

■ IST PUBREP UND IST DATAREP: DIE INSTITUTIONELLEN REPOSITORIEN AM IST AUSTRIA

von Barbara Petritsch, Jana Porsche

Zusammenfassung: Am IST Austria wurden 2013 ein Publikationsrepositorium und 2015, nach einer sorgfältigen Vorbereitungsphase, ein Datenrepositorium in Betrieb genommen – beide auf der Open-Source-Software EPrints basierend. Im ersten Teil unseres Erfahrungsberichts werden wir aus gegebenem Anlass unsere Erfahrungen mit Open-Source-Software allgemein und EPrints im Besonderen, bezüglich technischer Aspekte als auch im Hinblick ihrer Besonderheit – der User Communities, rekapitulieren. Der zweite Teil ist als Plädoyer für die Einbeziehung von EndbenutzerInnen in die Implementierungs-, Anpassungs- und Evaluierungsprozesse konzipiert.

Schlüsselwörter: Institute of Science and Technology Austria; Publikationsrepository; Datenrepository; EPrints; LibreCat

IST PUBREP UND IST DATAREP: THE INSTITUTIONAL REPOSITORIES AT IST AUSTRIA

Abstract: In 2013, a publication repository was implemented at IST Austria and 2015 after a thorough preparation phase a data repository was implemented – both based on the Open Source Software EPrints. In this text, designed as field report, we will reflect on our experiences with Open Source Software in general and specifically with EPrints regarding technical aspects but also regarding their characteristics of the user community. The second part is a pleading for including the end users in the process of implementation, adaption and evaluation.

Keywords: Institute of Science and Technology Austria; Publication repository; Data repository; EPrints; LibreCat

DOI: <http://doi.org/10.31263/voebm.v71i1.1993>



Dieses Werk ist lizenziert unter einer

[Creative-Commons-Lizenz Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/)

Inhalt

1. *Einleitung*
2. *IST PubRep – Automatisierter Workflow*
3. *IST DataRep*
4. *Software: Von EPrints und BibApp zu LibreCat*
5. *Conclusio*

1. Einleitung

Das IST Austria hat sich als junges Forschungsinstitut mit digital ausgerichteten Bibliotheksservices bereits kurz nach dessen Gründung an die Implementierung eines institutionellen Publikationsrepository und im Anschluss daran an die Implementierung eines Datenrepository gewagt. Bereits in der Anfangs-/Planungsphase waren die Projekte von zwei wesentlichen Ansprüchen geprägt: zum Einen standen unbestritten die BenutzerInnen im Fokus, zum anderen wurde nach technischen Hilfsmitteln gesucht, die leicht zu implementieren, pflegen, bedienen und nebenbei auch noch kostengünstig sein sollten. Nach mehreren Jahren Erfahrung ist eine Rekapitulation dieser Ansprüche im Lichte ihrer Umsetzung möglich.

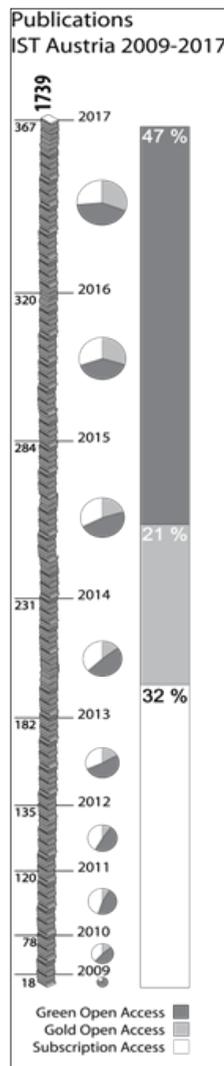
2. IST PubRep¹ – Automatisierter Workflow²

Das institutionelle Publikationsrepository IST PubRep wurde so konzipiert, dass die WissenschaftlerInnen im Normalfall lediglich den Upload der Volltexte über die institutionelle Publikationsdatenbank vornehmen müssen und die zugehörigen Metadaten im Zuge eines SWORD-Imports³ in das Repository eingespielt werden. Dadurch wird den AutorInnen eine mehrmalige händische Eingabe von Metadaten erspart und die Fehleranfälligkeit in der Dateneingabe minimiert. Die importierten Publikationen müssen allerdings im Review überarbeitet und vervollständigt werden, da die automatisierte Zuordnung der Felder nicht einwandfrei gelingt. Aufgrund der Tatsache, dass diese Option kaum proaktiv genutzt wurde, entwickelten wir 2016 eine Strategie um den Green Open Access-Anteil am Institut und die Benutzung des Publikationsrepositorys zu steigern: Einmal jährlich erhalten die AutorInnen ein persönliches Email mit einer Liste ihrer Publikationen, die zur Selbstarchivierung in Frage kommen, mit der zusätzlichen Information welche Version (Preprint, Accepted Version, Publisher Version) dafür zulässig ist. So konnten wir 2016 den Green Open

Access-Anteil affilierter Publikationen stark erhöhen. Etwa 1/3 der angefragten Publikationen wurden rückwirkend selbstarchiviert.⁴ Weiters legt die Bibliothek alle Publikationen unter freier Lizenz standardmäßig im institutionellen Repository ab. Mit einem momentanen Open Access Anteil von 68% (davon 47% Green Open Access und 21% Gold Open Access; siehe Grafik) aller IST Publikationen verstehen wir diese Maßnahmen als erfolgreich und das Repository hat sich zu einem bedeutenden Volltextnachweis institutioneller Publikationstätigkeit entwickelt. Das Auffinden der Volltexte in IST PubRep funktioniert über herkömmliche Suchmaschinen sehr gut und der Bestand ist in den fachlichen Suchmaschinen BASE⁵ und OALster⁶ nachgewiesen.

Als kleine Institution haben wir den Vorteil, umfassende Services anbieten zu können. Eines davon ist die jährliche Überprüfung zugangsbeschränkter Artikel von IST Austria-Angehörenden auf ihre anwendbaren Selbstarchivierungsrichtlinien hin, um anschließend die AutorInnen darüber zu informieren, welche Versionen ihrer Publikationen unter welchen Umständen selbstarchiviert werden können. Diesen Aufwand „intellektuell“ bei größeren Institutionen zu betreiben ist quasi ein Ding der Unmöglichkeit.⁷ Eine denkbare Variante davon wäre, Publisher bzw. Journals herauszufiltern, in denen die meisten Publikationen der eigenen Institution veröffentlicht werden und die Information über deren Selbstarchivierungsrichtlinien den betreffenden AutorInnen zukommen zu lassen.

Mit derartigen Maßnahmen ist es unserer Erfahrung nach möglich jene AutorInnen für die Sache zu gewinnen, die eine grundsätzlich positive Einstellung zu Open Access haben, „solange es nicht mit zusätzlichen Kosten oder Aufwand verbunden ist“. Open Access prinzipiell ablehnende AutorInnen werden damit wohl kaum erreicht werden können und sollten auch nicht im Fokus für erste Maßnahmen stehen. An ihnen kann man sich nach den ersten Erfolgen noch immer die Zähne ausbeißen.



3. IST DataRep⁸

Die Auswahl und der Implementierungsprozess des Datenrepositorys ist sehr gut dokumentiert und in Projektberichten nachlesbar.⁹ Wichtig war hier vor allem die Übereinstimmung der Metadatenschemas von Repository und der DOI-Vergabestelle DataCite.

Bereits zu Beginn des Projekts wurde großer Wert darauf gelegt, die künftigen NutzerInnen des Services miteinzubeziehen. So wurde vor der Sondierung von geeigneter Software eine Bestandserhebung in Form einer Umfrage bezüglich der Produktion, dem Handling und der Ablage von Forschungsdaten mit allen derzeitigen Forschungsgruppen durchgeführt. Anschließend wurde daraus ein Anforderungskatalog, sortiert nach Priorität und Durchführbarkeit, für infrage kommende Applikationen erstellt. In der nächsten Instanz wurden zwei Software-Lösungen getestet und als Ergebnis EPrints implementiert. Die Anpassungen an institutionelle Bedürfnisse wurden abermals im Anschluss an protokollierte BenutzerInnen-Tests durchgeführt, mit dem Ziel, die Benutzerfreundlichkeit der Oberfläche, Klarheit des Ablaufs und die Verständlichkeit der Hilfetexte sicherzustellen. Anhand der Ergebnisse der Testuploads wurden Änderungen und Anpassungen vom Bibliotheksteam und vor allem von der internen IT-Abteilung vorgenommen.

Für EPrints wird über EPrints Services Support angeboten. Da dieser jedoch nur für die Standardimplementierung gilt und für uns daher nicht in Frage kam, waren wir auf die sehr knappen internen IT-Ressourcen angewiesen. Bis dato sind wir im Grunde, abgesehen vom Handling großer Datenmengen (ab 1 GB), mit der Lösung zufrieden. Dennoch wird EPrints in Zukunft von einer anderen Software abgelöst werden. Die Gründe dazu werden an späterer Stelle näher ausgeführt.

Auch für das Datenrepository war es nötig, Strategien zu entwickeln um die WissenschaftlerInnen mit dem Service vertraut zu machen und sie so zur Nutzung zu motivieren. Das bedeutete anfangs, Open-Access-affine Forschungsgruppen am Institut direkt anzusprechen und sie zur Datenarchivierung zu motivieren. Beständigkeit war in diesem Fall die beste Strategie und ermöglichte uns weitere Feinanpassungen im Workflow vorzunehmen und nebenbei die WissenschaftlerInnen mit dem Service bekannt zu machen. Nach etwa einem Jahr war das Service zumindest soweit bekannt, dass wir proaktiv von den WissenschaftlerInnen angefragt wurden. Diese Entwicklung spiegelt sich auch in der kontinuierlichen Zunahme an Deposits über die Jahre wieder. So hatten wir 2015 ein Deposit, 2016 acht Deposits, 2017 dreizehn Deposits und mit 10. April 2018 haben wir bereits dreizehn Deposits für 2018 zu verzeichnen.¹⁰

4. Software: Von EPrints und BibApp zu LibreCat

Einige der wichtigsten Kriterien für eine Repository-Software waren, dass es eine kostengünstige, leicht implementierbare und leicht bedienbare Software sein sollte und somit fiel die Entscheidung auf die Open Source-Software EPrints. Für das Datenrepository wurde eine spezifische Recherche durchgeführt und weitere Applikationen begutachtet. Auch hier fiel die Entscheidung letztlich auf EPrints. Nicht nur weil dieses System bereits für das Publikationsrepository in Verwendung war, sondern auch weil für die Applikation ein Plugin zur Beschreibung von Daten zur Verfügung stand und unserem Kriterienkatalog dadurch am nächsten kam. Als das Datenrepository nach einer langen Anpassungsphase von zwei Jahren in Betrieb genommen wurde, zeichnete sich bereits ab, dass es große Unzufriedenheiten in der EPrints User Community gab. Somit kam die Weiterentwicklung fast zum Erliegen und das „neue“ Geschäftsmodell führte zu heftigen Diskussionen innerhalb der Community¹¹. Auch die instituti-
onsinterne IT-Abteilung bemängelte EPrints im Hinblick auf den veralteten Source Code und war ressourcentechnisch nicht erfreut über die Vielfalt an verschiedenen Bibliotheksapplikationen.

Mit Beginn dieses Jahres begann der Umstieg auf LibreCat. In diesem Prozess werden drei separat laufende Bibliotheksservices – die Publikationsdatenbank, das Publikations- und das Datenrepository – zusammengeführt. Der Anlass dazu war weniger die steigende Unzufriedenheit mit EPrints, als vielmehr die Notwendigkeit, die auf BibApp laufende Publikationsdatenbank mit einem neuen System abzulösen. BibApp ist eine Open Source-Software, deren Weiterentwicklung seit ca 2014 stagnierte und mittlerweile aufgrund der veralteten Rubyversion weder updatebar noch sicherheitstechnisch tragbar ist. Im Zuge der Besprechungen mit der IT-Abteilung über eine neue Publikationsdatenbank wurde der Wunsch geäußert, die verschiedenen laufenden Bibliothekssysteme auf ein Mindestmaß zu reduzieren, um die Instandhaltung und Aktualisierung effizienter gestalten zu können.

LibreCat¹² ist eine von der Universität Lund entwickelte Open Source Software, basierend auf Catmandu. Seit 2010 existiert eine Kollaboration zwischen den Universitäten Lund, Ghent und Bielefeld bei der Entwicklung der Software. Momentan sind etwa 15 Personen daran beteiligt und es herrscht ein offenes Klima zur Teilnahme am Projekt. Von unserer IT-Abteilung kam ein sehr positives Feedback bezüglich des Source Codes. Mit Beginn dieses Jahres wurde das Testsystem am Institut installiert. Der Anpassungsprozess und die Datenmigration gehen gut voran,

somit rechnen wir mit einer Inbetriebnahme des neuen System mit Ende des Jahres.

Diese Ereignisse bewegten uns ganz offensichtlich keineswegs dazu, auf lizenzierte Applikationen zu wechseln – denn auch hier besteht die Gefahr stagnierender Entwicklung, Weiterentwicklungen gegenläufig der institutionellen Bedürfnisse und/oder der Einstellung eines Produkts. Der aus unserer Sicht unbestreitbare Vorteil von Open Source-Lösungen bleibt dennoch gegeben: Der Source Code ist grundsätzlich offen zugänglich und lässt zumindest theoretisch zu, die Software institutions- oder community-intern selbstständig weiterzuentwickeln. Zumindest sind aber die Daten in einem offenen Format vorhanden und im besten Fall leicht zu exportieren. So existiert keine Beschränkung auf eine spezifische Applikation. Hat man das Glück, Teil einer großen Community zu sein, kann man zwischen mehreren Supportfirmen auswählen und ist nicht auf einen Sole-Supplier angewiesen.

5. Conclusio

Zum Abschluss möchten wir noch einzelne Aspekte festhalten, die aus unserer Sicht maßgeblich an einem erfolgreich geführten Repository beteiligt waren und sind:

Der persönliche Kontakt zu WissenschaftlerInnen, in diesem Fall zu den sogenannten Open Access-Champions, hat einen wichtigen Effekt. Mit ihrer Hilfe können Maßnahmen benutzerInnenfreundlich gestaltet und Workflows an die BenutzerInnenbedürfnisse angepasst werden. Außerdem ist man für die Open Access-SkeptikerInnen besser gewappnet. Eine kontinuierliche Dokumentation in Form von Projektberichten und Präsentationen in Usergruppen und Konferenzen ist von Vorteil, speziell wenn es während eines Projekts zu Personalwechsel bzw. Karenzvertretungen kommt. Diese Berichte im Repository oder über Fachzeitschriften zu veröffentlichen steigert ihren Wert, da sie von der interessierten Fachcommunity rezipiert werden können. Der informelle sowie institutionalisierte Erfahrungsaustausch mit der bibliothekarischen Fachcommunity von großer Bedeutung für Bereiche die noch erkundet und etabliert werden müssen. Insofern sind Konferenzen¹³ (und deren Kaffeepausen), User-Treffen, Netzwerke¹⁴ und Arbeitsgruppen die oft ergiebigsten Informationsquellen.

Auch intern sind Kooperationen mit anderen Abteilungen eine wichtige Quelle, um Feedback über Wirksamkeit und Kompatibilität der Maßnahmen einschätzen zu können und Kollaborationen zu entwickeln. Im Spezial-

len eine gute Beziehung zur IT-Abteilung ist im Falle der Repositories grundlegend für den Erfolg. Kurz gesagt, besteht die Arbeit mit/an Repositories neben den technischen Aspekten zu guten Teilen aus zwischenmenschlicher Beziehungsarbeit. Schließlich geht es ja unter anderem auch darum, eine Veränderung im Verhalten der beteiligten AkteurlInnen zu motivieren.

Mag.^a Barbara Petritsch

ORCID: <http://orcid.org/0000-0003-2724-4614>

Wien Museum, Bibliothek (ehem. IST Austria, Bibliothek)

E-Mail: barbara.petritsch@wienmuseum.at

Mag.^a Jana Porsche

Institute of Science and Technology Austria, Bibliothek

E-Mail: jana.porsche@ist.ac.at

- 1 <http://ist.ac.at/library/library-services/publication-repository/>
- 2 Siehe hierzu auch Tendenzen im internationalen Vergleich in den Ergebnissen einer australischen Studie aus 2017 zu Repositories: „Automated data flows are increasing (between Research Management systems and Repository systems). Self-deposit is decreasing in response to more automated workflows und ebenda: Challenges: Time poor researchers. Engagement with repositories by time poor researchers is an issue. If repository software is poor and doesn't allow easy integration, upload, bulk ingest, engagement or value, then libraries will continue to bear the burden of a lot of work without much support from key stakeholders (researchers and research offices).“ Simon Huggard, Kay Steel, Alexander Sussman (2017): Research Publications Repository Survey Report 2017, Australian National University, S. 4, 5+8. <http://www.caul.edu.au/content/upload/files/research/caul2017rprs-report.pdf>
- 3 Simple Web-Service Offering Repository Deposit: <http://www.swordapp.org>
- 4 Publikationen zwischen 2009 und 2015: von 278 Artikel wurden 91 selbstarchiviert, 2016: von 131 Publikationen wurden 50 selbstarchiviert, 2017: von 111 Publikationen wurden (bis einschließlich 10. April 2018) 16 selbstarchiviert.
- 5 Bielefeld Academic Search Engine: <https://www.base-search.net/>
- 6 The OAIster[®] database: <https://www.oclc.org/en/oaister.html>
- 7 Es bestehen vergleichbare Unternehmungen, wie zum Beispiel an der Universität Hannover, die dieses Service (Repository-Check-Up:

- <http://www.repo.uni-hannover.de/page/services>) anbietet. Hier müssen die AutorInnen das Service allerdings proaktiv einfordern. In einem Projekt des Kooperativen Bibliotheksverbundes Berlin-Brandenburg (KOBV) werden diesbezüglich automatisierte Methoden entwickelt und erprobt (<https://deepgreen.kobv.de/de/deepgreen/>).
- 8 <http://ist.ac.at/library/library-services/data-repository/>
 - 9 Petritsch, Barbara (2017): Implementing the institutional data repository IST DataRep. Project Report. IST Austria. <https://repository.ist.ac.at/724/>; Petritsch, Barbara (2017): Metadata for research data in practice. In: Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare. VÖB 70(2), S. 200–207. DOI: <https://doi.org/10.31263/voebm.v70i2.1678>; Porsche, Jana (2013): Technical requirements and features. Project Report. IST Austria. <https://repository.ist.ac.at/135/>; Porsche, Jana (2013): Initiatives and projects related to RD. Project Report. IST Austria. <https://repository.ist.ac.at/113/>; Porsche, Jana (2012): Actual state of research data @ ISTAustria. Project Report. IST Austria. <https://repository.ist.ac.at/103/>
 - 10 Siehe: <https://datarep.app.ist.ac.at/view/year/>
 - 11 Siehe hierzu die sog. Managed Availability von EPrints Updates. Diese Form der Lizenzierung stieß auf regelrechten Unmut, da sie als klares Statement gegen Open Source (GNPL) interpretiert wurde, auch wenn sich die Entwickler von EPrints vehement gegen diese Anschuldigung wehrten: <http://www.eprints.org/uk/index.php/eprints-3-4/managed-availability/> (zuletzt abgerufen am 14.3.2018)
 - 12 Mehr Information zu Libre Cat: <http://librecat.org/about.html>
 - 13 z.B. Die jährlich stattfindenden Konferenzen Open Repositories und die Open Access Tage.
 - 14 z.B. in Österreich das Repository ManagerInnen Netzwerk Repmannet (<https://ubifo.at/images/repmannet.pdf>).