

Integriertes Management und Publikation von wissenschaftlichen Artikeln, Software und Forschungsdaten am Helmholtz-Zentrum Dresden-Rossendorf (HZDR)

Edith Reschke¹, Dr. Uwe Konrad²

Helmholtz-Zentrum Dresden-Rossendorf

Zusammenfassung

Mit dem Ziel, das Publizieren von Artikeln, Forschungsdaten und wissenschaftlicher Software gemäß den FAIR-Prinzipien³ zu unterstützen, wurde am HZDR ein integriertes Publikationsmanagement aufgebaut. Insbesondere Daten- und Softwarepublikationen erfordern die Entwicklung bedarfsgerechter organisatorischer und technischer Strukturen ergänzend zu bereits sehr gut funktionierenden Services im Publikationsmanagement. In der Zusammenarbeit mit Wissenschaftlern des HZDR und internationalen Partnern in ausgewählten Projekten wurde der Bedarf an Unterstützung im Forschungsdatenmanagement analysiert. Darauf aufbauend wurde schrittweise ein integriertes System von Infrastrukturen und Services entwickelt und bereitgestellt. In einer seit Mai 2018 gültigen Data Policy wurden die Rahmenbedingungen und Regelungen sowohl für wissenschaftliche Mitarbeiter als auch für externe Messgäste definiert.

Zusammenfassend werden unsere Erfahrungen im integrierten Publikationsmanagement für Artikel, Forschungsdaten und Forschungssoftware vorgestellt. Es wird ein Ausblick auf die nächsten Schritte und Aufgaben gegeben und Aspekte der Integration im Kontext der europäischen und nationalen Forschungsorganisationen herausgearbeitet.

Abstract

With the aim of supporting the publication of articles, research data and scientific software in accordance with the FAIR principles, an integrated publication management system was established at the HZDR. In particular, data and software publications require the development of needs-based organizational and technical structures in addition to already well-functioning services in publication management.

¹ <https://orcid.org/0000-0002-9112-3326>

² <https://orcid.org/0000-0001-8167-9411>

³ <https://www.go-fair.org/fair-principles/>; 12.04.2019

In cooperation with HZDR scientists and international partners in selected projects, the demand for the support of research data management was analyzed. Building on this, an integrated system of infrastructures and services was gradually developed and made available. In a data policy valid since May 2018, the frameworks and regulations were defined for both scientific staff and external visitors.

In summary, our experiences in integrated publication management for articles, research data and research software are presented. It provides an outlook on the next steps and tasks and identifies aspects of integration in the context of European and national research organizations.

1 Rahmenbedingungen am HZDR – Ausgangssituation, Besonderheiten

Das Helmholtz-Zentrum Dresden-Rossendorf⁴ (HZDR) gehört zur Helmholtz-Gemeinschaft Deutscher Forschungszentren und hat neben dem Hauptstandort Dresden weitere Standorte in Freiberg, Leipzig, Schenefeld (bei Hamburg) und Grenoble (Frankreich). Im HZDR wird Spitzenforschung in den Forschungsbereichen Energie, Gesundheit und Materie betrieben. Die Zusammenarbeit mit den Partnerinstitutionen und die Arbeit in internationalen Forschungsprojekten sind am HZDR ein wesentlicher Bestandteil wissenschaftlicher Forschung. In den Labors der acht Forschungsinstitute und an den sieben Forschungsgroßgeräten, zum Beispiel am ELBE-Zentrum für Hochleistungs-Strahlungsquellen und am Hochfeld-Magnetlabor Dresden, haben HZDR-Wissenschaftler⁵ und externe Nutzer einzigartige Arbeitsbedingungen und Experimentiermöglichkeiten.

Seit 2011 haben wir am HZDR die Regelung zur „Sicherung guter wissenschaftlicher Praxis und Verfahren zum Umgang mit wissenschaftlichem Fehlverhalten“. Darin ist u.a. definiert, für welche Forschungsdaten die Archivierung Pflicht ist. Seitdem wird auch erfasst, wo und auf welche Art Forschungsdaten archiviert werden, die Grundlage einer Publikation sind. Wie in Abbildung 1 zu sehen ist, steigt der Anteil der zu archivierten Forschungsdaten seitdem kontinuierlich an und betrug zuletzt über 43%.

⁴ <https://www.hzdr.de/>

⁵ 2018 ca. 500 Wissenschaftler aus >60 Ländern

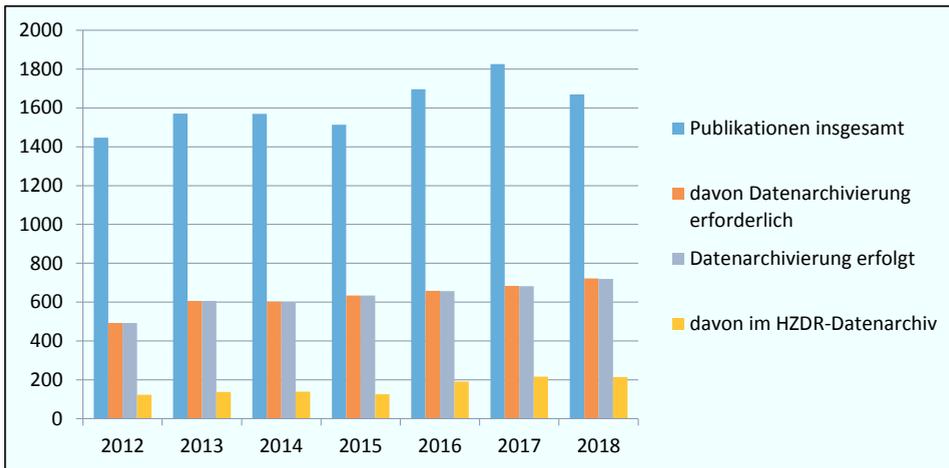


Abbildung 1: Forschungsdatenarchivierung am HZDR 2012-2018

Neben der Archivierung im HZDR-Datenarchiv wurden die Daten oftmals sehr individuell, in externen Repositorien und auch auf persönlichen „Lieblingsplätzen“ archiviert. Nicht immer waren die Langzeitverfügbarkeit und die Wiederauffindbarkeit gewährleistet. Ein wesentlicher Grund dafür waren u.a. auch das Fehlen eines Datenrepositoriums und einer Data Policy im HZDR.

Mit dem Ziel, Forschungsdaten und Forschungssoftware gem. den FAIR-Prinzipien und Open Access zu publizieren, zu archivieren, wieder auffindbar und nachnutzbar zu machen, haben wir unsere Publikationsinfrastruktur um das Datenrepositorium RODARE erweitert, welches auf dem am CERN entwickelten Invenio basiert.

2 Integriertes Publikationsmanagement am HZDR

2.1 Policies und Prozesse

Wissenschaftliche Reputation basiert zu einem bedeutenden Teil auf wissenschaftlichen Publikationen. Publiziert werden Artikel (Texte), Forschungsdaten und Forschungssoftware, wobei Forschungsdaten und Forschungssoftware erst in neuerer Zeit in den Publikationsfokus gerückt sind. Ein wichtiger Aspekt dabei sind die sich schnell verändernden rechtlichen Rahmenbedingungen, die mittelbar den Publikationsprozess beeinflussen. Dazu sind u.a. das Urheberrecht und der Datenschutz zu zählen. Die Internationalität und Vernetzung der Forschung und des Publikationsmarktes sind weitere Faktoren, die es zu berücksichtigen gibt.

Wissenschaftliches Publizieren gemäß den FAIR-Prinzipien erfordert ein effizientes Publikationsmanagement, das drei Komponenten integriert:

- die Bereitstellung der notwendigen technischen Infrastrukturen,
- Services für alle Publikationen (Text, Daten, Software) und
- eine passende Struktur für die Unterstützung und Beratung der Wissenschaftler.

Das integrierte Publikationssystem muss flexibel angepasst werden können und die Akzeptanz durch den Wissenschaftler inkl. der Gäste finden.

Jedes Publikationsmanagement unterliegt natürlich rechtlichen Rahmenbedingungen. Ergänzend zu den allgemein gültigen Gesetzen (Urheberrecht, Datenschutz, u.a.) wurden HZDR-internen Regelungen erstellt:

- „Sicherung guter wissenschaftlicher Praxis und Verfahren zum Umgang mit wissenschaftlichem Fehlverhalten“
- „Publikationsordnung“
- „Terms and Conditions for the Storage, Access and Curation of Research Data“ (HZDR Data Policy)
- “Terms and Conditions for User Access to the Experimental Facilities”

Diese Regelungen sind eine solide Grundlage für die Entwicklung eines rechtssicheren integrierten Publikationsmanagements.

Verantwortlich für die Entwicklung und den Aufbau eines Publikationsmanagements am HZDR ist die Zentralabteilung Informationsdienste und Computing, zu der auch die Bibliothek gehört. Beteiligt sind natürlich weitere Abteilungen wie die wissenschaftliche Infrastruktur, die Stabsabteilung Recht und Patente und die Stabsabteilung Programmplanung und Internationale Projekte.



Abbildung 2: Bereiche, die ins Publikationsmanagement integriert werden müssen

2.2 Integrierte Systemlandschaft

Die Publikationsdatenbank (ROBIS) am HZDR gibt es seit 1993. Bis Mitte 2018 war sie für folgende Services verfügbar:

- Nachweis aller Publikationen seit 1993
- Verwertungsprüfung
- Open Access Finanzierung
- Genehmigungsworkflow
- Erstellung von institutionellen und persönlichen Publikationslisten
- Statistik, wissenschaftliches Controlling
- Langzeitarchivierung der elektronischen Volltexte
- Verlinkung von Publikationen zur Datenarchivierung

ROBIS ist eine Eigenentwicklung, basierend auf der Nutzung einer Oracle-Datenbank mit Weboberfläche, die für Publikationen im Laufe der Zeit immer wieder an den lokalen Bedarf angepasst und weiterentwickelt wurde. Da die Anforderungen an ein Datenrepositorium insbesondere unter technischem Aspekt sich nicht mit einer

Erweiterung von ROBIS abdecken ließen, wurde entschieden, ein separates Datenrepositorium aufzubauen.

In Interaktion mit ROBIS sollen alle notwendigen Publikationsservices für Textpublikationen, Forschungsdaten und Forschungssoftware bereitgestellt werden.

Auf der Basis der Invenio⁶ Framework library wurde das Rossendorf Data Repository RODARE entwickelt. RODARE ist auf GitLab⁷ eingestellt und steht unter der GNU General Public License by the Free Software Foundation (GPLv3) zur Nachnutzung zur Verfügung. Seit Mitte 2018 ist die Verlinkung der Workflows des Publikationsrepositorium ROBIS und des Forschungsdatenrepositoriums RODARE aktiv geschaltet (siehe Abb. 3). Die Autoren können selbst entscheiden, in welchem Repository sie mit ihrer Metadatenerfassung und Datenspeicherung beginnen. Die beiden Repositorien übergeben die „Kern“- Metadaten automatisch an das jeweils andere Repository, so dass doppeltes Datenerfassen ausgeschlossen wird. Es bleibt die Notwendigkeit, dass die Autoren die Repository-spezifischen Einträge anschließend vornehmen müssen.

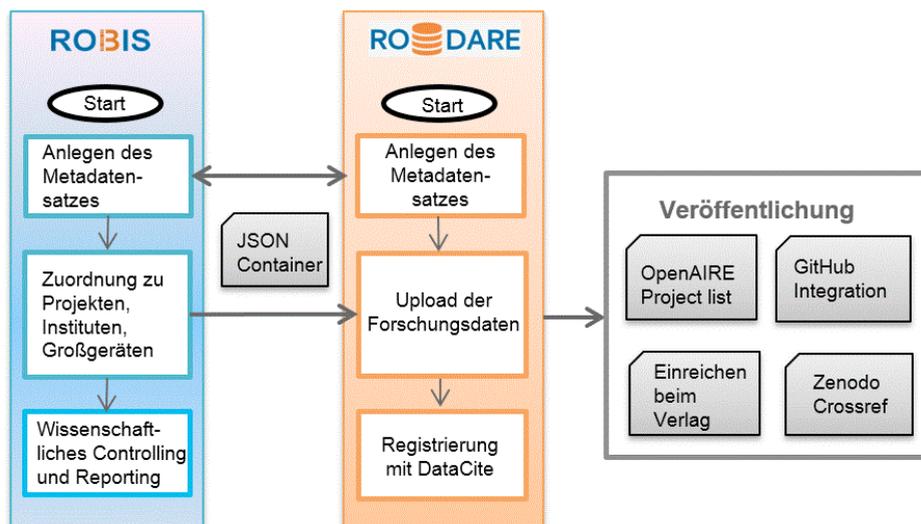


Abbildung 3: Interaktionen zwischen ROBIS und RODARE

⁶ <https://inveniosoftware.org/> - 09.04.2019

⁷ <https://gitlab.hzdr.de/rodare/rodare> - 09.04.2019

2.3 *Daten und Software*

Daten und Software gehören seit jeher zum wissenschaftlichen Publikationsprozess. Nur führten sie lange Zeit ein Schattendasein. Mit der Forderung nach guter wissenschaftlicher Arbeit rückte das Publizieren von Forschungsdaten und wissenschaftlicher Software in seiner Bedeutung auf zu den wissenschaftlichen Textpublikationen. Forschungsdaten und Forschungssoftware sind sehr oft die Grundlage einer Publikation. Sie liegen daher viel eher vor als die Textpublikation oder auch als eigenständige Publikation. Deshalb war die Entwicklung eines Datenrepositoriums zusätzlich zu unserem Publikationsrepositorium notwendig. Forschungsdaten und Software bedingen spezifische Metadaten, die sich nur zum Teil mit den Metadaten in ROBIS decken. In RODARE wurden zusätzliche Funktionen gebraucht; zum Beispiel der Umgang mit großen bzw. komplexen Datenstrukturen, die automatische Registrierung einer DOI, APIs zu OpenAIRE, GitHub bzw. GitLab, die Zuordnung von Lizenzen u.a.m.

Um die im Punkt 2.2. dargestellten Funktionen zu erfüllen, ist es notwendig, dass alle Publikationen, also Textpublikationen, Forschungsdaten und Forschungssoftware im Publikationsrepositorium ROBIS nachgewiesen werden. Mit der „Verzahnung“ von ROBIS und RODARE haben wir es vermeiden können, dass die Autoren ihre Daten mehrfach eingeben müssen.

2.4 *Publikationssupport aus einer Hand*

Die Bereitstellung von Publikationsinfrastruktur und zugehörigen Services sind Bausteine im Gesamtkonzept „Wissenschaftliches Publizieren“. Akteure sind insbesondere die wissenschaftlichen Institute (als Betreiber der Forschungsgeräte), die Abteilung Informationsdienste und Computing und die Bibliothek. Die Koordinierung der publikationsspezifischen, abteilungsübergreifenden Aktivitäten aller Akteure ist die Kernaufgabe des Data Librarian in der Bibliothek. Der Data Librarian ist für folgende Services zuständig:

- Anwenderberatung zu ROBIS und RODARE
- Unterstützung für die Erstellung und Pflege der Daten-Management-Pläne
- Qualitätssicherung und Kuratierung der Metadaten
- Beratung zu Open Access und Lizenzierungsoptionen.

Entscheidend, ob diese Angebote von den potentiellen Autoren angenommen werden oder nicht, sind die Qualität der Information, Beratung und Unterstützung, idealerweise projektbegleitend vom Projektstart bis zur Veröffentlichung der Publikationen.

2.5 Weiterentwicklung von Services im digitalen Forschungsprozess

Entscheidend für die Akzeptanz der Prozesse und Infrastrukturen durch die Nutzer ist eine optimale Einbettung in die tägliche Arbeit und die Vermeidung einer aufwändigen Eingabe und Pflege von Zusatzdaten. Dies trifft ganz besonders für den Lebenszyklus von Daten- und Softwarepublikationen zu, da diese oft in komplexen Strukturen vorliegen, in größeren Teams über einen langen Zeitraum bearbeitet werden und der Aufwand für Reproduzierbarkeit sowie Nachnutzbarkeit wesentlich größer als für Textpublikationen ist.

Deshalb ist es erforderlich, Informationen und Metadaten frühzeitig, kontinuierlich und systematisch in einem durchgängigen digitalen Forschungsprozess zu erfassen. Dieser beginnt bereits mit der Anmeldung eines Experimentes / Projektes für ein wissenschaftliches Experiment, z.B. an einem Forschungsgerät. Am HZDR erfolgt diese Anmeldung über das System GATE, über das der Nutzer per Webportal den Zugang beantragt und nach einem Reviewprozess im Erfolgsfall bekommt. Bereits in dieser frühen Phase wird die Data Policy verbindlich eingeführt und es werden viele Metadaten abgefragt, die für die Daten-Management-Pläne und schließlich für alle Publikationsarten von Interesse sind und automatisiert verarbeitet werden können.

Während der Zeit der wissenschaftlichen Experimente spielen digitale Laborbücher für Forschungsdaten eine wichtige Rolle. In diesen erfassen die Forscher Informationen und Bilder zu verwendeten Geräten, Proben, Einstellungen und besonderen Vorkommnissen. Werden diese Daten strukturiert erfasst und gespeichert, können sie später die grundlegenden Metadaten (z.B. nach DataCite Schema 4.x) ergänzen und sachlich erschließbar machen. Am HZDR wird dafür derzeit in Pilotprojekten das System openBIS (Quelle ETH Zürich) eingesetzt.

Für Forschungssoftware sind die Entwicklungsumgebungen die Plattform kontinuierlicher Änderungen und Tests. Am HZDR wird den Entwicklern dafür zukünftig das lokale System GitLab angeboten, das seinerseits mit der führenden Entwicklungsplattform GitHub (betrieben von Microsoft) vernetzt ist. Von diesen Systemen kann die Publikation direkt in das Publikationssystem RODARE erfolgen, so dass dem Softwareentwickler keine großen zusätzlichen Aufwände entstehen.

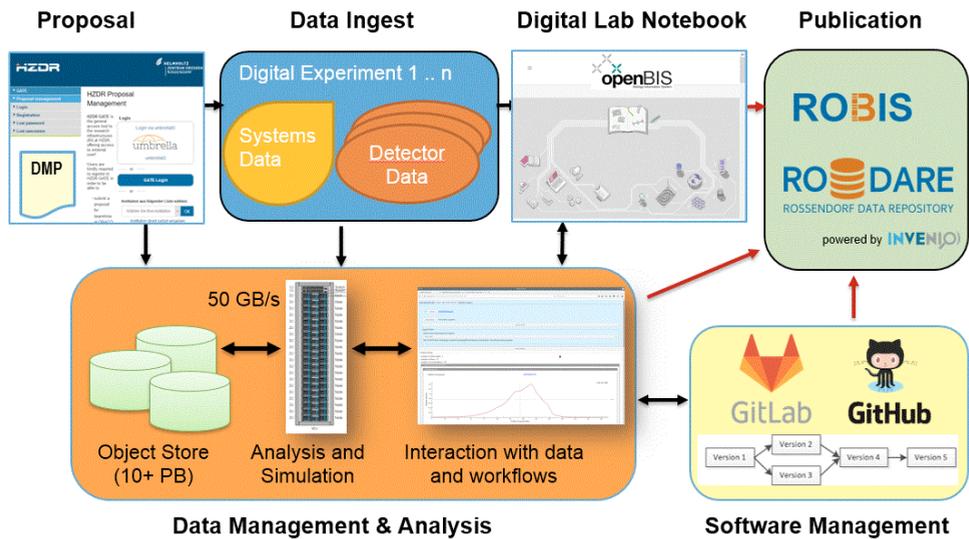


Abbildung 4: Daten und Software im Lebenszyklus am HZDR

In Abbildung 4 ist gut zu erkennen, dass man im Publikationsprozess den gesamten Lebenszyklus aller Arten von Publikationen mit allen beteiligten Partnern betrachten muss, um eine Qualität im Sinne von FAIR zu erreichen. Sobald man eine Komponente vernachlässigt hat der publizierende Wissenschaftler erhebliche Aufwände zur Eingabe und Pflege von Metadaten und wird Argumente finden, dies nicht zu tun. Man kann die Qualität von textuellen Veröffentlichungen noch vor der Publikation verbessern, bei Daten und Software ist es am Ende oft zu spät.

Einen großen Einfluss auf die Nachhaltigkeit der Ergebnisse haben auch Metadaten-Standards, die von Geräteherstellern, der Industrie und der Wissenschaft selbst entwickelt werden. An dieser Stelle spielen insbesondere die Research Data Alliance (RDA) und die geplanten community-spezifischen Instanzen der Nationalen Forschungsdaten-Infrastruktur (NFDI) in Deutschland eine besondere Rolle. Hier haben die führenden Wissenschaftler und großen Daten-Infrastrukturen der jeweiligen Communities eine große Verantwortung zur Vereinheitlichung von Prozessen und Metadaten sowie zur Aus- und Weiterbildung der Forscher.

Die Helmholtz-Gemeinschaft trägt dieser Verantwortung mit dem „Helmholtz Information & Data Science Framework“ Rechnung, in dem Konzepte zur gemeinschaftsweiten Zusammenarbeit entwickelt werden.

Beginnend 2019 werden 5 Serviceplattformen eingerichtet, die diese Prozesse in allen Zentren vorantreiben und unterstützen werden⁸. Das sind:

- die „Helmholtz Data Science Academy (HIDA)
- die “Helmholtz Federated IT Services (HIFIS)“
- das Helmholtz Metadata Center (HMC)
- die Helmholtz Artificial Intelligence Coordination Unit (HAICU)
- die Helmholtz Imaging Plattform (HIP)

Speziell die ersten drei genannten Plattformen sind darauf ausgerichtet, die Prozesse der Aus- und Weiterbildung, der Softwareentwicklung und der Metadatendefinition und -erfassung zu entwickeln und zu vereinheitlichen.

Fazit

Um wissenschaftliche Artikel, Forschungsdaten und Forschungssoftware nach den FAIR Prinzipien publizieren zu können, muss der digitale Forschungsprozess als Ganzes betrachtet und optimiert werden. Durch die nahtlose Integration der Teilprozesse und IT-Infrastrukturen vom Antrag eines Forschungsvorhabens bis zur Langzeitarchivierung kann der publizierende Wissenschaftler von aufwändigen Metadateneingaben entlastet und gleichzeitig die Qualität und Nachhaltigkeit der Ergebnisse gesichert werden.

Diese ganzheitliche Betrachtungsweise erfordert ein enges Zusammenwirken der beteiligten Servicebereiche; zum Teil müssen die klassischen Aufgabenzuordnungen überwunden werden. Am HZDR wurde dieser Prozess 2015 mit der Fusion des Rechenzentrums mit der Bibliothek begonnen und seitdem konsequent umgesetzt. Die technische Integration der Systeme (GATE, Invenio, openBIS, GitHub, GitLab u.a.) ist ein erheblicher Aufwand, bedeutet aber für den Nutzer einen spürbaren Mehrwert. Die positiven Rückmeldungen aus der Wissenschaft bestätigen uns, dass wir auf dem richtigen Weg sind.

⁸ https://www.helmholtz.de/forschung/information_data_science/helmholtz_inkubator/ - 01.04.2019

Referenzen

Barillari, C. (2018): openBIS – an open resource for academic laboratories. Vortrag, [https://os.helmholtz.de/fileadmin/user_upload/os.helmholtz.de/Workshops/elN18hzi_barillari.pdf - 15.04.2019](https://os.helmholtz.de/fileadmin/user_upload/os.helmholtz.de/Workshops/elN18hzi_barillari.pdf-15.04.2019)

Konrad, U.; Görzig, H.; Juckeland, G. (2018): Forschungsdatenmanagement am Helmholtz-Zentrum Dresden-Rossendorf und am Helmholtz-Zentrum Berlin (RDM@DB). Konferenzbeitrag, DOI: [10.14278/rodare.62](https://doi.org/10.14278/rodare.62)

Konrad, U. (2017): Research Data Management to increase research quality. Konferenzbeitrag, DOI: [10.5281/zenodo.1040289](https://doi.org/10.5281/zenodo.1040289)

Konrad, U. (2017): Wissenschaftliche Software – Anspruch und Realität im Forschungsprozess. Konferenzbeitrag, [10.5281/zenodo.1040289](https://doi.org/10.5281/zenodo.1040289)