von Dataverse im Heidelberger Campus durch Gespräche mit anderen Repositorienbetreibern. Hinzu kommt der persönliche Austausch mit dem Entwicklungsteam konkret zu einzelnen Softwarefeatures oder Anwendungsfragen.

Aber damit nicht genug, im Anschluss an die Konferenz hatte ich am nächsten Tag zudem noch die großartige Möglichkeit, die Widener Library am Harvard Yard zu besuchen und dort nicht nur eine Gutenberg-Bibel zu bewundern, sondern auch Peter Suber, den Leiter des Harvard Office for Scholarly Communication¹⁹ und einen der zentralen Protagonisten der internationalen Open-Access-Bewegung, zu treffen und mich mit ihm intensiv über die Entwicklung des Open Access sowohl in den USA als auch in Deutschland auszutauschen. Auch dies mit wertvollen Inspirationen für die weitere Arbeit vor Ort in Heidelberg.

Ich danke BI International herzlich für die großzügige Unterstützung meiner Teilnahme an dieser Veranstaltung durch die Gewährung eines Stipendiums im Rahmen des Förderprogramms zur Kongressteilnahmen.

¹⁹ <https://osc.hul.harvard.edu/>