

Von Forschungsdaten und deren Aufbewahrung

Die Open Repositories Konferenz 2022 in Denver

Stipendiatenbericht von Johanna Staudinger

Nach zwei Jahren virtueller Veranstaltungen fand in diesem Jahr die Open Repositories Konferenz¹ vom 06.06. bis 09.06.2022 an der University of Colorado in Denver statt. Seit 2006 wird die internationale Konferenz jährlich veranstaltet, um Bibliothekar*innen, Manager*innen von Repositorien sowie Entwickler*innen einen Einblick in weltweit verwendete Techniken und Technologien im Bereich der Repositorien zu bieten. Die diesjährige Konferenz stand unter dem Motto: „Building trust together - Integrating, Collaborating and Sharing“.

Im Rahmen meiner Masterarbeit habe ich mich mit der Anreicherung von bibliographischen Metadaten in das Forschungsinformationssystem (FIS) der Universität Bamberg beschäftigt. Seit meinem Studienabschluss bin ich nun als Entwicklerin an der Universitätsbibliothek Bamberg für ein Upgrade des FIS verantwortlich, für das in naher Zukunft eine Integration des in der Masterarbeit entstandenen Prototypen durchgeführt werden soll. Ab dem kommenden Herbst werde ich zudem das Referendariat zur wissenschaftlichen Bibliothekarin zunächst an der Universitätsbibliothek in Bamberg und im Anschluss an der BSB München beginnen.

Zusammen mit meinem Vorgesetzten Steffen Illig und meinem Kollegen Florian Gantner habe ich ein Poster zu den technischen Hintergründen und der zukünftigen Weiterentwicklung des Prototypen entworfen. Dieses Poster durfte ich auf der Open Repositories Konferenz vorstellen und mich mit interessierten Forschenden über die Funktionsweise des Prototyps austauschen. Vor allem bei der Schwierigkeit, die Forschenden bei der Anreicherung ihrer Forschungsdaten in Repositorien zu unterstützen, soll das Poster einen kleinen Beitrag zur Besserung leisten.

Die Schwerpunkte der Open Repositories Konferenz lagen in diesem Jahr vor allem auf den Bereichen Open Access, der Erhaltung von Forschungsdaten und neuen Technologien der Datenverwaltung wie beispielsweise unter Einsatz von Künstlicher Intelligenz. In verschiedensten Präsentationen, Diskussionen sowie Workshops wurde den etwa 300 Teilnehmenden ein breites Angebot an Möglichkeiten zur Information und Vernetzung geboten.

Die Konferenz begann mit einem Begrüßungspanel, in dem über das Motto der Veranstaltung „Building trust together“ mit allen Anwesenden diskutiert wurde. Die Begriffe „trust“ sowie „openness“ wurden zusammen mit dem Publikum genauer definiert und es wurden Erfahrungen aus dem Arbeitsalltag ausgetauscht. Eine gelungene Zusammenarbeit zwischen Forschenden und Bibliotheksvertreter*innen werde demnach durch das gegenseitige Vertrauen in die Fähigkeiten des Gegenübers wie auch einer offenen Kommunikation ermöglicht.

Im Anschluss konnten die Forschenden wählen, zu welchem Forschungsbereichen und in welcher Vortragsart sie sich die nachfolgenden Präsentationen anhören wollten.

¹ <https://www.openrepositories.org>



Die Open Repositories Konferenz 2022 war gut besucht. © Julian Lehrhuber CC BY-SA 4.0
<https://creativecommons.org/licenses/by-sa/4.0/deed.de>

Ich habe mich im Rahmen dessen für den Developer Track unter dem Oberthema „Repository Systems Updates“ entschieden, um mich über Neuerungen von Verwaltungssystemen für Forschungsdaten wie das FIS zu informieren. Bei den Vorträgen der Softwareentwickler 4Science und LYRISIS wurden die neuesten Veröffentlichungen für das System DSpace² präsentiert, die auch an der Universitätsbibliothek Bamberg Anwendung finden. Zudem referierten Forschende der University of Virginia, der British Library sowie der Ubiquity Press zur Hyku³ Open Source Repositorien Plattform als Alternative zu DSpace.

Anschließend hörte ich mir die 24x7-Reihe an, in der die Referenten 7 Minuten Zeit haben, ihre maximal 24 Folien dem Publikum zu präsentieren. Vor allem der Vortrag „Author Names and Identity Management in Institutional Repositories“ weckte dabei mein Interesse, da er sich ähnlich meiner Masterarbeit mit der Schwierigkeit befasste, Forschende anhand ihres Namens zu identifizieren. Der Referent Ellis D. Ging nannte dabei die Transliteration von Namen oder die globalen Unterschiede in der Reihenfolge von Namensbestandteilen als eine große Ursache des „Name Disambiguation Problems“. Mögliche Lösungsvorschläge dafür wären die Nutzung von Namensfeldern, deren Typ nicht spezifiziert ist, oder die einheitliche Großschreibung von Nachnamen. Daneben wurde im Vortrag „RightLib - electronic access to publications under copyright“ über Zugangsvoraussetzungen auf urheberrechtlich geschützte Werke referiert. So werden im tschechischen Kramerius-System⁴ die Publikationen mit verschiedenen Farben markiert als: frei zugänglich (blau), urheberrechtlich geschützt (rot) und zugänglich durch Authentifizierung (grün).

Nach der Mittagspause nahm ich an Präsentationen unter den Oberthemen „Integrating tools“ und „Infrastructures“ teil. Forschende der Oklahoma State University stellten in diesem Rahmen ihre Version eines DSpace Repositoriums zur Nutzung als Forschungsinformationssystem vor. Analog zum Bamberger FIS sollen in diesem System Publikationen von Forschenden eingereicht und verwaltet werden können. Sollte ein Duplikat einer Publikation im Repositorium aufgefunden werden, so werden die beiden Publikationseinträge zusammengeführt und das Duplikat wird als Versionierung

² <https://dspace.lyrasis.org>

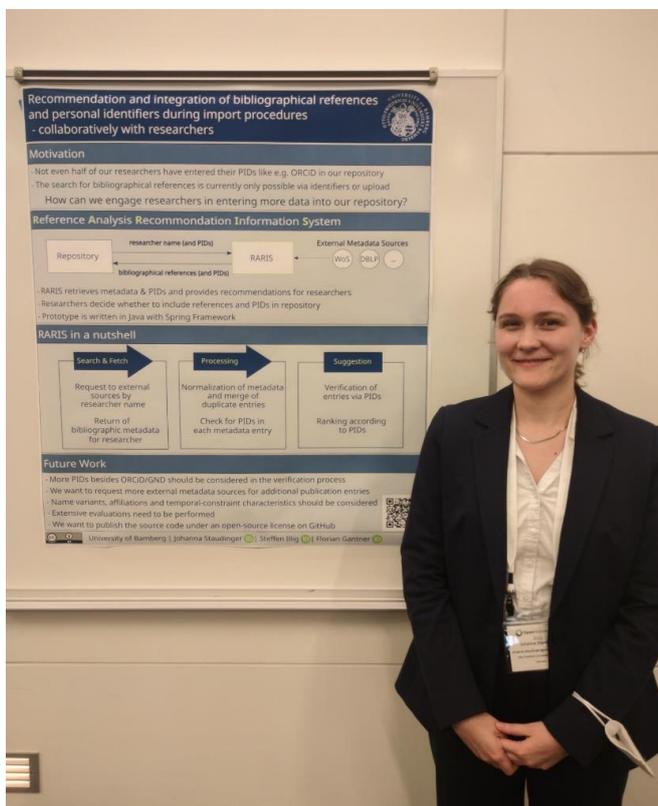
³ <https://hyku.samvera.org>

⁴ <https://kramerius.nkp.cz>

im älteren Publikationseintrag angelegt. Anschließend wurde das Aggregator-System CORE⁵ vorgestellt, das Open Access Publikationen aus Repositorien und Zeitschriften sammelt und zur Verfügung stellt. Zudem wurde im darauffolgenden Vortrag über Identifikatoren für Forschungsdaten referiert, die anders als der Digital Object Identifier (DOI) die Daten einmalig und persistent definieren sollen. So können Open Archives Initiative (OAI)-Identifikatoren genutzt werden, die als Personal Object Identifiers (PIDs) jedem Objekt zugeordnet und im OAI-PMH des Repositoriums hinterlegt werden. Durch einen OAI-Resolver wie CORE sollen die Objekte schließlich global identifiziert werden können, wobei immer auf das Objekt im Stamm-Repositorium verwiesen werden soll.

Anschließend fand das „Repository Rodeo“-Panel statt, in dem sich Vertreter*innen aus der Repositorien-Community den Fragen und Antworten des Publikums stellten. In diesem Jahr wurden folgende Repositorien-Systeme vertreten: Dataverse, DSpace, EPrints, Fedora, Haplo, Invenio, Islandora und Samvera.

Den Abschluss des ersten Kongresstages machte dann die Postersession, in der die Forschenden ihre Poster dem Publikum innerhalb von 60 Sekunden vorstellen durften. Auch ich stellte als Vertreterin aus Bamberg unser Poster mit dem Titel „Recommendation and integration of bibliographical references and personal identifiers during import procedures - collaboratively with researchers“ vor. Im Anschluss konnte ich mich dann mit Interessenten über den Vortrag austauschen und einige Anregungen und Vorschläge für die Weiterentwicklung des Prototypen erhalten. Zudem erhielt ich im Gespräch mit den anderen Vortragenden spannende Einblicke in die Forschungsdatenverwaltung mittels des InvenioRDM-Repositoriums oder die Nutzung von Künstlicher Intelligenz beim Zugang zu elektronischen Publikationen.



Ich durfte ein Poster zum Prototyp meiner Masterarbeit auf der Konferenz vorstellen. © Julian Lehrhuber CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0/deed.de>

⁵ <https://core.ac.uk>

Der zweite Tag der Open Repositories Konferenz begann für mich mit dem Developer Track zum Oberthema „Acknowledged: All about notifications“. Hierbei wurde das „Notify“⁶ Protokoll vorgestellt, das von der Confederation of Open Access Repositories entwickelt wurde. Es soll die Verknüpfung von Forschungsdaten mit externen Ressourcen wie Peer-Review-Diensten ermöglichen und dabei Benachrichtigungen über die verknüpften Daten senden. Im Vortrag wurde dabei vor allem auf die technische Umsetzung wie die Nutzung bestimmter Kommunikationsprotokolle eingegangen. Im anschließenden Vortrag wurde auf die Implementation des Protokolls in DSpace eingegangen, das beispielsweise eine Verlinkung von DSpace-Items mit der Dataverse⁷ Plattform ermöglicht. Auf beiden Seiten der Verlinkung werden Benachrichtigungen erzeugt, die den Nutzer über den Status seiner eingereichten Daten informieren.

Nach einer kurzen Pause beschäftigten sich die darauf folgenden 24x7-Vorträge mit dem Oberthema „Building trust in data repositories“. Besonderer Fokus wurde dabei auf die benötigten Eigenschaften eines Repositoriums gelegt, die gegeben sein müssen, damit Forschende ihre Forschungsdaten mit einem guten Gefühl darin hinterlegen. Zu nennen sind hier die TRUST Prinzipien: transparency, responsibility, user focus, sustainability und technology. Ein Forscher der Max Planck Digital Library erklärte in diesem Zusammenhang das Vorgehen bei der Wahl eines neuen Betriebssystems für ihr Datenrepositorium EDMOND⁸. Vor allem die Interaktion mit den Nutzer*innen spielte dabei eine große Rolle, da neben Feedback-Runden auch User Tests sowie virtuelle Tutorials genutzt wurden, um die Meinung der Forschenden zu evaluieren.

Nach der Mittagspause folgten weitere Präsentationen zum Thema „Adapting and Integrating Repository Systems“. Von Interesse für mich waren dabei vor allem die Vorträge, die sich mit der Personalisierung und Weiterentwicklung des DSpace-Systems beschäftigten. So entwickeln Forschende der University of Oklahoma mit SAF-GUI eine Anwendung, die Daten aus dem CSV-Schema in das Simple Archive Format (SAF) konvertiert und validiert, bevor diese in ein DSpace-Repositorium importiert werden. So werden Batch-Einspielungen der Daten für Administratoren und Bibliothekare erleichtert, was möglicherweise auch im FIS der Universität Bamberg Anwendung finden könnte. Zudem wurde die Personalisierung und Zusammenführung zweier DSpace-Instanzen am Georgia Institute of Technology vorgestellt, und der Softwareentwickler 4Science demonstrierte Neuerungen im aktuellen Release der Software DSpace-CRIS.

Den Abschluss des zweiten Konferenztags bildeten die Präsentationen zum Oberthema „Digital Preservation“, die sich mit der langfristigen Erhaltung von Forschungsdaten befassten. Demnach gibt es eine hohe Nachfrage an gut dokumentierten, plattformunabhängigen und vertrauenswürdigen AIPs (archival information packages), die in globalen Netzwerken hinterlegt sein sollen. Allerdings sind auch hier die Wahl eines geeigneten Hosts und einer passenden Plattform sowie der Aufbau von Vertrauen zwischen Nutzer*innen, Stakeholdern und Anbietenden der Systeme von hoher Relevanz.

Den Abschluss der Open Repositories Konferenz 2022 bildete die Closing Keynote von Kyle K. Courtney, der als Jurist über rechtliche Gegebenheiten im Bereich der Repositorien wie Urheberrecht, Lizenzierung oder Open Access sprach.

⁶ <https://www.coar-repositories.org/notify/>

⁷ <https://dataverse.org>

⁸ <https://edmond.mpdl.mpg.de>



An der University of Colorado in Denver fand die Konferenz statt. © Julian Lehrhuber CC BY-SA 4.0
<https://creativecommons.org/licenses/by-sa/4.0/deed.de>

Zusammenfassend war die Teilnahme an der Open Repositories Konferenz eine tolle Erfahrung, bei der ich vielen spannenden Vorträge folgen, mich mit anderen Forschenden austauschen und allgemein mein bescheidenes Wissen im Bereich der Repositorien erweitern konnte. Für meine zukünftige Tätigkeit als wissenschaftliche Bibliothekarin im Fachbereich Informatik werden mir die erlernten Kenntnisse definitiv von Nutzen sein, um die Verwaltung von Forschungsdaten gut betreuen zu können und den Bibliotheken im aktuellen Technologiezeitalter zur Seite zu stehen.

Ich bedanke mich sehr herzlich bei BI-International und dem Goethe Institut e.V., die mir mit ihrem BII-Stipendium die Teilnahme an der Open Repositories Konferenz 2022 ermöglicht und mir damit zu einem einmaligen Erlebnis verholfen haben.

Kontakt:

Johanna Staudinger

johanna.staudinger@uni-bamberg.de