

5. RADAR-Workshop

(27./28. November 2019, Halle/Saale)

Zusammenfassung

Der 5. RADAR-Workshop fand am 27. und 28. November 2019 am Leibniz-Institut für Pflanzenbiochemie (IPB) in Halle an der Saale statt. Die 19 Teilnehmenden umfassten Vertreterinnen und Vertreter aus Hochschulen, wissenschaftlichen Forschungseinrichtungen und Initiativen aus ganz Deutschland.

Der erste Tag richtete sich mit der grundlegenden Einführung in das Forschungsdatenrepository RADAR (www.radar-service.eu), einer Live-Demonstration des Systems, Informationen zu Preisen und Verträgen sowie einem Vortrag zu rechtlichen Aspekten rund um RADAR und FDM vor allem an Interessenten und Neueinsteiger. RADAR bietet akademischen Institutionen und Forschenden ein umfangreiches, disziplinübergreifendes und verlässliches Angebot zur Archivierung und Veröffentlichung digitaler Forschungsdaten. Der zweite Tag legte den Schwerpunkt auf den Erfahrungsaustausch und die gemeinsame Planung der zukünftigen funktionalen Weiterentwicklung von RADAR. Neben einem Rückblick auf die letzten Releases, einem Ausblick auf das anstehende Release und der Vorstellung und gemeinsamen Diskussion des RADAR Product Backlogs boten sechs Teilnehmerbeiträge zur FDM-Praxis an den jeweiligen Institutionen Gelegenheit zum gegenseitigen Austausch. Ein Update zur NFDI und der Rolle generischer Repositorien in diesem Kontext bildete den Abschluss des Workshops.

Im Anschluss an den öffentlichen Workshop tagte erstmalig auch der Nutzerinnen- und Nutzerbeirat, der sich aktuell aus Vertreterinnen und Vertretern von fünf nutzenden Institutionen zusammensetzt. Der Beirat wird FIZ Karlsruhe zukünftig bei der Planung der RADAR Product Roadmap sowie der Identifizierung und Priorisierung zukünftiger Funktionalitäten beraten.

Nachfolgend eine Zusammenfassung der einzelnen Workshop-Beiträge:

Einführung in RADAR und grundlegende Konzepte

Der Vortrag gab einen Einblick über die Entstehung von RADAR im Rahmen des DFG-Projekts (2013-2016) unter dem Motto „Aus der Community für die Community“ durch das interdisziplinäre Projektkonsortium. Weiterhin wurden das RADAR-Dienstleistungsspektrum, die grundlegenden Konzepte der Datenübernahme und -speicherung, die schematische Architektur, das generische Metadatenschema, das Rollen- und Rechtekonzept für die delegierte Administration von RADAR durch nutzende Einrichtungen und die kundenspezifischen Anpassungsmöglichkeiten des Dienstes vorgestellt.

Preise und Verträge

Im Vortrag wurden das RADAR-Geschäftsmodell, die Preisstruktur (inkl. der rabattierten Preisoption für Datenvolumina > 50 TB), der RADAR-Dienstvertrag (inkl. der juristischen Beziehungen der Akteure), die Kernregelungen des RADAR-Vertrags, die Nutzungshinweise für Datengebende / Datennutzende sowie RADAR in Bezug auf die DSGVO erläutert.

FIZ Karlsruhe verfolgt mit RADAR keine kommerziellen Interessen, vielmehr betrachtet das Institut den Dienst als Teil seines öffentlichen Auftrags. Daher übernimmt FIZ Karlsruhe als Betreiber die Hälfte der operativen Fixkosten von RADAR aus der eigenen Grundfinanzierung. Lediglich der verbleibende Fixkostenanteil und die variablen Betriebskosten werden über Nutzungsgebühren verrechnet. Trotz mehrfacher Anfragen gibt es zurzeit keine Einmalzahlung für archivierte Daten. Die in diesem Fall notwendige Einpreisung eines Risikozuschlags für die langfristige Zusage der Datenarchivierung würde voraussichtlich zu unattraktiven Konditionen führen. Die Einmalzahlung würde auch eine Bindung über die gesamte Haltefrist bedingen, was zulasten der Flexibilität einer Institution ginge.

Abschließend wurde ein Ausblick auf die alternativen Betriebsmodelle gegeben, die zukünftig zusätzlich neben dem Betrieb von RADAR als vollständiger Cloud-Service angeboten werden. Im Betriebsmodell „RADAR Hybrid“ (in Arbeit für 2020) wird der permanente Speicher vom Kunden gestellt, beim Betriebsmodell „RADAR Local“ (in Planung für 2021) stellen die Kunden sowohl den permanenten Speicher als auch die Ablaufumgebung. In beiden Fällen wird die RADAR-Software weiterhin von FIZ Karlsruhe betrieben und gepflegt. Die neuen Betriebsmodell-Alternativen werden über Pauschalpreise ohne variable Kostenanteile abgerechnet.

Rechtliche Aspekte rund um RADAR und FDM

Der Vortrag hatte die rechtlichen Anforderungen an die Veröffentlichung von Forschungsdaten zum Schwerpunkt und nahm dabei vor allem die Perspektive der Datengebenden ein. Dabei wurde auch ein Entscheidungsbaum¹ von [forschungsdaten.info](https://www.forschungsdaten.info) vorgestellt. Zur Lizenzvergabe für Forschungsdaten kann aktuell in RADAR zwischen Creative Commons-Lizenzen, einer proprietären Lizenz und der Option „Alle Rechte vorbehalten“ gewählt werden. Im Sinn von FAIR-Data und Nachnutzbarkeit sollten im Regelfall keine (zu) restriktiven Lizenzen vergeben werden. Bei der Vergabe einer restriktiven Lizenz bei nicht schutzfähigen Daten besteht die Gefahr der Schutzrechtsberührung zu Unrecht (Copyfraud). Dieses Risiko lässt sich z.B. durch die Verwendung von CCO vermeiden. RADAR wird für solche Daten mit dem nächsten Release zusätzlich die „Public Domain Mark 1.0“ als weitere Option ergänzen.

Rückblick auf bisherige RADAR Releases / Vorstellung der neuesten Erweiterungen

Im Vortrag wurde die Chronologie der bisherigen RADAR-Releases aufgezeigt, die kontinuierliche Weiterentwicklung des Dienstes bis zur aktuellen Version 1.6 erläutert und die neuen Funktionalitäten vorgestellt. Das aktuelle Softwarerelease zielt auf die bessere Einbindung und Anpassbarkeit des Dienstes an die lokalen Gegebenheiten und Arbeitsabläufe bei nutzenden Einrichtungen. Es bietet ein optimiertes Dashboard zur zentralen Verwaltung der institutionellen Einstellungen und weitere Möglichkeiten zur Angleichung der Benutzungsoberfläche an das eigene Corporate Design (z.B. Schmuckfarbe). Für publizierte Forschungsdatensätze lassen sich nun DOIs mit einem institutseigenen Präfix vergeben. Neue Funktionen zur Datenintegritätsprüfung und zum Export von Zitationen runden das Softwarerelease ab.

¹ Siehe <https://doi.org/10.5281/zenodo.3368293>

Abschließend erfolgte ein Ausblick auf die kommenden Meilensteine: die Einführung weiterer Betriebsmodelle, in denen kundeneigene Infrastruktur genutzt werden kann (siehe auch “Preise und Verträge”) und noch weitergehende kundenspezifische Anpassungsmöglichkeiten von RADAR, etwa einer institutionellen Suche/Sicht auf Datenpakete oder die Verwendung einer eigenen Domain. Dazu wurde das Szenario von RADAR4KIT vorgestellt, das auf diesen Meilensteinen aufsetzen wird. Die neuen Features werden mit dem nächsten Release 1.7 allen nutzenden Institutionen zur Verfügung stehen.

Vorstellung des RADAR Product Backlogs / Diskussion zukünftiger Erweiterungen / Anforderungen aus der Community

Das RADAR-Team sammelt eigene Ideen und Anforderungen aus der Community in einem so genannten Product Backlog. Die darin enthaltenen Feature-Beschreibungen lassen sich grob in folgende Kategorien clustern: weitere kundenspezifische Einstellungen, disziplinspezifische Erweiterungen des generischen Dienstes, vereinfachte Benutzung für Forschende und die verbesserte Unterstützung der FAIR-Kriterien.

Die Weiterentwicklung von RADAR soll dabei so bedarfsorientiert wie möglich im engen Austausch mit nutzenden Einrichtungen erfolgen. Im Vortrag wurde eine Vielzahl möglicher Ansatzpunkte für die Weiterentwicklung von RADAR aus unterschiedlichen Themenfeldern vorgestellt: Authentifizierung, Rollen- und Rechtemodell, Metadaten, Datenupload/-download, Daten im temporären Speicher („in Bearbeitung“ bzw. „in Review“), Daten im permanenten Speicher („archiviert“ bzw. „publiziert“), Benutzungsoberfläche etc. Die Teilnehmenden hatten die Möglichkeit, diese mit Hilfe eines Fragebogens unverbindlich zu bewerten und für die anschließende tiefergehende Diskussionsrunde zu priorisieren. Basierend auf der Wichtung der Teilnehmenden wurde anschließend über folgende vier Feature-Ideen diskutiert:

Unterstützung fachspezifischer Metadaten

Die Implementierung ermöglicht es Nutzenden, neben den RADAR-Metadaten zusätzlich auch eigene, den fachspezifischen Anforderungen entsprechende Metadaten für Datenpakete und darin enthaltene Dateien und Ordner zu übermitteln, sie gegen hinterlegte Metadatenschemata zu validieren, mit den Datenpaketen zu archivieren und gezielt abzurufen. Unter den Begriff „fachspezifische“ Metadaten fallen dabei sowohl in einer Disziplin allgemein anerkannte Metadatenstandards (z.B. im Rahmen der NFDI erarbeitet), als auch projektspezifische Metadatenformate.

Zur Erleichterung des Erstellens von Metadatenschemata wird seitens der Teilnehmenden der Metadatenmanager der RWTH Aachen² genannt. Impulse für die Verwendung geeigneter Metadatenschemata könnte auch das Research Data Alliance Metadata Directory³ bieten.

RADAR ist hier bereits in konkreten Planungen. Die Funktionalität wird nach Beauftragung durch eine nutzende Einrichtung als kundenspezifische Weiterentwicklung kostenpflichtig entwickelt und anschließend in 2020 allgemein auf RADAR verfügbar gemacht.

² <https://doc.itc.rwth-aachen.de/display/FDM/MetadatenManager>

³ <https://rd-alliance.github.io/metadata-directory/standards/>

Versionierung von Forschungsdaten

In der Praxis ergibt sich regelmäßig die Notwendigkeit der Aktualisierung bereits publizierter oder archivierter Forschungsdaten (Hinzufügen weiterer Daten zu Datensammlungen, Errata fehlerhafter Daten, etc.). Da RADAR aktuell bisher keine vollwertige und einfach zu bedienende Versionierung unterstützt, ist in diesen Fällen die Verlinkung des neuen Datenpakets mit der vorherigen Version manuell über das optionale RADAR-Metadatenfeld „Verwandter Identifier“ nachzuführen. Die Verlinkung der Vorgängerversion zur aktuellen ist jedoch aufgrund deren Unveränderlichkeit nach erfolgter Archivierung bzw. Publikation bisher nicht möglich. Auch die Darstellung der Versionsgeschichte in der Benutzungsoberfläche von RADAR ist stark verbesserungsbedürftig. Schließlich soll es möglich sein, für die Gesamtheit aller Versionen eines Datenpakets eine eigene DOI zu vergeben, über die sie gemeinsam identifiziert und zitiert werden können. Das Feature „Versionierung von Forschungsdaten“ wird von allen Teilnehmenden als notwendig erachtet.

Es werden zwei Handlungsschwerpunkte festgestellt und betont, den Fokus jeweils auf die Usability zu legen: Einerseits sollte für Datengebende durch intuitive Benutzerführung einfach zu erschließen sein, dass und wie aktualisierte Versionen für Forschungsdaten hinterlegt werden können. Andererseits sollte auch für Datennehmende über bidirektionale Referenzierung klar ersichtlich sein, sobald Forschungsdaten in mehreren Versionen vorliegen und welches der aktuellste Datensatz ist.

Von der Mehrheit der Teilnehmenden wird präferiert, als ersten Schritt eine leicht zu bedienende Lösung noch in 2020 verfügbar zu machen, die die gängigsten Use Cases abdeckt: Bei Änderungsbedarf wird jeweils das komplette aktualisierte Datenpaket hochgeladen. Zur Steigerung des Nutzerkomforts werden dabei die Metadaten des bisherigen Datensatzes als Defaultwerte übernommen. Die neue Version wird eine neue DOI erhalten. Die bidirektionale Verlinkung in der Benutzungsoberfläche (und damit einhergehend die in der Datenbank hinterlegten Metadaten) wird systemseitig unterstützt.

Lösungen, die das Hochladen von Appends oder nur eines Daten-Offsets (was z.B. zur Speicherplatzvermeidung bei großen Datenvolumina vorteilhaft wäre) erlauben, könnten in den Folgejahren angedacht werden. Die Core-Funktionalität der RADAR Software und dessen einfache Benutzbarkeit sollte jedoch bei keiner Funktionserweiterung aus dem Fokus geraten.

Aktualisieren von Metadaten

Das nachträgliche Ändern von Metadaten kann aus vielerlei Gründen erforderlich sein (vergessene Eintragungen, falsche Eintragungen, Tippfehler etc.). Selbst die gängige wissenschaftliche Veröffentlichungspraxis macht das Ändern von Metadaten erforderlich, denn im Tagesgeschäft stellt sich oft das Problem, dass Verlage für die Zeitschriftenpublikation einen bereits vorab publizierten Forschungsdatensatz einfordern (d. h. die bloße Mitteilung einer vorab in RADAR reservierten DOI reicht in diesem Fall nicht aus).

Dies stellt nutzende Institutionen vor Schwierigkeiten, denn das Eintragen der DOI der Zeitschriftenpublikation im RADAR-Metadatenfeld „Verwandter Identifier“ wird zwar unterstützt, jedoch nur vor der Publikation eines Forschungsdatensatzes. Sobald veröffentlicht, ist ein Nachtragen der DOI der Zeitschriftenpublikation nicht mehr möglich. Genau dieses wird aber angesichts oben geschilderter Praxis von den Nutzenden in RADAR gewünscht. Andere Akteure wie DataCite oder Zenodo erlauben die nachträgliche Editierung von Metainformationen.

Die Teilnehmenden stimmen überein, dass RADAR die Änderung und transparente Versionierung von Metadaten ebenfalls unterstützen sollte. Diese Änderung sollte aber nur die in der Datenbank und auf der Landing-Page vorhandenen Metainformationen betreffen. Eine Änderung der im eigentlichen BagIt-Archivpaket vorliegenden Metadaten soll nicht erfolgen, da damit das gesamte Paket inklusive seiner Checksums verändert würde.

Schnittstelle zum Upload eines Datenpakets aus einem Git-Repository

Teilnehmende weisen auf die bei Zenodo mögliche Upload-Option hin und raten, eine vergleichbare Funktion auch in RADAR anzubieten. Dabei sollten nicht ausschließlich GitHub, sondern auch GitLab und lokale Git-Repositorien als Quelle berücksichtigt werden.

Einsatz von RADAR in der Praxis

Folgende Teilnehmende berichteten aus ihren eigenen Einrichtungen, die RADAR bereits nutzen oder z.B. über einen Einsatz erwägen:

- Jan Kenter, Universitäts- und Landesbibliothek Bonn
- Dr. Jürgen Rohrwild, Universitätsbibliothek der FAU Erlangen-Nürnberg
- Axel Plähn, Helmholtz-Zentrum für Infektionsforschung, Braunschweig
- Juliane Watson, Deutsches Archäologisches Institut, Berlin
- Christoph Rzymiski, Max-Planck-Institut für Menschheitsgeschichte, Jena
- Dr. Marco Gronwald, Universitätsbibliothek, Osnabrück

Aus den Vorträgen lassen sich übergreifende Erkenntnisse, Best Practices bzw. Herausforderungen ableiten:

Übergreifende Erkenntnisse:

- Die Vertreter der Universitäten berichten, dass dort z. T. bereits vor Jahren Bedarfserhebungen im Bereich FDM initiiert und durchgeführt wurden.
- Nicht alle Institutionen haben bereits eine finale Forschungsdaten-Policy, diese ist jedoch in der Regel in Arbeit.
- An den Einrichtungen, vor allem an Hochschulen, sind die verwendeten Datentypen und -volumina sowie die verwendeten Standards zur Strukturierung und Beschreibung von Forschungsdaten sehr heterogen.

Best Practices:

- Als wichtige Auswahlkriterien für die Entscheidung, RADAR als Dienst für die Archivierung und Publikation von Forschungsdaten einzusetzen, werden die vorliegende Exit-Strategie und der geringe Implementierungsaufwand genannt.
- Die Bereitstellung nicht-projektfinanzierter Personalstellen seitens der Hochschulleitungen wird als maßgeblich für den Aufbau dauerhafter Services im FDM-Bereich (z.B. Servicestellen, Referenten) erachtet.
- Neben Individualberatungen bauen Hochschulen vermehrt Schulungsangebote auf. Neben proaktiven Informationsveranstaltungen in Fachbereichen (z.B. Master-Seminaren) wird das

Thema FDM z.B. fest im Veranstaltungsprogramm für Promovenden/PostDocs verankert und z.T. durch fachspezifische Schulungen oder e-Learning Angebote ergänzt.

- Beim Thema FDM in der Antragsberatung / Drittmittelakquise bietet sich übereinstimmend eine Kooperation von Bibliothek und anderen Organisationsbereichen (z.B. Forschung, Hochschulentwicklungsplanung) an. Bereits bei Antragsstellung wird angesichts des aktualisierten DFG-Kodex zur „Guten wissenschaftlichen Praxis“ Bewusstsein für FDM aufgebaut.
- Unterstützt wird die Bewusstseinsbildung auch mit eigenen Webseiten, auf denen über die angebotenen Dienstleistungen und Infrastrukturen (wie z.B. RADAR) entlang des Forschungsprozesses informiert wird.
- RADAR ergänzt meist eine ganze Reihe von Diensten zum Forschungsdatenmanagement an den Einrichtungen und wird dort vorrangig für die Archivierung und Publikation von Daten des „Long Tails“ verwendet, für die es kein disziplinäres Datenzentrum gibt.
- Viele Administrierende nehmen auch die Rolle als Kurator bzw. Kuratorin ein.
- Zur Vermeidung des internen Verwaltungsaufwands räumen die Einrichtungen teilweise den Datengebernden ein kostenfreies Speicherkontingent ein. Andere Institutionen verzichten ganz auf die interne Umlegung der nutzungsbasierten Kosten auf die Forschenden/Lehrstühle, sondern übernehmen die Kosten für Datenarchivierung/-publikation aus dem eigenen Budget.

Herausforderungen:

- Während Auflagen durch Förderer und Verlage zum Teil starke Motivatoren für eine Datenveröffentlichung darstellen, gehen von Fachebenen/Lehrstühlen selbst noch wenige Anreize hierzu aus.
- RADAR solle Versionierung und eine transparentere Nutzerverwaltung für Administrierende anbieten, um wiederkehrende Barrieren für die Nutzung abzubauen.
- Einrichtungen, die andere Tools (z.B. zur Software-Veröffentlichung) nutzen, bemängeln z.B. das dort fehlende Rollen- und Rechtmanagement
- Zunehmende Nachfragen nach rechtlichen Aspekten machen die Beratung von Forschenden im Einzelfall schwierig.

Generische Repositorien im NFDI-Kontext

Ogleich der Schwerpunkt der NFDI auf fach- oder formatspezifischen Repositorien liegt, existieren nicht für alle Datentypen passende Lösungen bzw. ist eine Archivierung nicht immer mitgedacht. Generischen Repositorien kommt in diesen Fällen als „Catch-all-Lösung“ eine wichtige Rolle zu. Weiterhin bieten sich diese als institutionelle Lösung zum Nachweis des eigenen wissenschaftlichen Outputs an. Eine Interoperabilität mit der NFDI muss in diesem Fall in beide Richtungen unterstützt werden.

RADAR hat den Anspruch, sich als Infrastrukturbaukasten in die NFDI einzubringen. Der Dienst ist in der ersten Antragsrunde als generische Infrastruktur in vier Konsortien vertreten: MaRDI (Mathematik), NFDI4Chem (Chemie), NFDI4Culture (Kunstgeschichte, Musikwissenschaft, Ethnologie, Performing Arts, Architektur) und MobilTech (Autonomes Fahren). Die Interoperabilität mit der NFDI wird dabei durch die Implementierung von Schnittstellen, die Berücksichtigung empfohlener Datentypen und die Unterstützung disziplinspezifischer Metadaten sichergestellt.