

Forschungsdaten managen – Bausteine für eine dezentrale, forschungsnahe Unterstützung

Dominik Schmitz, RWTH Aachen University Universitätsbibliothek

Marius Politze, RWTH Aachen University IT Center

Zusammenfassung

(Inter-)Nationale Infrastrukturen für Forschungsdaten sind aktuell im Entstehen begriffen. Zentrale Infrastruktureinrichtungen müssen aber bereits jetzt den Bedarf der Wissenschaftlerinnen und Wissenschaftler der eigenen Universität nach forschungsnahe Unterstützung erfüllen. Gemeinsam ist beiden, dass erst die fachspezifische Anpassung und Ausprägung des Forschungsdatenmanagements (FDM) zu einer wirklichen Unterstützung für die Forschenden führt. Aktuell sind gerade die Plattformen erfolgreich, die ganz auf die Bedürfnisse einer wissenschaftlichen Community ausgerichtet sind. Analog ist die Situation an der RWTH Aachen University ebenfalls durch Vielfalt und Dezentralität geprägt. Die einzelnen Institute und Lehrstühle sind unterschiedlich weit bei der Unterstützung ihrer Forschenden. Wie können zentrale Infrastruktureinrichtungen wie die Bibliothek und das Rechenzentrum in diesem Kontext Unterstützung leisten? Dieser Beitrag stellt das bausteinbasierte Lösungskonzept der RWTH vor, benennt konkrete Dienste, die zentral angeboten werden, und argumentiert, warum technologieunabhängige, prozessorientierte Schnittstellen geeignet sind, um forschungsnahe Unterstützung sicherzustellen und sich zugleich in die gerade in der Entstehung befindlichen (inter-) nationalen Forschungsdaten-Infrastrukturen einzubringen.

Summary

At present, (inter-)national research data infrastructure services are about to be established. Central infrastructure facilities such as libraries and IT centers should not wait for this, but even now should meet the needs of the scientists at their University for research support. It is apparent that only research data services that are tailored to the specific needs of a scientific community are really considered supportive by scientists. This is in line with the findings at RWTH Aachen University. The decentral institutes are very diverse with regard to the support they already provide to their researchers. How can central infrastructure facilities such as libraries and IT centers develop suitable support for researchers in this context? This paper presents the modular approach that has been adopted by RWTH Aachen University and explains practical services which are offered centrally. We also discuss why technology independent, process oriented interfaces are well suited both for research support for scientists and the future integration into the emerging (inter-)national research data infrastructures.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/2018H3S76-91>

Autorenidentifikation: Schmitz, Dominik: GND 1014088844, ORCID: <http://orcid.org/0000-0002-2362-4372> ; Politze, Marius: ORCID: <https://orcid.org/0000-0003-3175-0659>

Schlagwörter: Forschungsdatenmanagement

1. Einführung

Forschungsdaten und die dazugehörigen Infrastrukturen stehen aktuell im Fokus. Förderungen und Forderungen gibt es seitens der Deutschen Forschungsgemeinschaft (DFG), dem Bundesministerium für Bildung und Forschung (BMBF), der EU und Verlagen. Insbesondere hat sich der Rat für Informationsinfrastrukturen (RfII) dafür stark gemacht, nationale Strukturen aufzubauen: „Als künftiges neues Rückgrat für das Forschungsdatenmanagement in Deutschland empfiehlt der RfII die Etablierung einer Nationalen Forschungsdateninfrastruktur (NFDI). Diese soll als bundesweites, verteiltes und wachsendes Netzwerk arbeitsteilig angelegt sein.“¹

Der Aspekt der Forschungsdaten-Infrastruktur bezieht sich dabei vor allem auf nationale, europäische und internationale Strukturen wie die angedachte nationale Forschungsdaten-Infrastruktur (NFDI) oder die European Open Science Cloud (EOSC)² und entsprechende Projekte wie der EOSC-Hub.³

Unabhängig von (inter-)nationalen Strukturen sind an einer Universität die zentralen Infrastruktureinrichtungen wie Bibliothek und Rechenzentrum in der Pflicht. Zugleich stehen sie aufgrund der großen Vielfalt von Disziplinen vor enormen Herausforderungen. So dominieren an der RWTH Aachen University zwar zumindest zahlenmäßig Natur- und Ingenieurwissenschaften. Aber wie sich aus den bisherigen Erfahrungen im Projekt „Forschungsdatenmanagement an der RWTH“ gezeigt hat, divergieren die Bedürfnisse selbst innerhalb z.B. der Ingenieurwissenschaften stark. Wie kann nun auf die vielen unterschiedlichen fachspezifischen Bedürfnisse reagiert werden?

Der vorliegende Beitrag baut auf dem Grundkonzept einer umfassenden Service-Suite an der RWTH auf.⁴ In den Vordergrund treten jedoch modulare, integrierbare Bausteine als Grundlage des Lösungskonzepts. Sie adressieren insbesondere die reale Dezentralität. Oft übernehmen einzelne Disziplinen oder Lehrstühle Vorreiterrollen und schaffen eigene Strukturen. Andere setzen hingegen auf überregionale, fachspezifische Dienste. Damit sorgen neben Disziplinspezifika auch unterschiedliche Entwicklungsstände für eine hohe Heterogenität innerhalb einer Universität.

Im Beitrag wird argumentiert, dass auch die hochschul- und länderübergreifenden Strukturen absehbar heterogene Eigenschaften aufweisen werden. Daher ist ein bausteinbasierter Ansatz mit möglichst technologieunabhängigen, prozessorientierten Schnittstellen auch mit Blick auf diese Situation eine verlässliche Grundlage für die forschungsnahe Unterstützung einer sich im Aufbau befindlichen Gesamt-Forschungsdaten-Infrastruktur. Denn ob unterschiedliche Entwicklungsstände

- 1 Rat für Informationsinfrastrukturen (RfII): Leistung aus Vielfalt. Empfehlung zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Göttingen 2016. Online: <<http://nbn-resolving.de/urn:nbn:de:101:1-201606229098>>, Stand: 01.07.2018.
- 2 European Commission: EOSC Declaration: European Open Science Cloud. New Research & Innovation Opportunities, Online: <https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf#view=fit&pagemode=none>, Stand: 01.07.2018.
- 3 EOS-hub: Integrated services for the European Open Science Cloud, <<http://eosc-hub.eu/eosc-hub-integrated-services-european-open-science-cloud>>, Stand: 01.07.2018.
- 4 Eifert, Thomas; Muckel, Stephan; Schmitz, Dominik: Introducing research data management as a service suite at RWTH Aachen University, in: Müller, Paul (Hg.): 9. DFN-Forum Kommunikationstechnologien, Bonn 2016, S. 55-64. Online: <<http://subs.emis.de/LNI/Proceedings/Proceedings257/55.pdf>>, Stand: 01.07.2018.

vor Ort oder (inter-) national adressiert werden, macht keinen Unterschied – es ist nur eine Frage geeigneter Schnittstellen.

2. Strukturen des Forschungsdatenmanagements

In der Idealvorstellung des Forschungsdatenmanagements (FDM) ist der Datenlebenszyklus (s. Abb. 1) vollständig vorab geplant und durchgängig durch geeignete IT-Systeme unterstützt. Entstehende Daten werden zum frühest möglichen Zeitpunkt mit einem Identifier versehen, der es erlaubt, jeden Umgang und jeden weiteren Verarbeitungsschritt zu dokumentieren und zu referenzieren. Peter Wittenburg setzt den Persistent Identifier (PID) von der Bedeutung her mit der IP im Internet gleich.⁵ Erst mit dieser Basistechnologie sind alle darauf aufsetzenden Dienste möglich geworden und natürlich können auf diese Weise – möglichst automatisiert – Metadaten über die Daten und die ausgeführten Verarbeitungsschritte erfasst, dokumentiert und suchbar gemacht werden. Im besten Fall sind all diese Informationen FAIR,⁶ d.h. von überall aus für jeden zugänglich und können so z.B. auch über Disziplinergrenzen hinweg gefunden werden. Tabelle 1 zeigt exemplarisch den Prozess „Daten zu einer Publikation“, der sehr früh im Projekt analysiert wurde.

Tabelle 1: Prozess „Daten zu einer Publikation“

| Nr. Schritt | |
|-------------|--|
| 1 | Erstellung eines Nachweis oder Open Access Veröffentlichung einer wissenschaftlichen Publikation |
| 2 | Hinweis auf die Möglichkeiten zugehörige Forschungsdaten a) ebenfalls zu veröffentlichen (lokal oder in einem Fachrepositorium) oder b) zumindest zu archivieren (10 Jahre) und mit der Publikation zu verknüpfen. |
| 3a | Link auf Repositorium ergänzt um Verknüpfungsinformation |
| 3b | Link auf Archiv zum Anlegen eines Archivknotens inkl. Verknüpfungsinformation |
| 4a | Erstellung eines Nachweis oder Open Access Veröffentlichung von Daten zur Publikation (analog zu Schritt 1) |
| 4b | Erzeugung der Archivdatei, Speicherung im Archiv, Erstellung einer PID inkl. Verknüpfungsinformation zur Publikation |
| | 5b Hinweis auf die Möglichkeit Metadaten zu den Forschungsdaten zu erfassen |
| | 6b Optional: Link auf Metadaten-Werkzeug mit einrichtungsspezifischem Metadatenschema |
| | 7b Optional: Verknüpfung der Metadaten (öffentlich oder auch nicht) mit der PID und so auch der Publikation |

5 Wittenburg, Peter: Data Foundation & Terminology WG Data Fabric IG. Präsentation auf dem RDA-DE-DINI Workshop „Aktuelle Resultate der Research Data Alliance (RDA) und deren zukünftige Bedeutung“, 2015-05-28/29 in Karlsruhe, Deutschland, S. 10. Online: <http://www.forschungsdaten.org/images/fc/RDA-DE-2015_Wittenburg_Peter_III.pdf>, Stand: 01.07.2018.

6 Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan u.a.: The FAIR Guiding Principles for scientific data management and stewardship, in: Scientific data 3 (160018), 2016. Online: <<https://doi.org/10.1038/sdata.2016.18>>, Stand: 01.07.2018.

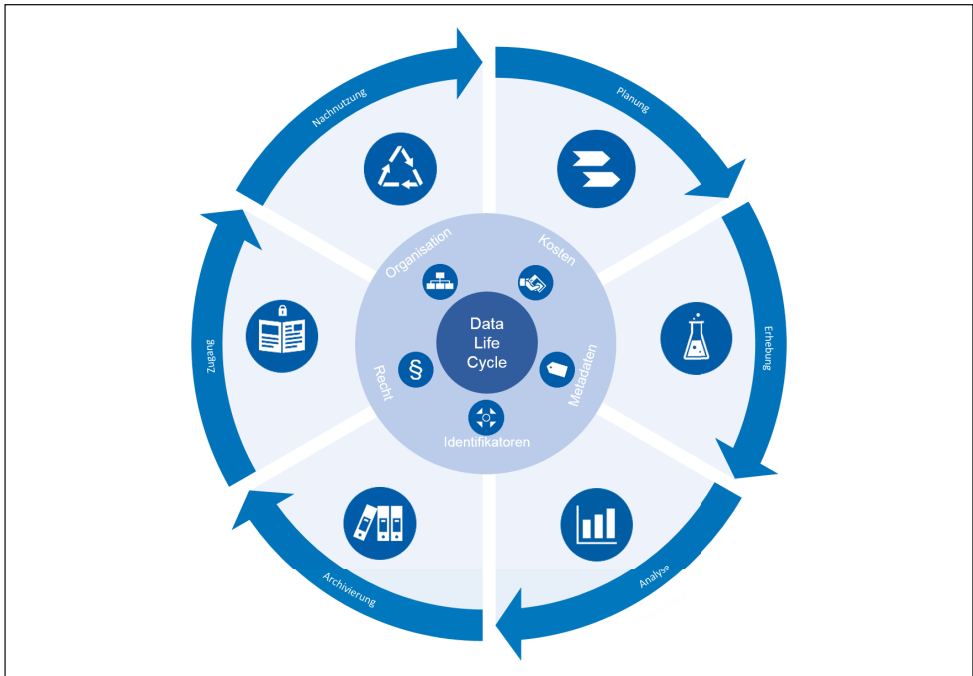


Abb. 1: Datenlebenszyklus

Der Rat für Informationsinfrastrukturen hat in seinem Übersichtspapier vom Juni 2017 die weltweiten Aktivitäten zusammengefasst.⁷ Einzelne Länder, die früh begonnen haben und daher auch Mitbegründer wichtiger Initiativen wie der Research Data Alliance⁸ sind, z.B. Australien, sind bereits gut aufgestellt. Dies ist jedoch nicht gleichbedeutend damit, dass die Strukturen schon flächendeckend existieren. Gerade die Research Data Alliance mit ihren „interest-“ und „working-groups“ setzt auf die Bildung fachspezifischer Interessengruppen, die den Umgang mit Forschungsdaten in der jeweiligen Disziplin oder Community voranbringen.

Innerhalb von Deutschland sind Leuchttürme vornehmlich auf Basis von Projektförderungen entstanden, etwa im Bereich Erd- und Umweltwissenschaften (PANGAEA)⁹ oder der Biologie (gfbio).¹⁰ Diese Strukturen rücken das ernstzunehmende Problem der aktuellen Situation in den Vordergrund. Nach Ablauf der Förderung ist es ungewiss, wie die geschaffenen Strukturen weiter existieren und gegebenenfalls wachsen können. Die Bereiche, in denen die Nachnutzung von Daten Tradition hat

7 Rat für Informationsinfrastrukturen (RfII): Entwicklung von Forschungsdateninfrastrukturen im internationalen Vergleich: Bericht und Anregungen, Göttingen 2017. Online: <<http://www.rfii.de/download/rfii-fachbericht-laenderanalysen-2017/>>, Stand: 01.07.2018.

8 Research Data Alliance, <<https://rd-alliance.org/node>>, Stand: 01.07.2018.

9 Pangaea, <<https://www.pangaea.de/>>, Stand: 01.07.2018.

10 Research Data Management, German Federation for Biological Data, <<https://www.gfbio.org/>>, Stand: 01.07.2018.

und methodisch fundiert ist, allen voran GESIS und RatSWD, sind hier schon einen Schritt weiter. An diesen Leuchttürmen orientiert sich auch die Konzeption der nationalen Forschungsdateninfrastruktur (NFDI). Wie ausgeführt, wird das Thema sowohl in Deutschland als auch in der EU politisch forciert. Es steht zu erwarten, dass die aktuell entstehende nationale Forschungsdateninfrastruktur zunächst einmal die Probleme der bisher primär projektbezogenen Förderung löst. Es ist zu hoffen, dass endlich die erforderlichen Mittel zur Verfügung gestellt werden, um aus den vielen erfolgreichen Projekten tatsächlich dauerhaft die Strukturen zu schaffen, die den geforderten „nachhaltigen Zugang“ sicherstellen können. In seinen diesbezüglichen Empfehlungen weist der Rat ganz prominent daraufhin, dass es dabei nicht um eine übergreifend vereinheitlichende Infrastruktur geht. Der gewählte Titel „Leistung aus Vielfalt“ macht das explizit. Zugleich impliziert dies aber auch, dass schon auf dieser übergreifenden Ebene Diversität und Heterogenität der Angebote und Daten bestehen bleiben werden.

3. Forschungsnahe Unterstützung

3.1. Interessen und Bedarfe an der RWTH Aachen University

Die disziplinspezifische Heterogenität bedingt, dass die Universität diese Vielfalt unterstützen muss. Hierzu addiert sich die unterschiedliche Ausstattung, Interessens- und Leistungsbereitschaft einzelner Einrichtungen. Es gehören oftmals spezifische Geräte wie Laborausstattungen oder IT-Systeme zur lokalen Infrastruktur, wodurch sich die Strukturen zur Unterstützung der Forschenden noch stärker unterscheiden.

Konkrete Erfahrungen wurden in Aachen sowohl mit einzelnen Einrichtungen als auch mit größeren Verbundprojekten gesammelt. Im Lehrstuhl für Technische Thermodynamik, einem der Pilotpartner im initialen FDM-Projekt, hat die Beschäftigung mit dem Thema FDM zur Einführung neuer Datenbanken und Informationsstrukturen geführt, die zum einen auf bestehende Strukturen (Intranet, Gerätedatenbank) zurückgreifen und zum anderen bewusst neue Dienste (EPIC-Persistent Identifier) integrieren. Im SFB 985 „Mikrogele“ war von Beginn an das Probenmanagement als zu adressierende, übergreifende Datenmanagement-Dienstleistung identifiziert worden. Auf Basis der Kollaborationsplattform MS SharePoint und in Verbindung mit QR-Codes wurde eine Umgebung aufgebaut, bei der Proben leichter über Einrichtungsgrenzen bis hin zu den Publikationen, in die sie Eingang gefunden haben, verfolgt werden können. Ein weiterer Anwendungsfall ist das Gemeinschaftslabor für Elektronenmikroskopie, das als Dienstleister für die Hochschule tätig ist. Hier geht es vor allem darum, effiziente Strukturen zur Datenverwaltung und -übergabe zu finden, die die erforderlichen engen Begrenzungen von Datensichtbarkeit gemäß Rechtemanagement beachten. Daher wird der Einsatz eines ObjectStore-Speichers in Verbindung mit dem EPIC-Dienst pilotiert.

Diese Anforderungen werden durchgängig durch entsprechende Interviews, Umfragen, Workshops, Beratungen sowie Weiterbildungsangebote aber auch in der Literatur bestätigt.¹¹ Wie schnell sich

¹¹ Vgl. Curdt, Constanze; Hoffmeister, Dirk; Jekel, Christian u.a.: Implementation of a centralized data management system for the CRC Transregio 32 „Patterns in Soil-Vegetation-Atmosphere-Systems“, in: Curdt, Constanze; Bareth, Georg (Hg.): Proceedings of the Data Management Workshop, 29.-30.10.2009, University of Cologne, Germany, Köln 2010, S. 27-33. (Kölner Geographische Arbeiten 90); Kirsten, Toralf; Kiel, Alexander; Wagner, Jonas u.a.: Selecting,

lokale und nationale Infrastrukturen vernetzen, ist am Projekt SMITH (Smart Medical Information Technology for Healthcare) in der Medizininformatik zu sehen.¹² Gemeinsam mit Leipzig und Jena entstehen hier sogenannte Datenintegrationszentren (DIZs), die insbesondere eine Lösung für die kritische Verknüpfung von Forschungs- und Patientendaten bieten. Nach abgeschlossener Realisierung für die im Projekt vorgesehenen Use Cases sollte sich die Lösung im Prinzip auf gleichartige Anwendungsfälle übertragen lassen. Schließlich ist die RWTH – vielleicht in besonderer Weise und historisch bedingt – durch hohe Autonomie der Lehrstühle und Institute geprägt. Zentrale Angebote wie Backup und Datenarchiv sind optional, ihre Nutzung ist nur in engen Grenzen verpflichtend.

3.2. Lösungskonzept: Modulare, integrierbare Bausteine

Wo Vielfalt und Dezentralität überwiegen, müssen zentrale Angebote durch ihre Qualität und Integrierbarkeit überzeugen. Genau darauf zielt das *bausteinbasierte Konzept* an der RWTH ab. Ausgehend von den Diensten, die jetzt bereits als verlässliche Angebote angenommen werden, sollen neue Schnittstellen, neue Angebote und neue kombinierte Dienste den Forschenden das Forschungsdatenmanagement erleichtern. Ziel ist es eine grundlegende Infrastruktur vernetzter Dienste und Services zu schaffen, die insgesamt das Forschungsdatenmanagement für ein Institut und damit für die gesamte RWTH unterstützen kann. Die Einrichtungen haben aber die Freiheit, mehrere oder einzelne Komponenten auszutauschen und im Bedarfsfall durch eigene, spezifischere und daher die Forschenden noch besser unterstützende Komponenten zu ersetzen (z.B. eine ObjectStore für die Mikroskop-Infrastruktur oder ein elektronisches Laborbuch).

Zwei Fragen stehen am Anfang: welche Dienste eignen sich als Bausteine für ein solches Lösungskonzept und wie muss die Schnittstelle eines Dienstes gestaltet sein, damit er über verschiedene Disziplinen hinweg verwendet werden kann? Während im nächsten Kapitel die einzelnen, an der RWTH identifizierten Dienste vorgestellt werden, wird hier vorab die Frage der Schnittstelle geklärt. Lösung an der RWTH ist die *Definition von technologieunabhängigen und prozessorientierten Schnittstellen*. An die Stelle von spezifischen Formaten und Protokollen tritt eine unabhängige und möglichst offene Spezifikation, die sich an den zu unterstützenden Prozessen orientiert. Diese abstrakte Zwischenschicht übernimmt die Übersetzung zwischen den Prozessen der Forschenden und der Implementierung der Hersteller oder Dienstbetreiber. Die Konzeption steht im Einklang mit aktuellen Standards in der Softwarearchitektur wie (Micro-/Web-) Service-orientierte Architekturen.¹³ Dieses Vorgehen bedeutet nicht unerheblichen Mehraufwand, denn die notwendige Modellierung der Prozesse erfordert die Zusammenarbeit zwischen technischen sowie organisatorisch dienstleistenden und den nutzenden Einrichtungen. Er wird jedoch damit vergolten, dass entsprechend modellierte Systeme besser zu den forschungsnahen Abläufen in der jeweiligen Disziplin und Einrichtung passen. Als weiterer wesentlicher Vorteil neben der Vereinfachung von Migrations- und Exit-Szenarien bietet dieses Vorgehen die Möglichkeit, bestehende, heterogene Systemlandschaften zu integrieren.

Packaging, and Granting Access for Sharing Study Data, in: Eibl, Maximilian; Gaedke, Martin (Hg.): INFORMATIK 2017: Digitale Kulturen. Beiträge der 47. Jahrestagung der Gesellschaft für Informatik e.V. (GI), S. 1381-1392.

12 Smart Medical Information Technology for Healthcare, <<http://www.smith.care/>>, Stand: 01.07.2018.

13 Vgl. Dumas, Marlon (Hg.): Process-aware information systems: Bridging people and software through process technology, Hoboken NJ 2005; Wolff, Eberhard: Microservices: Grundlagen flexibler Softwarearchitekturen, 1., korr. Nachdr. Heidelberg 2016.

Aus Sicht des Nutzenden verschwimmen die Grenzen zwischen den einzelnen Systemen oder verschwinden sogar ganz. Es ist diese Bündelung und Integration von bestehenden Diensten, die dem Forschenden Mehrwert, Entlastung und Nutzen bringt.¹⁴

4. Bausteine für eine forschungsnahe Unterstützung

In der ersten Projektphase sind bereits mehrere Bausteine für eine forschungsnahe Unterstützung des Forschungsdatenmanagements an einzelnen Lehrstühlen und Instituten entstanden, die im Folgenden vorgestellt werden.

4.1. Persistent Identifier (PID)

Die Vergabe von Persistent Identifiern (PID) nimmt einen besonderen Stellenwert ein. Die PID ist Dreh- und Angelpunkt für die Annotation und Verknüpfung von Forschungsdaten. Hervorzuheben ist an dieser Stelle, dass eine PID nicht mit der DOI gleichzusetzen ist. Deutlich machen das die Klimarechenzentren, die weltweit Simulationen von Klimamodellen koordinieren müssen. Ein automatisiert vergebener Persistent Identifier hilft hier unter anderem, die komplexen Beziehungen zwischen den Simulationsergebnissen zu dokumentieren. Aber längst nicht alle Simulationen führen zu verwertbaren Informationen. Einen zitierbaren Persistent Identifier (= DOI) erhalten später die Daten, die tatsächlich in den Weltklimabericht eingehen.¹⁵

Analog zum Vorgehen im Bereich Klimaforschung setzt der Baustein *Persistent Identifier* an der RWTH ebenfalls auf die Kerneigenschaften Eindeutigkeit und Dauerhaftigkeit und fokussiert darauf, diesen elementaren Dienst möglichst einfach und vielseitig einsetzbar zu machen. Hervorzuheben in diesem Kontext sind die im Vergleich zur DOI stark reduzierten Anforderungen an Metadaten. Die Konfiguration des European Persistent Identifier Consortium (EPIC)-Dienstes¹⁶ an der RWTH (genutzt über die Universität Göttingen) erzwingt neben der ID selber nur genau eine zusätzliche Angabe: den Verweis auf den Speicherort der zu identifizierenden Daten. Die EPIC wird primär im Kontext der Bitstream Archivierung im Archivsystem der RWTH genutzt. Grundsätzlich kann der Verweis aber auf einen beliebigen Speicherort erfolgen. Insbesondere ist es so denkbar, dass der Verweis nur für das jeweilige Institut erreichbar oder sogar nur für dieses interpretierbar ist. Mit einer so starken Dezentralität geht natürlich auch eine große Verantwortung zum Betrieb der dezentralen Systeme (z.B. Fileserver) einher, auf die verwiesen wird. An dieser Stelle werden die Anforderungen und Voraussetzungen in Zukunft mit Sicherheit steigen und entsprechend werden verlässliche Aussagen hierzu gemacht werden müssen. In der technischen Umsetzung löst die EPIC für externe Nutzerinnen und Nutzer auf eine Landing Page auf, die eben keine Informationen preisgibt, sondern lediglich eine Kontaktaufnahme ermöglicht.

14 Vgl. Juling, Wilfried: Vom Rechnernetz zu e-Science, in: PIK - Praxis der Informationsverarbeitung und Kommunikation, 32, 2009, S. 33-36. Online: <<https://doi.org/10.1515/piko.2009.007>>, Stand: 01.07.2018.

15 Buurmann, Merret: Tutorial Session PIDs in der Forschung. Präsentation auf dem RDA-Deutschland-Treffen 2015, Potsdam 25./26.11.2015, S. 21. Online: <https://os.helmholtz.de/fileadmin/user_upload/os.helmholtz.de/Workshops/rda_de_15_buurman.pdf>, Stand: 01.07.2018.

16 Kálmán, Tibor; Kurzawe, Daniel; Schwardmann, Ulrich: European Persistent Identifier Consortium – PIDs für die Wissenschaft, in: Altenhöner, Reingard; Oellers, Claudia (Hg.): Langzeitarchivierung von Forschungsdaten: Standards und disziplinspezifische Lösungen, Berlin 2012, S. 151-164.

Neben diesen verpflichtenden Daten stehen nur zwei Datenfelder zur Verfügung, die optional gefüllt werden können: der Hinweis auf einen Speicherort für zusätzliche Metadaten sowie die Option, auf relevante Publikationen, die in der Anwendung RWTH Publications¹⁷ verzeichnet sind, zu verweisen. Die RWTH war eine der ersten Einrichtungen, die die Metadaten im PID-System minimiert und einen Verweis auf zusätzliche, außerhalb des PID-Systems gespeicherte Metadaten ermöglicht und noch dazu optional gemacht hat. So viel Freiheit lässt kaum eine andere Universität zu. Daher muss man sich die Chancen der hohen Flexibilität dieser Lösung zunächst einmal bewusst machen. Die Metadaten nicht direkt im PID-System zu speichern, sondern zu erlauben auf einen externen Speicherort zu verweisen, der sowohl eine statische Datei, ein physisches Buch oder ein anderes IT-System mit entsprechendem Parameter sein kann, macht es leicht, *existierende Systeme oder fachspezifische Systeme* einzusetzen. Wenn also ein Großgerät bereits Metadaten zu Experimenten dokumentiert, kann leicht darauf verwiesen werden. Wichtig ist, dass die Sichtbarkeit der Metadaten außerhalb dieses Kontextes definiert wird. Insbesondere wenn das verlinkte System ein ausgefeiltes Rechte-Management besitzt, greift das an dieser Stelle ohne weiteres Zutun. Wer dem Link folgt, muss dem erreichten System gegenüber erst einmal nachweisen, dass er oder sie eine Berechtigung für den Zugriff auf die (Meta-)Daten hat.

Natürlich birgt dieser Ansatz auch Risiken. So begrenzt er die Möglichkeiten zentral verfügbarer Angebote und Informationen. Selbst generische Metadaten wie Autor, Titel, Jahr sind im Gegensatz zu einem zentralisierten Datenrepositorium oder -archiv auf diese Weise nicht per se universitätsweit such- oder zugreifbar. Für die eingeschränkte Sichtbarkeit spricht, dass die Entscheidung, wer auf welche (Meta-) Daten zugreifen darf, für jeden Einzelfall und ggf. unter Berücksichtigung juristischer Aspekte zu treffen ist. Aber einer Herausgabe aller Daten oder auch nur eines Ausschnitts der Metadaten steht grundsätzlich technisch nichts im Wege. Mit geeigneten Schnittstellen können diese Informationen über die Links automatisiert abgerufen und in einen geeigneten Suchraum kombiniert werden. Grundlage dafür ist ein Einvernehmen in der Hochschule und eine beherrschbare Umsetzung mit Standard-Technologien wie Indexierung und Rechercheoberfläche.

Warum aber sollte man erlauben, dass gar keine Metadaten verlinkt werden? Zunächst einmal stellt die Freiwilligkeit den aktuellen Ist-Zustand dar und es ist nicht festgelegt, dass das immer so bleiben wird. Wichtig ist für die RWTH aktuell, die Schwelle für die Nutzung des Dienstes möglichst niedrig zu halten. Dabei sind zwei Szenarien im Fokus.

1. Schon jetzt unterstützt die Vernetzung der Anwendungen (s. Tabelle 1 und Kapitel 4.5) den Standardfall, dass zu Textpublikationen, die im Repository verzeichnet oder sogar als Volltext publiziert sind, die Forschungsdaten, die nicht veröffentlicht werden können oder sollen, aber im zentralen Archivsystem gespeichert sind, über die EPIC verknüpft werden. Das heißt von der Publikation (z.B. der Dissertation) aus sind die Forschungsdaten über die EPIC erreichbar. Häufig genügt den Forschenden dieser Link, denn sie argumentieren, dass die jeweilige Publikation die Forschungsdaten ausreichend beschreibt. Der Idealvorstellung des FDMs

17 RWTH Publications, <<http://publications.rwth-aachen.de>>, Stand: 01.07.2018.

entspricht das nicht, aber eben der Realität. Entsprechend sind Metadaten, die ohnehin weitestgehend den Metadaten der verknüpften Publikation entsprechen, nicht erforderlich.

2. Das zweite Szenario adressiert die missliche, aber immer noch verbreitete Situation, bei der die Daten bisher nicht im Archivsystem, sondern auf Festplatten im Schrank des Lehrstuhlinhabers oder Systemadministrators lagern. Dem gegenüber ist ein Archivknoten, in dem die Daten sicher für die nächsten 10 Jahre liegen und der immerhin mit der Person verknüpft ist, eindeutig vorzuziehen. Zumal, siehe Szenario 1, der Bezug zu den jeweiligen Publikationen im Gegensatz zur physischen Festplatte sehr leicht via EPIC hergestellt werden kann.

4.2. Publizieren

Wie in den vorangegangenen Kapiteln ausgeführt, sind dauerhaft heterogene Strukturen für die sinnvolle Publikation von Forschungsdaten zu erwarten. Sicherlich ist die aktuelle Vielfalt an Repositorien, die in re3data.org verzeichnet werden, weder zielführend noch infrastrukturell plausibel und dauerhaft am Leben zu erhalten. Aber die Disziplinen und daher mindestens einmal die Anforderungen an die fachspezifischen Beschreibungen sind zu unterschiedlich, als dass jemals ein einzelnes Repository alle Forschungsdaten erfassen wird.

An der RWTH hat die UB die Aufgabe übernommen, Forschende bei der Auswahl eines geeigneten Repositoriums zu unterstützen. Hier gilt der Grundsatz „anerkanntes Fachrepository zuerst“. Dem liegt die Annahme zugrunde, dass dort sowohl die Sichtbarkeit der Daten vor allem innerhalb der Disziplin als auch die inhaltliche Beschreibung und Betreuung besser gegeben ist. Die Wahrscheinlichkeit, dass jemand auf der Suche nach sozialwissenschaftlichen Forschungsdaten das institutionelle Repository der RWTH aufsucht, scheint extrem gering, vor allem wenn mit GESIS Strukturen existieren, die sich offensichtlich besser als Suchstartpunkt eignen. Wer dagegen will, dass die selbst entwickelte Software gefunden wird, sollte sie aktuell auf GitHub¹⁸ veröffentlichen. Die einfache Möglichkeit, über die Zenodo-Integration an eine DOI zu gelangen, ist sicherlich empfehlenswert.¹⁹ Denn wenn in Zukunft eine neue bessere Plattform für Software entsteht, bleibt die DOI stabil und kann auf das neue Ziel verweisen.

Auch für die langfristige Vision, dass Forschungsdaten interdisziplinär nachgenutzt werden, ist es sinnvoller, wenn eine geringe Zahl fachspezifischer Repositorien sich als Infrastruktureinrichtungen die Mühe machen, Datenbestände kompatibel zu beschreiben und übergreifende Suchen zu ermöglichen, statt dass an jeder Einrichtung nur minimale generische, bibliographische Metadaten zu Forschungsdaten erfasst werden. Die Publikation von Forschungsdaten auf dem institutionellen Repository RWTH Publications ist somit vor allem dann eine Option, wenn (noch) kein geeignetes Fachrepository existiert. Hierfür bietet der Invenio²⁰-basierte und in der JOIN²-Kooperation²¹

18 GitHub, <<https://github.com>>, Stand: 01.07.2018.

19 Making your code citable, GitHub, <<https://guides.github.com/activities/citable-code/>>, Stand: 01.07.2018.

20 INVENIO, <<http://invenio-software.org>>, Stand: 01.07.2018.

21 JOIN², <<http://join2.de>>, Stand: 01.07.2018.

entwickelte Dienst ausreichend Platz, nötigenfalls unter Rückgriff auf die Bausteine ObjectStore oder hierarchische Speichersysteme (HSM).

4.3. Archivieren

Für publizierte Daten gibt es kein Verfallsdatum; sie stehen jedem jederzeit zur Nutzung zur Verfügung. Daher steht die unmittelbare Verfügbarkeit abgesichert durch geeignete Maßnahmen wie Backup im Vordergrund. Im Hintergrund ist es sinnvoll auf Werkzeuge zur Langzeitarchivierung zurückzugreifen, um die Zugreif- und Lesbarkeit durch etwaige Formatumschreibungen auch dauerhaft sicherzustellen.

Anders sieht es bei Daten aus, die nur archiviert werden. Als Minimum sind hier die von der DFG geforderten 10 Jahre zu sehen.²² Für diese Zwecke genügt eine Bitstream Preservation, die nicht berücksichtigt, ob das Format fortgeschrieben werden muss. Es ist zu beachten, dass eine echte Longterm Preservation ungleich höheren Aufwand verursacht, wie an den entsprechenden Zertifizierungsanforderungen zu erkennen ist.²³ Die Entscheidung zwischen Bitstream und Longterm Preservation muss bewusst und informiert getroffen werden.

An der RWTH stehen grundsätzlich beide Technologien zur Verfügung. Die Erfahrungen mit der Bitstream Preservation sind bereits sehr intensiv, da die Hochschule Konsortialführer für die Backup- und Archividienste in NRW ist. Die technischen Systeme werden kompetent beherrscht. Das schließt insbesondere auch die Kenntnisse ihre Defizite in diesem Umfeld ein. Eine unmittelbare Archivierung in das Archivsystem ist Forschenden aufgrund der unzureichenden Usability des entsprechenden Clients nicht zumutbar. Daher hat das IT Center eine einfache webbasierte Oberfläche analog zur bekannten Anwendung „GigaMove“ geschaffen. Mit Hilfe von *SimpleArchive* können Daten bequem über das Web in Archivknoten gesichert und von dort wiederhergestellt werden.²⁴ Zugleich sind damit Mehrwerte verknüpft: Die Vergabe einer EPIC-PID sowie die Nutzung des DFN-Zeitstempel-Dienstes. Der Einsatz von einer am HBZ (Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen) gehosteten Rosetta-Instanz für die Longterm Preservation ist im Pilotstadium. Die Einbindung ist trivial: eine EPIC kann als Speicherort natürlich auf ein System der Langzeitarchivierung verweisen.

4.4. Metadaten

Auch wenn die bisher beschriebenen Komponenten primär abgeschlossene Forschungsdaten (im Datenlebenszyklus (s. Abb. 1) die Phasen Archivierung, Zugang und Nachnutzung) adressieren, ist die Idealvorstellung des Forschungsdatenmanagements an der RWTH nicht vergessen. Insbesondere

22 Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten, 2015, S. 1. Online: <http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf>, Stand: 01.07.2018.

23 nestor-Arbeitsgruppe Zertifizierung: Erläuterungen zum nestor-Siegel für vertrauenswürdige digitale Langzeitarchive - Version 2.0, Frankfurt am Main 2016 (nestor-materialien 17,2). Online: <<http://nbn-resolving.de/urn:nbn:de:0008-2016111106>>, Stand: 01.07.2018.

24 Politze, Marius; Krämer, Florian: simpleArchive - Making an Archive Accessible to the User, in: EUNIS 2017 - Shaping the Digital Future of Universities, Proceedings of the 23rd EUNIS Congress. Münster 2017, S. 121-123. Online: <http://www.eunis.org/download/2017/EUNIS2017_Book_of_Proceedings_1.pdf>, Stand: 01.07.2018.

das Domänenmodell²⁵ führt klar vor Augen, dass mit wachsender Sichtbarkeit die Anforderungen an begleitende, erläuternde Metadaten steigen. Zugleich ist bekannt, dass die nachträgliche Erhebung wesentlich höhere Aufwände verursacht, als wenn dieselben Daten ggf. sogar automatisiert direkt zu Beginn erhoben werden. Entsprechend sieht die Konzeption an der RWTH ein generisches Werkzeug zum Erzeugen und standardisierten Dokumentieren von fachspezifischen Metadaten vor. Das RDF²⁶-basierte Werkzeug ist aktuell noch in der Entwicklung, wird aber bereits mit konkreten Anwendern in der Hochschule pilotiert. Primär wichtig ist, dass es sich zur Bildung eines institutsspezifischen Metadatenmodells auf mehrere Metadatenstandards z.B. auf die entsprechenden Metadatenstandard-Directories oder die PID information types, die im Rahmen der Research Data Alliance entstanden sind, stützen kann. Um der Verantwortung zur Vernetzung bei Bedarf gerecht werden zu können, haben sich die zentralen Einrichtungen dazu entschieden, das RADAR-Modell als Minimaldatenmodell in allen spezifischen Metadatenmodellen als verpflichtend vorauszusetzen.²⁷ Essentiell ist aber, dass die Metadatenmodelle auf die konkreten Institute und Anwendungsfälle angepasst werden. Die Forschenden bestimmen, welche Informationen für sie relevant sind. Die Bibliothek unterstützt durch die Recherche nach geeigneten, möglicherweise fachspezifischen Metadatenstandards. Das IT Center setzt das individuell auf die Einrichtung angepasste Metadatenmodell in RDF um und stellt es der Einrichtung als Webanwendung zur Verfügung.

Erwähnenswert ist, dass das Werkzeug natürlich eine Oberfläche zur manuellen Eingabe von Daten bietet, dass hier aber nicht der Fokus gesehen wird. De facto ist die Oberfläche nur eine Anwendung, die, wie viele andere auch, die zugrundeliegende API ansprechen kann, um Metadaten zu erzeugen. Es ist absehbar, dass mehr und mehr Daten automatisiert erzeugt oder zumindest annotiert werden. D.h. der jeweilige Dienst muss disziplin- oder einrichtungsspezifisch entsprechend anprogrammiert werden können. Diese Anforderung erfüllt die Anwendung bereits im Prototyp-Status. Erst so lässt sich das eigentliche Ziel des Forschungsdatenmanagements erreichen: eine möglichst nahtlose Integration in die täglichen Arbeitsprozesse der Forschenden. Diese Anpassbarkeit an den konkreten Bedarf wird von den zentralen Einrichtungen als wesentlich angesehen, um bei Forschenden Akzeptanz zu erzielen.

Wichtig ist zu bemerken, dass das hier beschriebene Metadatentool nur eine von mehreren Möglichkeiten darstellt, wie Metadaten zu Forschungsdaten erfasst werden können. Existieren in einzelnen Lehrstühlen, Fakultäten oder Einrichtungen andere, für bestimmte Zwecke geeignetere Systeme (z.B. ein elektronisches Laborbuch, eine angepasste Simulationssoftware etc.), dann sollten diese Verwendung finden. Aufgrund der generischen Ausführung kann das Metadatentool in diesem Kontext aber immer als Komponente einer Exit-Strategie gesehen werden: Wenn ein spezifisches Werkzeug doch das Ende seines Lebenszyklus erreicht, realisiert das Metadatentool auf solider

25 Klar, Jochen; Enke, Harry: DFG-Projekt RADIESCHEN - Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur, Report „Organisation und Struktur“, 2013. Online: <http://dx.doi.org/10.2312/RADIESCHEN_005>, Stand: 01.07.2018.

26 Resource Description Framework, <<https://www.w3.org/RDF/>>, Stand: 19.02.2018.

27 Kraft, Angelina; Razum, Matthias; Pothhoff, Jan u.a.: The RADAR Project—A Service for Research Data Archival and Publication, in: ISPRS International Journal of Geo-Information 5 (3), 2016, Nr. 28. Online: <<https://dx.doi.org/10.3390/ijgi5030028>>, Stand: 01.07.2018.

informationstechnischer Basis mit Filesystem und standardisiertem Dateiformat (RDF) einen Fallback-Metadatenpeicher. Dieser Schritt erfordert natürlich eine in der Regel aufwändige Migration, aber zumindest besteht über ein zentral angebotenes System diese Option. Die Zwischenstufe Persistent Identifier erlaubt es, die erforderlichen Anpassungen in der Dokumentation ohne Informationsverlust umzusetzen.

Die nächsten Jahre werden zeigen, ob sich (einzelne) dezentrale Systeme zur (Meta-)Datenspeicherung bewähren oder ob sich langfristig doch verlässliche, zentralisierte Systeme durchsetzen werden. In der jetzigen Situation hilft an dieser Stelle nur Beratung. Dass die erforderliche Verlässlichkeit den Beteiligten durchaus bewusst ist, zeigt sich auch darin, dass in Verbindung mit EPIC-Service und Metadatentool aktuell mit großer Mehrheit Archivknoten als Speicherort von den Instituten und Lehrstühlen gewählt werden. Wenn Einrichtungen ein zentrales Angebot nutzen, müssen sie möglicherweise damit einhergehende Einschränkungen in Kauf nehmen. Dafür geben Sie Verantwortung und Aufwand für die technische Umsetzung und deren Erhalt ab. Abhängig von der konkreten Situation wird dies, wie es sich hier andeutet, aus rationaler Sicht häufig die bessere Wahl sein.

4.5. Beispiel für das Zusammenwirken von Diensten

Auf Basis der vorgestellten Technologien lassen sich verschiedene Prozesse definieren, die ein dezentrales Datenmanagement unterstützen. Realisiert ist an der RWTH bereits der in Tabelle 1 vorgestellte Prozess: Der Nachweis einer wissenschaftlichen Publikation und der zugrundeliegenden Daten unabhängig davon, ob sie veröffentlicht oder nur archiviert sind (s. Abb. 2).

Die Anwendung RWTH Publications ist Nachweis und Publikationsserver der RWTH. Jeder wissenschaftliche Output der RWTH wird hier verzeichnet. Daher liegt es nahe, dass die Nutzenden nach Hinzufügen einer Text-Publikation auch nach den zugehörigen Forschungsdaten gefragt werden. Der Hilfe-Text enthält einen Hinweis auf die Beratung zu Fachrepositorien, einen Link für einen Datennachweis oder eine Datenveröffentlichung in RWTH Publications sowie einen Link auf die Archivierung mit Hilfe des Dienstes SimpleArchive. Als Besonderheit enthält dieser Link, der aus dem Nachweis der Text-Publikation auch zu jedem späteren Zeitpunkt heraus aufgerufen werden kann, genug Informationen, damit die archivierte Datei automatisiert im Nachweis-Datensatz verlinkt werden kann. Damit besteht der Aufwand für den Forschenden für diesen einfachsten Weg, um der guten wissenschaftlichen Praxis Genüge zu tun, darin, ein Archiv mit den relevanten Daten zu erzeugen und in SimpleArchive hochzuladen. Analog lässt sich auch das Metadatentool wie in Tabelle 1 beschrieben als optionalen Schritt integrieren.

Die entsprechenden Schnittstellen orientieren sich am Prozess und sind aus technologischer Sicht einfach gehalten: REST-basierte APIs, die entsprechende Informationen als Aufrufparameter übertragen oder zurückmelden, ohne zu eng an die realisierende technische Plattform gekoppelt zu sein. Essentiell ist in diesem Kontext natürlich das Rechtemanagement. Hier konnte auf die umfangreichen

Erfahrungen mit der Authentifizierung via Shibboleth in Kombination mit der Autorisierung via OAuth aus dem Kontext Lehre und RWTHApp zurückgegriffen werden.²⁸

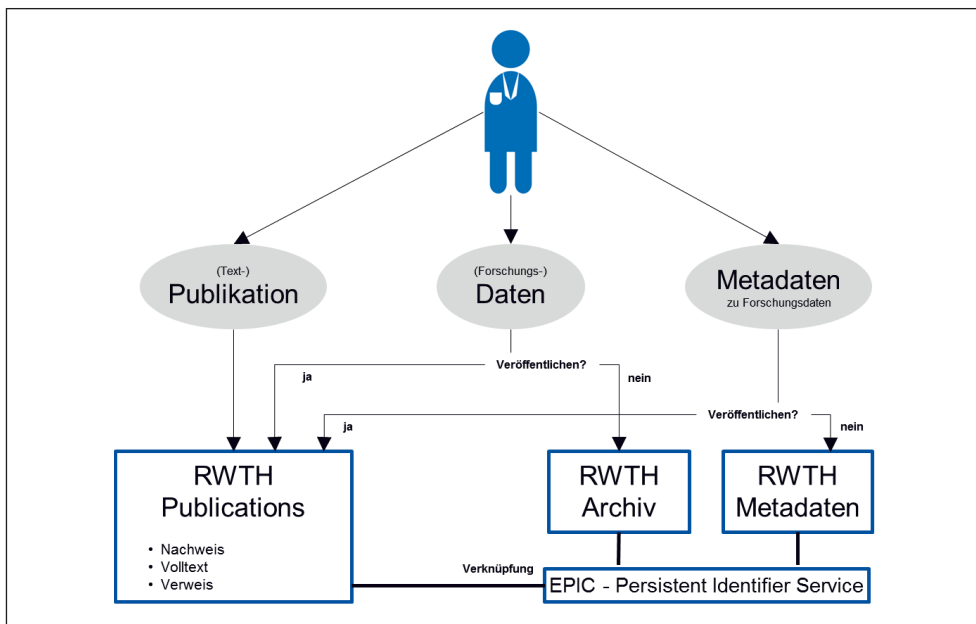


Abb. 2: Nachweis wissenschaftlicher Publikationen und zugrundeliegender Daten

5. Fazit und Ausblick

Das Beispiel „SimpleArchive“ macht deutlich, dass die angebotenen Dienste durchaus zu größeren Einheiten verschmelzen können. Primär ist SimpleArchive ein web-basierter und komfortabel nutzbarer Archivdienst. Aber darüber hinaus ist im Gegensatz zum zugrundeliegenden Basis-Archiv-Dienst die Vergabe von Persistent Identifiern hier fest integriert. SimpleArchive ohne EPIC-PID gibt es nicht. Das hat vor allem Vorteile für die Nutzerinnen und Nutzer. Zugleich bietet es aber auch der RWTH insgesamt die Möglichkeit, das Level der Verbindlichkeit zu erhöhen. Wenn absehbar wird, dass zentral auch Metadaten zu jedem archivierten Datensatz gesammelt werden müssen, kann eine entsprechende Komponente, z.B. das Metadatentool, ebenfalls an das Angebot SimpleArchive gekoppelt werden und so die Erfüllung der globalen Anforderungen sicherstellen. Gleichzeitig ist darauf zu achten, dass die Flexibilität bestehen bleibt. Wenn ein Fachbereich wie z.B. die Medizin aus zwingenden Gründen eine eigene Infrastruktur für die Verwaltung von Metadaten betreiben muss,

28 Politze, Marius; Schaffert, Steffen; Decker, Bernd: A secure infrastructure for mobile blended learning applications, in: European Journal of Higher Education IT, 1, 2016. Online: <http://www.eunis.org/download/2016/EUNIS2016_paper_19.pdf>, Stand: 01.07.2018.

dann soll das weiter möglich sein. Ähnliches gilt für die Einbettung eines Bausteins für die Erstellung von Datenmanagementplänen.

Schon in seinem Länderbericht 2017 stellt der Rat für Informationsinfrastrukturen heraus:

„Wissenschaftliche Informationsinfrastrukturvorhaben müssen auf Dynamik eingestellt sein, um nicht zu scheitern.“²⁹

In der aktuellen Situation mit einer gerade erst in der Entstehung befindlichen nationalen Forschungsdaten-Infrastruktur gilt das erst recht. Hier liegen aber zugleich auch große Chancen. Denn wie in diesem Beitrag ausgeführt, gibt es hinsichtlich der realen Inhomogenität und Dezentralität keinen Unterschied zwischen lokalen und (inter-)nationalen Strukturen. Ein Lösungsansatz, der diese Spezifika unmittelbar in den Blick nimmt, kann Angebote sinnvoll verknüpfen und von Entwicklungen auf allen Ebenen profitieren. So wurde bei der Vorstellung des bausteinbasierten Lösungskonzepts auf Basis von technologieunabhängigen, prozessorientierten Schnittstellen auf den erforderlichen Mehraufwand zur Erfassung der fachspezifischen Prozesse hingewiesen. Umgekehrt kann es aber sein, dass die erforderlichen Beschreibungen für fachspezifische Prozesse in (inter-)nationalen Strukturen ohnehin verstärkt in den Blick genommen werden und daher als Ausgangspunkt für die hier erforderliche Analyse zur Verfügung stehen. Da von vorn herein von einer heterogenen Systemlandschaft ausgegangen wurde, macht es konzeptionell keinen Unterschied, ob die Lösungskomponenten lokal oder (inter-)national betrieben werden, sodass auch über Einrichtungs- und Ländergrenzen hinweg fließend gewechselt werden kann. Die technische Realisierbarkeit ist dann nur noch eine Frage der Schnittstellen. Im besten Fall ist eine nahtlose Integration in die gerade erst im Entstehen begriffene nationale Forschungsdaten-Infrastruktur so bestens vorbereitet.

Literaturverzeichnis

- Buurmann, Merret: Tutorial Session PIDs in der Forschung. Präsentation auf dem RDA-Deutschland-Treffen 2015, Potsdam 25./26.11.2015. Online: <https://os.helmholtz.de/fileadmin/user_upload/os.helmholtz.de/Workshops/rda_de_15_buurman.pdf>, Stand: 01.07.2018.
- Curdt, Constanze; Hoffmeister, Dirk; Jekel, Christian u.a.: Implementation of a centralized data management system for the CRC Transregio 32 „Patterns in Soil-Vegetation-Atmosphere-Systems“ in: Curdt, Constanze; Bareth, Georg (Hg.): Proceedings of the Data Management Workshop, 29.-30.10.2009, University of Cologne, Germany, Köln 2010, S. 27-33. (Kölner Geographische Arbeiten 90.)

29 Vgl. Rat für Informationsinfrastrukturen (RfII): Entwicklung von Forschungsdateninfrastrukturen im internationalen Vergleich: Bericht und Anregungen, Göttingen 2017. S. 2, Online: <<http://www.rfii.de/download/rfii-fachbericht-laenderanalysen-2017/>>, Stand: 01.07.2018.

- Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten, 2015. Online: <http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf>, Stand: 01.07.2018.
- Dumas, Marlon (Hg.): Process-aware information systems: Bridging people and software through process technology, Hoboken NJ 2005.
- Eifert, Thomas; Muckel, Stephan; Schmitz, Dominik: Introducing research data management as a service suite at RWTH Aachen University, in: Müller, Paul (Hg.): 9. DFN-Forum Kommunikationstechnologien, Bonn 2016, S. 55–64. Online: <<http://subs.emis.de/LNI/Proceedings/Proceedings257/55.pdf>>, Stand: 01.07.2018.
- European Commission: EOSC Declaration: European Open Science Cloud. New Research & Innovation Opportunities, Online: <https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf#view=fit&pagemode=none>, Stand: 01.07.2018.
- Juling, Wilfried: Vom Rechnernetz zu e-Science, in: PIK - Praxis der Informationsverarbeitung und Kommunikation, 32, 2009, S. 33-36. Online: <<https://doi.org/10.1515/piko.2009.007>>, Stand: 01.07.2018.
- Kálmán, Tibor; Kurzawe, Daniel; Schwardmann, Ulrich: European Persistent Identifier Consortium – PIDs für die Wissenschaft, in: Altenhöner, Reingard; Oellers, Claudia (Hg.): Langzeitarchivierung von Forschungsdaten: Standards und disziplinspezifische Lösungen, Berlin 2012, S. 151-164.
- Kirsten, Toralf; Kiel, Alexander; Wagner, Jonas u.a.: Selecting, Packaging, and Granting Access for Sharing Study Data, in: Eibl, Maximilian; Gaedke, Martin (Hg.): INFORMATIK 2017: Digitale Kulturen. Beiträge der 47. Jahrestagung der Gesellschaft für Informatik e.V. (GI), S. 1381-1392.
- Klar, Jochen; Enke, Harry: DFG-Projekt RADIESCHEN - Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur, Report „Organisation und Struktur“, 2013. Online: <http://dx.doi.org/10.2312/RADIESCHEN_005>, Stand: 01.07.2018.
- Kraft, Angelina; Razum, Matthias; Potthoff, Jan u.a.: The RADAR Project—A Service for Research Data Archival and Publication, in: ISPRS International Journal of Geo-Information, 5, 2016, Nr. 12 . Online: <<https://dx.doi.org/10.3390/ijgi5030028>>, Stand: 01.07.2018.
- nestor-Arbeitsgruppe Zertifizierung: Erläuterungen zum nestor-Siegel für vertrauenswürdige digitale Langzeitarchive - Version 2.0, Frankfurt am Main 2016 (nestor-materialien 17,2). Online: <<http://nbn-resolving.de/urn:nbn:de:0008-2016111106>>, Stand: 01.07.2018.

- Politze, Marius; Krämer, Florian: simpleArchive – Making an Archive Accessible to the User, in: EUNIS 2017 – Shaping the Digital Future of Universities, Proceedings of the 23rd EUNIS Congress. Münster 2017, S. 121-123. Online: <http://www.eunis.org/download/2017/EUNIS2017_Book-of_Proceedings_1.pdf>, Stand: 01.07.2018.
- Politze, Marius; Schaffert, Steffen; Decker, Bernd: A secure infrastructure for mobile blended learning applications, in: European Journal of Higher Education IT, 1, 2016. Online: <http://www.eunis.org/download/2016/EUNIS2016_paper_19.pdf>, Stand: 01.07.2018.
- Rat für Informationsinfrastrukturen (RfII): Entwicklung von Forschungsdateninfrastrukturen im internationalen Vergleich: Bericht und Anregungen, Göttingen 2017. Online: <<http://www.rfii.de/download/rfii-fachbericht-laenderanalysen-2017/>>, Stand: 01.07.2018.
- Rat für Informationsinfrastrukturen (RfII): Leistung aus Vielfalt. Empfehlung zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Göttingen 2016. Online: <<http://nbn-resolving.de/urn:nbn:de:101:1-201606229098>>, Stand: 01.07.2018.
- Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan u.a.: The FAIR Guiding Principles for scientific data management and stewardship, in: Scientific data 3 (160018), 2016. Online: <<https://doi.org/10.1038/sdata.2016.18>>, Stand: 01.07.2018.
- Wittenburg, Peter: Data Foundation & Terminology WG Data Fabric IG. Präsentation auf dem RDA-DE-DINI Workshop “Aktuelle Resultate der Research Data Alliance (RDA) und deren zukünftige Bedeutung” 2015-05-28/29 in Karlsruhe, Deutschland. Online: <http://www.forschungsdaten.org/images/f/fc/RDA-DE-2015_Wittenburg_Peter_III.pdf>, Stand: 01.07.2018.
- Wolff, Eberhard: Microservices: Grundlagen flexibler Softwarearchitekturen, 1., korr. Nachdr. Heidelberg 2016.