

Chancen und Herausforderungen im Zusammenspiel von Forschungsdatenmanagement und Forschungsinformationssystemen

Dr. Reingis Hauck ^a, Dr. Sandra Broll ^a

^a Leibniz Universität Hannover, Dezernat Forschung und EU-Hochschulbüro, Technologietransfer, Welfengarten 1a, 30167 Hannover

Der vorliegende Artikel bezieht sich auf den gleichnamigen Vortrag auf den E-Science-Tagen 2017 in Heidelberg.

Abstract

Erfolgreiches Forschungsdatenmanagement (FDM) hängt im großen Maße von der Bereitstellung der entsprechenden IT-Infrastruktur ab, die Wissenschaftlerinnen und Wissenschaftler im Forschungsprozess angemessen unterstützt. Hierbei ist nicht nur auf technische Aspekte zu achten sondern auch auf wissenschaftsfreundliche Workflows, die darauf Rücksicht nehmen, dass eine Vielfalt von IT-Systemen im Alltag des Hochschullebens zu bedienen ist (Simons et al. 2016). Im Folgenden sollen die Chancen und Herausforderungen im Zusammenspiel von Volltextrepositorium, Datenrepositorium und Forschungsinformationssystem exemplarisch am Beispiel der Leibniz Universität Hannover (LUH) beleuchtet werden.

Die LUH führt zurzeit ein Forschungsinformationssystem (Pure) ein, das Informationen zu den Forschungsaktivitäten ihrer Mitglieder zusammenführt. Diese sogenannten Forschungsinformationen umfassen neben Metadaten zu (Text-)Publikationen und Projekten auch Metadaten zu publizierten Forschungsdatensätzen. Parallel zur Einführung des Forschungsinformationssystems wurde der Aufbau eines institutionellen Volltextrepositoriums umgesetzt. Im Sommer 2016 wurde der Aufbau eines institutionellen Forschungsdatenrepositoriums als Teil des institutionellen Konzepts zum Forschungsdatenmanagement an der LUH beschlossen. In diesem Zusammenhang wird die Frage der Schnittstelle zwischen Daten- und Volltextrepositorium sowie zum Forschungsinformationssystem im Hinblick auf wissenschaftsfreundliche Workflows aufgegriffen. Die Datenpublikation soll nicht als weiterer administrativer Overhead von Wissenschaftlerinnen und Wissenschaftlern verbucht werden müssen. Neben dem Primat der Eingabefreundlichkeit gilt, dass Forschungsinformationssysteme per definitionem das zentrale Berichtstool für das Forschungsmanagement darstellen. Auf institutioneller Ebene besteht das Interesse, das tatsächliche Verhalten hinsichtlich des Forschungsdatenmanagements mit Zahlen zu hinterlegen, u.a. zu Art und Umfang der Daten und Anzahl der Datenpublikationen. Eine ganzheitliche Systemarchitektur erfordert, sich entwickelnde Anforderungen mitzudenken. Der Blick nach Großbritannien kann hier weiterhelfen: Sowohl hinsichtlich der Umsetzung von

Datenmanagementkonzepten als auch in der Nutzung von Forschungsinformationssystemen besteht dort ein erheblicher Erfahrungsvorsprung, der zu nutzen gilt.

Schlagwörter: Forschungsdaten, Forschungsinformationssysteme, Workflow, Repositoryum

1 Ausgangssituation und Systemarchitektur

In 2011 startete an der Leibniz Universität Hannover das Projekt „Aufbau eines Forschungsinformationssystems und einer Dienstleistungsinfrastruktur zum Digitalen Publizieren“. Das Projekt sollte die Infrastrukturdefizite aufgreifen, die im Bereich von Forschungsinformationen und im Bereich des digitalen Publizierens deutlich geworden waren: Weder existierte eine Hochschulbibliographie noch ein institutionelles Volltextrepositorium.

Die Leibniz Universität Hannover hat sich für das Forschungsinformationssystem Pure der Firma Elsevier entschieden, das Informationen zu den Forschungsaktivitäten der Wissenschaftlerinnen und Wissenschaftler der Hochschule zusammenführt. Diese sogenannten Forschungsinformationen umfassen neben Metadaten zu (Text-)Publikationen und Projekten auch Metadaten zu publizierten Forschungsdatensätzen. Das Forschungsinformationssystem synchronisiert soweit wie möglich Daten aus bestehenden Systemen wie SAP HR (SAP Modul Personalwirtschaft) und dem SAP Folders Management (elektronische Drittmittelakte) als auch aus bestehenden Onlinedatenquellen wie z.B. dem GVK, Scopus oder Crossref¹.

Forschungsinformationen können durch Verknüpfungen miteinander und zu Personen in Bezug gebracht werden. Diese verknüpften Informationen sollen in einem Portal der interessierten Öffentlichkeit präsentiert und zugänglich gemacht werden. Das Portal soll maßgeblich zu einer besseren Außendarstellung der Universität führen. Darüber hinaus wird das Portal für Webcrawler zugänglich sein und so für die Auffindbarkeit der Forschungsinformationen im Netz sorgen. Vergleichbare Portale wurden bereits an anderen Hochschulen umgesetzt². Die Implementierung des Forschungsinformationssystems Pure wird an der Leibniz Universität Hannover voraussichtlich im Herbst 2017 beendet sein. Ein Volltextrepositorium (DSpace) ging Ende 2015 in den Produktivbetrieb³.

¹ GVK ist die Abkürzung für den Gemeinsamen Verbundkatalog des Gemeinsamen Bibliotheksverbunds (GBV)

² S. hierzu z.B. <https://pure.strath.ac.uk/portal/en/>

³ www.repo.uni-hannover.de

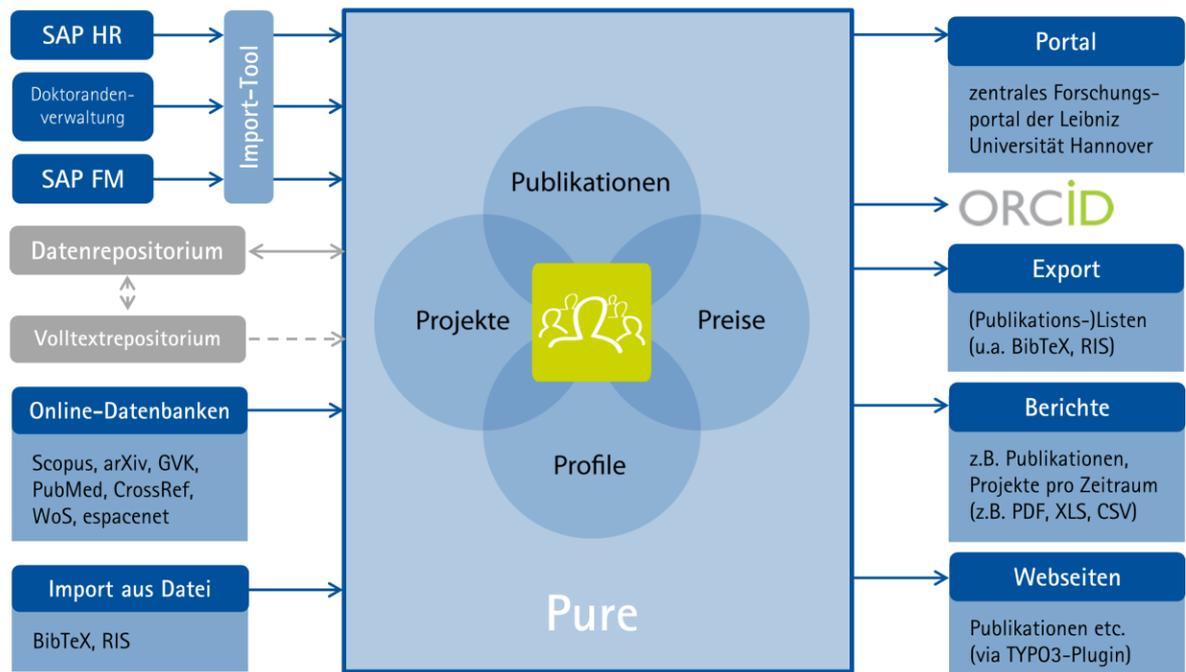


Abbildung 1: Forschungsinformationssystem Systemarchitektur (Eskera 2017)

Ein Projekt zur Konzeptionierung eines Services zum Forschungsdatenmanagement startete in 2015 mit einer universitätsweiten Umfrage zum Umgang mit Forschungsdaten (Hauck et al. 2016). Seit Herbst 2016 wird das erarbeitete institutionelle Konzept zum Forschungsdatenmanagement umgesetzt, das neben dem Ausbau eines integrierten Beratungs- und Schulungsservices zum Thema FDM den Aufbau eines Datenrepositoriums (CKAN) vorsieht⁴.

Im Rahmen des Aufbaus des Datenrepositoriums wird nicht nur die Frage der Schnittstelle zum Volltextrepositorium sondern auch zum Forschungsinformationssystem im Hinblick auf wissenschaftsfreundliche Workflows aufgegriffen. Die Datenpublikation soll nicht als weiterer administrativer Overhead unter Wissenschaftlerinnen und Wissenschaftlern verbucht werden: Ohne großen Aufwand sollte möglichst sowohl die Datenpublikation veröffentlicht als diese auch mit der Volltextpublikation, dem Projekt, in dessen Rahmen die Daten entstanden ist, und ggf. auch der Infrastruktur (Geräte) verknüpft werden.

⁴ www.fdm.uni-hannover.de

Für das weitere Vorgehen erwies es sich als vorteilhaft, dass beide Projekte federführend in einer Abteilung (hier Forschungs- und Transferservice) angesiedelt sind. Das Thema der Systemarchitektur wurde als eigene Teilaufgabe mit weiteren zugeordneten Arbeitspaketen (Workflowanalyse, Schnittstellen, Seamless Integration, administrative Berichtsanforderungen) definiert, die sowohl im Implementierungsprojekt des Forschungsinformationssystems als auch im Umsetzungsprojekt für das institutionelle Forschungsdatenmanagement verankert ist.

2 Ergebnisse der Workflowanalyse

Im Arbeitspaket Workflowanalysen werden mögliche Workflows im Zusammenhang mit der Datenpublikation aufgenommen und auf Vereinfachungen und Anpassung von Schnittstellen analysiert. Darüber hinaus sollen hieraus auch Ansprüche bezüglich der Gestaltung des Backends der - Anwendungen abgeleitet werden. Die Priorisierung der resultierenden Aufgaben soll in eine Roadmap münden.

Abbildung 2 dient zur Veranschaulichung des Workflows der Publikation eines Open Access Volltextes in Verbindung mit der Publikation eines Datensatzes- und verdeutlicht das komplexe Zusammenspiel zwischen Wissenschaftler/innen, Repository Manager, den Repositories und dem Forschungsinformationssystem. Insgesamt wurden in der Bestandsaufnahme über neun Zwischenschritte im Workflow identifiziert. Im Folgenden werden einige Aspekte näher betrachtet:

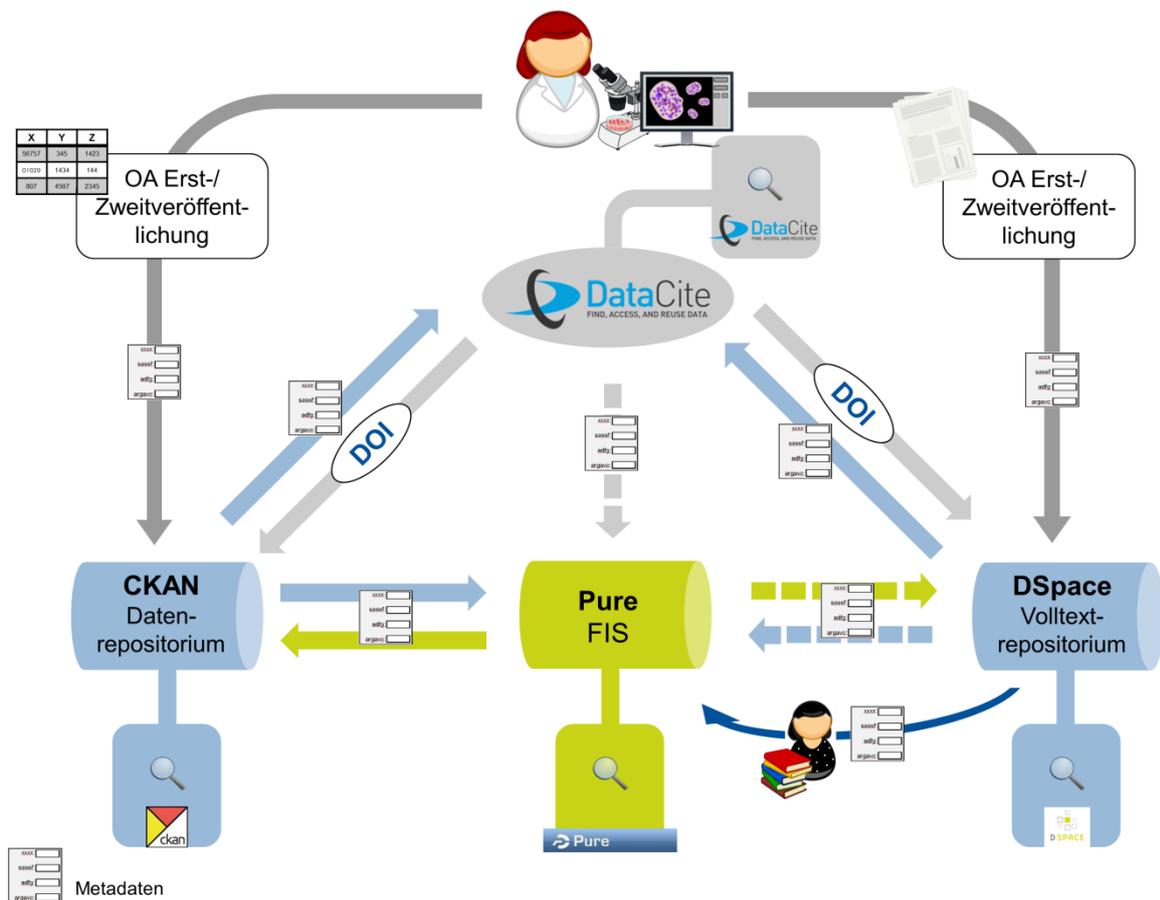


Abbildung 2: Workflowanalyse (Broll, Soßna 2017)

2.1 Verbindung Forschungsinformationssystem und Volltextrepositorium

Es gibt keine Möglichkeit der automatischen Synchronisation von Metadaten aus dem Volltextrepositorium nach Pure. Der Übertrag der Metadaten nach Pure erfolgt im Hintergrund durch den Repository Manager. Für die weitere Ausgestaltung des Workflows sind zeitliche Verzögerungen zu berücksichtigen, so kann z.B. die Prüfung der rechtlichen Randbedingungen der Zweitveröffentlichung eine Zeitverzögerung darstellen ebenso wie die Qualitätsprüfung des Uploads des Datensatzes.

2.2 Verbindung Forschungsinformationssystem und Datenrepositorium

Das Forschungsinformationssystem Pure bietet die Möglichkeit, Metadaten zu Forschungsdaten regelmäßig per xml-Datei zu synchronisieren. Um die Verknüpfung des Datensatzes mit einer Publikation bereits an dieser Stelle zu ermöglichen, soll der Webservice des Forschungsinformationssystems genutzt werden, die in Pure vorhandenen Metadaten zu den eigenen Publikationen in das Datenrepositorium zu übermitteln, um dort dem jeweiligen Autor bzw. Autorin für eine mögliche Verknüpfung angezeigt zu werden.

Die Ausgestaltung der Systemhinweise an die Wissenschaftlerin bzw. den Wissenschaftler im Sinne der Begleitung des Publikationsworkflows muss auf zeitliche Verzögerungen Rücksicht nehmen und die Wiederaufnahme des Workflows vereinfachen bzw. zu geeigneten Zeitpunkten wieder anstoßen. Aus Abbildung 2 geht hervor, dass die Anbindung von DataCite als Quelle für Metadaten die Befüllung des Forschungsinformationssystem deutlich vereinfachen würde. Wissenschaftlerinnen und Wissenschaftler werden (und sollten), wenn möglich Fachrepositorien nutzen, so dass DataCite als Importquelle für Metadatensätze ins Forschungsinformationssystem von hoher Bedeutung ist. Die Leibniz Universität Hannover wird sich in der Pure User Group entsprechend hierfür engagieren und es gibt bereits erste Hinweise auf den Anschluss dieser Quelle bis Februar 2018.

3 Ausblick

In der Ausgestaltung der Systemarchitektur müssen außerdem die Anforderungen der administrativen Berichtspflichten berücksichtigt werden. Auf institutioneller Ebene besteht das Interesse das tatsächliche Verhalten im Forschungsdatenmanagement jenseits der qualitativen Aussagen in Befragungen besser zu verstehen und mit Zahlen zu hinterlegen, u.a. zu Art und Umfang der Forschungsdaten, Anzahl der Datenpublikationen, Zitationen, Orte der Publikation, beantragte FDM-relevante Drittmittel, sowie die Anwendung von Datenmanagementplänen in Projekten. Hiermit kann der erhoffte Kulturwandel im Forschungsdatenmanagement bewertet werden, und die Hochschulleitung kann bei Bedarf steuernd in die Entwicklung der Datenmanagementservices eingreifen. Bei der weiteren Gestaltung der Systemarchitektur muss daher überprüft werden, an welcher Stelle diese Daten generiert werden und für die Berichterstattung zusammengezogen werden können (Khokhar et al. 2016). In Großbritannien wurde hierfür die Software DMAonline entwickelt⁵. Auch die Einbindung eines webbasierten Tools zur Datenmanagementplanung wie z. B. DMPonline muss in diesem Rahmen auf die Vereinfachung von Workflows bzw. Vermeidung von Mehrfacheingaben überprüft werden⁶.

Eine ganzheitliche Systemarchitektur erfordert Voraussicht, die sich entwickelnde Anforderungen mitdenkt und ihnen gegenüber anschlussfähig bleibt. Der Blick nach Großbritannien kann hier

⁵ <http://dmao.info/>

⁶ <https://dmponline.dcc.ac.uk/>

weiterhelfen: Sowohl hinsichtlich der Umsetzung von Datenmanagementkonzepten als auch in der Nutzung von Forschungsinformationssystemen besteht ein erheblicher Erfahrungsvorsprung, der zu nutzen gilt.

Literatur

[1] Hauck, Reingis; Kaps, Reiko; Krojanski, Hans Georg; Meyer, Anneke; Neumann, Janna und Soßna, Volker. 2016. Der Umgang mit Forschungsdaten an der Leibniz Universität Hannover : Auswertung einer Umfrage und ergänzender Interviews 2015/16. Hannover: Institutionelles Repositorium der Leibniz Universität Hannover. DOI: <https://doi.org/10.15488/265>

[2] Khokhar, Masud, Hardy Schwamm, John Krug and Adrian Albin-Clark. 2016. "UK Data Management Administration Online (DMAOnline)". Communicating and Measuring Research Responsibly: Profiling, Metrics, Impact, Interoperability": Proceedings of the 13th International Conference on Current Research Information Systems (2016): Procedia Computer Science (2016, In Press). <http://hdl.handle.net/11366/499>.

[3] Simons, Ed, Mijke Jetten, Marnix van Berchum, Maaïke Messelink, Hans Schoonbrood und Marion Wittenberg. 2016. "The important role of CRIS's for registration and archiving of research data". Communicating and Measuring Research Responsibly: Profiling, Metrics, Impact, Interoperability": Proceedings of the 13th International Conference on Current Research Information Systems (2016): Procedia Computer Science (2016, In Press). <http://hdl.handle.net/11366/524>.