

Management digitaler Forschungsdaten im akademischen Umfeld – Lessons learned aus der Einführung von RADAR

Kerstin Soltau, Matthias Razum¹
Dorothea Strecker²

¹ FIZ Karlsruhe - Leibniz-Institut für Informationsinfrastruktur

² Humboldt-Universität zu Berlin

RADAR: vom DFG-geförderten Projekt zum verstetigten Dienst

Mit RADAR nahm 2017 ein disziplinübergreifender Cloud-Dienst für die langfristige Archivierung und Publikation digitaler Forschungsdaten den Betrieb auf. RADAR wird von FIZ Karlsruhe - Leibniz-Institut für Informationsinfrastruktur betrieben und wendet sich derzeit primär an Hochschulen und außeruniversitäre Forschungseinrichtungen, die keine eigene Forschungsdateninfrastruktur betreiben oder die RADAR ergänzend zu existierenden disziplinspezifischen Angeboten nutzen möchten.

Die Entwicklung von RADAR erfolgte im Rahmen eines DFG-geförderten Projekts (2013-2016) unter Beteiligung von fünf Forschungseinrichtungen aus den Natur- und Informationswissenschaften. (RADAR, 2016) Zu den Projektpartnern zählten neben FIZ Karlsruhe die Technische Informationsbibliothek (TIB), die Fakultät für Chemie und Pharmazie der Ludwig-Maximilians-Universität München (LMU), das Leibniz-Institut für Pflanzenbiochemie Halle (IPB), sowie das Steinbuch Centre for Computing (SCC) des Karlsruher Instituts für Technologie (KIT). Die Kooperation von Informationsinfrastruktureinrichtungen mit Forschenden an Hochschulen und außeruniversitären Forschungseinrichtungen zielte darauf ab, ein bedarfsgerechtes Angebot „aus der wissenschaftlichen Community für die wissenschaftliche Community“ zu ermöglichen. Bereits während der Projektlaufzeit wurde ein nachhaltiges Geschäftsmodell erarbeitet, das nach einer fünfjährigen Anlaufphase ein sich selbst tragendes Dienstangebot sichern soll. Ein wichtiger Aspekt ist dabei der Verzicht auf Projektförderung für den laufenden Betrieb von Anfang an.

Seit Ende der Projektphase wird RADAR von FIZ Karlsruhe betrieben, das als Vertragspartner für die nutzenden Einrichtungen auftritt. Die technische Infrastruktur von RADAR wird vom SCC und dem Zentrum für Informationsdienste und Hochleistungsrechnen (ZIH) der TU Dresden bereitgestellt.

Die Speicherung der Forschungsdaten erfolgt in drei Kopien an geographisch getrennten Standorten in den Rechenzentren des SCC und des ZIH. Als deutsches DataCite-Mitglied registriert die TIB Hannover DOIs für Datensätze, die über RADAR publiziert werden. Der komplette RADAR-Service und seine Infrastruktur unterliegen damit deutschem Recht.

Seit Projektbeginn zielte RADAR insbesondere auf das institutionelle Forschungsdatenmanagement (FDM) im „Long Tail“ der Forschung ab, also auf Disziplinen, die mit kleineren Datenmengen umgehen und für die bisher keine disziplinspezifischen Forschungsdateninfrastrukturen zur Verfügung stehen. Als disziplinübergreifender Dienst steht RADAR somit vor der Herausforderung, fachspezifische Anforderungen in einer generischen Infrastruktur zusammenzuführen. Bei der Sammlung von Bedarfsanforderungen verfolgte das Projektteam verschiedene Ansätze.

Die Konzeption während der Projektphase wurde durch Anforderungsanalysen begleitet, die die bedarfsgerechte, standardkonforme und dem Stand der Technik entsprechende Entwicklung der Infrastruktur und die Definition der zentralen Diensteigenschaften, wie z. B. Dienstleistungsmodell, Rechtekonzept Datenmanagement oder Systemarchitektur sicher stellten. (Hahn, Neumann, & Razum, 2014) Die Analyse fachwissenschaftlicher Anforderungen umfasste sowohl einen Vergleich mit etablierten internationalen Dienstleistern als auch eine Analyse nationaler, überwiegend disziplinspezifischer Forschungsdatenrepositorien und zeigte Lücken in der bestehenden Infrastrukturlandschaft auf, beispielsweise das Fehlen disziplinübergreifender Angebote unter deutschem Recht. Insbesondere die Projektpartner aus den Bereichen Chemie und Biochemie trugen dazu bei, dass das RADAR-Dienstleistungsspektrum bestmöglich an die Bedarfe und Workflows von Forschenden an akademischen Einrichtungen angepasst wurde. Das RADAR-Metadatenchema wurde neben den Projektpartnern auch von Akteurinnen und Akteuren aus weiteren Fachdisziplinen (Material-, Geistes-, Sport- und Gesundheitswissenschaften) mit vorhandenen Forschungsdaten auf Anwendbarkeit geprüft. Die Konzeption und Entwicklung von RADAR orientierte sich zudem an den Ergebnissen nationaler und internationaler Initiativen.

Um die nutzergerechte Ausrichtung von RADAR zu gewährleisten, zielte das Projektteam auf eine enge Abstimmung mit den späteren Nutzenden und Institutionen ab.

Während der Projektphase wurden drei öffentliche Workshops mit mehr als 80 Teilnehmenden durchgeführt. Zum Ende des zweiten Projektjahres erfolgte die Freischaltung eines Testsystems, über das mehr als 30 Einrichtungen das System evaluieren und Anregungen für dessen Weiterentwicklung geben konnten. Für die Kommunikation mit der wissenschaftlichen Community wurde vielfältige Publikationen zum Projekt veröffentlicht und Präsentationen gehalten sowie eine zweisprachige Website mit Informationsmaterialien (beispielsweise in Form eines Glossars, FAQs oder der Dokumentation von Metadatenschema und API) erstellt.¹ Auf Anregung des Ausschusses für wissenschaftliche Bibliotheken und Informationssysteme (AWBI) wurde für das RADAR-Projekt ein wissenschaftlicher Beirat etabliert, dessen Vertreterinnen und Vertreter nationale Infrastruktureinrichtungen repräsentieren. Mit dem Beirat wurde insbesondere das Geschäftsmodell, die Integration des Dienstes in das wissenschaftspolitische Umfeld und die funktionale Ausgestaltung von RADAR diskutiert.

RADAR - Dienstleistungen, Dienstmerkmale und Geschäftsmodell

Funktional bietet RADAR seit Aufnahme des Dienstes drei zentrale Dienstleistungen an:

Datenarchivierung: Die Datenarchivierung dient der langfristigen und formatunabhängigen Aufbewahrung von Forschungsdaten für die jeweils festgelegten Haltefristen (z. B. 10 Jahre, wie von der DFG im Rahmen guter wissenschaftlicher Praxis empfohlen, aber auch abweichende Haltefristen sind möglich). Forschungsdaten werden in Form von paketierte Zusammenstellungen gesichert (bitstream preservation) und erhalten einen eindeutigen Identifier. Sofern von der Datengeberin bzw. dem Datengeber nicht anders vorgesehen, werden archivierte Forschungsdaten und zugehörige Metadaten nicht veröffentlicht. Durch eine flexible Zugriffsverwaltung können Datengeberinnen und Datengeber Datensätze gezielt mit benannten Dritten oder der Allgemeinheit teilen.

Datenpublikation: Publierte Datensätze werden für mindestens 25 Jahre gesichert. Jedes publizierte Datenpaket erhält einen persistenten Identifier (DOI), wird

¹ <https://www.radar-service.eu/de>

automatisch bei DataCite indexiert und über standardisierte Protokolle (OAI-PMH) zum Harvesting angeboten. Dies sorgt für maximale Verbreitung und Auffindbarkeit der Forschungsdaten.

Über die DOI ist das Datenpaket eindeutig und dauerhaft identifizierbar, zitierfähig und kann mit wissenschaftlichen Publikationen verknüpft werden. Bei Bedarf können DOIs bereits vor der Datenpublikation reserviert werden. Falls Forschungsdaten nicht sofort veröffentlicht werden sollen, kann ein Embargo von bis zu einem Jahr festgelegt werden. Das Datenpaket ist dann erst nach Ablauf dieser Sperrfrist öffentlich zugänglich, die Metadaten sind sofort nach der Veröffentlichung einsehbar. Für jedes publizierte Datenpaket muss die Datengeberin bzw. der Datengeber eine Lizenz angeben (z. B. Creative Commons 4.0).

Peer-Review von Daten: Forschungsdaten können vor der Veröffentlichung im Rahmen eines Peer-Review-Prozesses begutachtet werden, beispielsweise durch externe Gutachterinnen und Gutachter, Verlage oder wissenschaftliche Zeitschriften. Über die Peer-Review-Funktion wird eine sichere URL generiert, die an die Gutachterinnen und Gutachter weitergeleitet werden kann. Nach Abschluss des Peer-Review-Prozesses verliert der Link seine Gültigkeit. Während des Peer-Review-Prozesses kann der betreffende Datensatz nicht verändert werden.

Zu den wesentlichen Dienstmerkmalen von RADAR zählen die Systemarchitektur, das Metadatenschema sowie das Rollen- und Rechtemodell. Eine weitere Eigenschaft des Dienstes ist das Geschäftsmodell.

Die RADAR-Systemarchitektur (Abb. 1) ist modular aufgebaut und besteht aus dem User Interface, der Management Layer und der Storage Layer (Archiv). Die Schichten kommunizieren über Application Programming Interfaces (API) miteinander. Dieser offene Aufbau ermöglicht die Integration von RADAR in bestehende Systeme und Arbeitsprozesse, wobei einzelne Komponenten von RADAR gegen eigene Lösungen ausgetauscht oder parallel betrieben werden können. Anwendungsfälle umfassen beispielsweise den Betrieb eines institutionseigenen Frontends, das automatisierte Hochladen von Forschungsdaten in die Management Layer oder die Übertragung beziehungsweise den Abruf von (Meta-)Daten aus anderen Anwendungen.

Die gesamte entwickelte Software ist ausschließlich aus Open Source Komponenten aufgebaut und implementiert ein OAIS-konformes Langzeitarchivierungssystem.

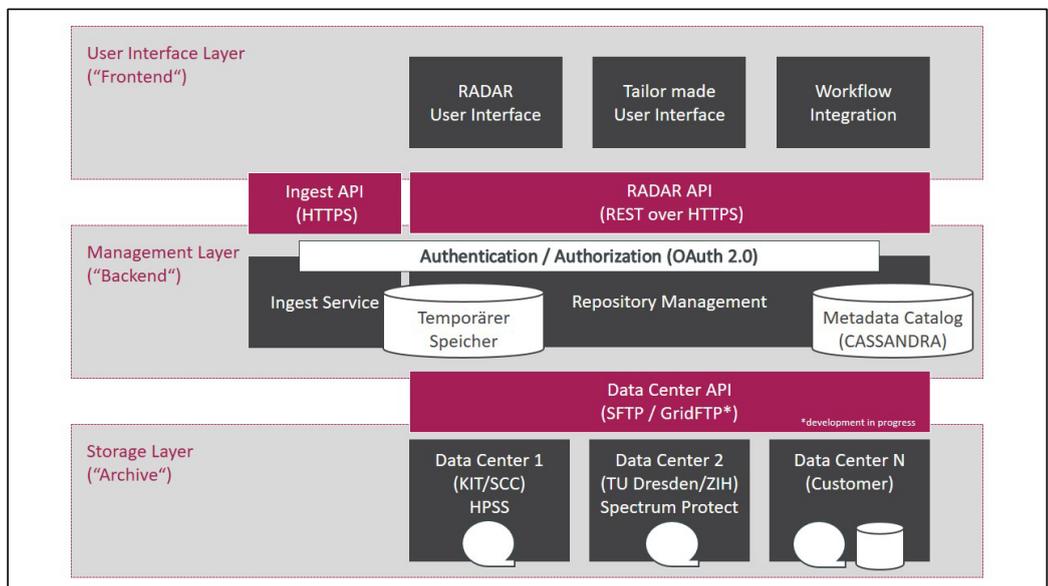


Abbildung 1: RADAR-Systemarchitektur

Das RADAR Metadatenschema² ist disziplinübergreifend angelegt und kompatibel mit dem DataCite Metadata Schema und DublinCore. Es erlaubt die Verwendung von Normdaten für Personen (ORCID iD) und Förderorganisationen (CrossRef Open Funder Registry). Das Metadatenschema umfasst neben 10 Pflichtfeldern, die für die DOI-Registrierung notwendig sind, 13 optionale Felder zur Beschreibung der Datenerhebung und -aufbereitung. Durch die Kombination aus kontrollierten Vokabularen und Freitext-Einträgen ermöglicht RADAR die Interoperabilität der beschriebenen Forschungsdaten, während gleichzeitig der Heterogenität der Daten aus einer Vielzahl von Disziplinen Rechnung getragen wird.

Ein klar definiertes Rollen- und Rechtemodell (Abb. 2) erlaubt die delegierte Administration durch die nutzende Einrichtung. Damit können entsprechend der jeweiligen Bedürfnisse Prozesse strukturiert, Aufgaben verteilt und interne Verantwortlichkeiten definiert werden. Von der nutzenden Einrichtung eingesetzte

² <https://www.radar-service.eu/de/radar-schema>

Administratorinnen und Administratoren verwalten die RADAR-Arbeitsbereiche, die als zentrale Einstiegspunkte für Forschende einer Arbeitsgruppe oder eines Projekts dienen. Administratorinnen und Administratoren können Kuratorinnen und Kuratoren bestimmen, welche Forschungsdaten in einem Arbeitsbereich ablegen, mit Metadaten beschreiben und nach qualitätssichernden Maßnahmen archivieren oder publizieren. Optional können Subkuratorinnen und Subkuratoren benannt werden, die Forschungsdaten bearbeiten und beschreiben, aber nicht archivieren oder publizieren können. Über eine integrierte Registrierung wird die Nutzung von RADAR durch Forschende administrativ erleichtert. Sofern die Einrichtung an DFN-AAI teilnimmt, erfolgt die Authentifizierung mit der institutionellen Nutzererkennung.

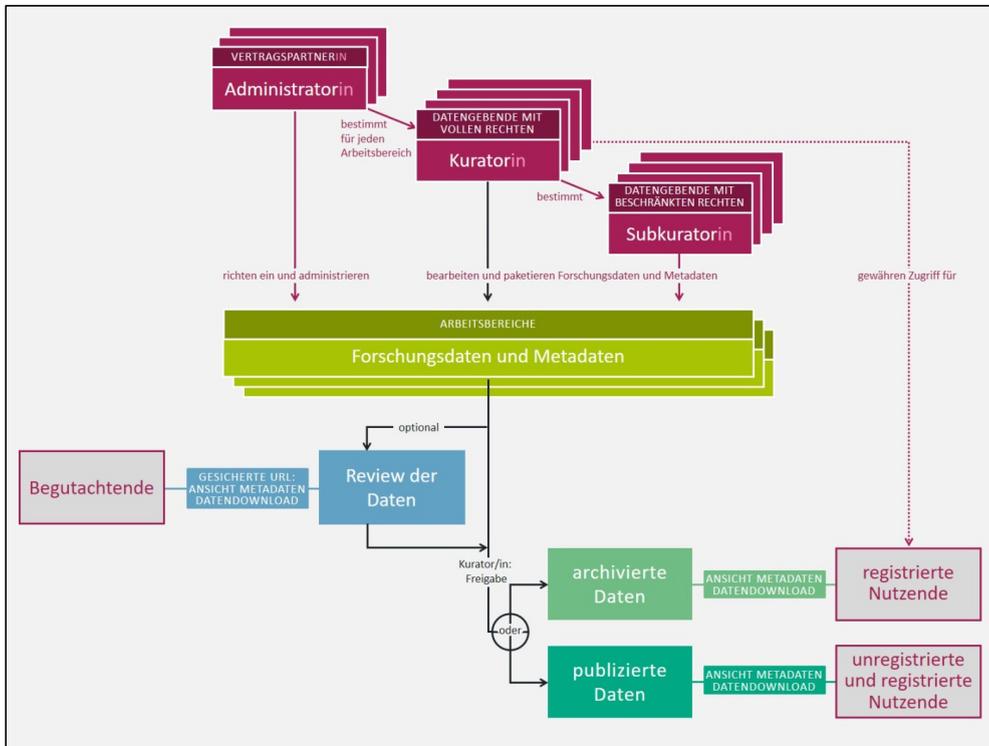


Abbildung 2: RADAR Rollen- und Rechtemodell

Geschäftsmodell: RADAR ist auf langfristigen Betrieb ausgelegt und operiert nicht gewinnorientiert. Die Nutzung von RADAR setzt den Abschluss eines Dienstleistungsvertrags voraus, für den eine jährliche Grundgebühr anfällt.

Die Dienstleistungen Datenarchivierung und Datenpublikation werden nutzungsbasiert abgerechnet. Für archivierte Daten fallen innerhalb der gewählten Haltefrist jährliche Kosten an. Nutzende Einrichtungen können archivierte Daten nach Vertragsende aufgrund der verwendeten Standardformate für die Datenpakete einfach an andere Dienstleister weitergeben und damit das Risiko eines „Vendor Lock-ins“ ausschließen. Publierte Daten müssen unabhängig von der Vertragssituation verfügbar bleiben. Aus diesem Grund bietet RADAR ein Modell basierend auf Einmalzahlungen für veröffentlichte Daten mit einer garantierten Haltefrist von mindestens 25 Jahren, selbst nach Vertragsende.

Mit RADAR können nutzende Einrichtungen Kosten einsparen, die für den Aufbau, die Weiterentwicklung und den Unterhalt eigener Forschungsdatendienste anfallen. Gleichzeitig ermöglicht die delegierte Administration an der nutzenden Einrichtung die Verwaltung von Forschungsdaten durch vorhandene Stellen, beispielsweise durch Bibliothekspersonal, die auch die Beratung der Forschenden vor Ort übernehmen können. Durch die Transparenz des Preismodells können Kosten, die für die Nutzung von RADAR-Diensten anfallen, vorab kalkuliert und über Drittmittel mit eingeworben werden. Über Quotas lassen sich zudem Kostenobergrenzen flexibel festlegen, um notwendige Mittel besser planen zu können.

Aktuelle Entwicklungen im FDM-Umfeld

Seit RADAR 2017 den Produktivbetrieb aufnahm, hat sich das Umfeld für Forschungsdatenrepositorien dynamisch weiterentwickelt. Im neu ausgerichteten Förderprogramm "Fachinformationsdienste für die Wissenschaft" der DFG entwickeln einige FIDs disziplinspezifische Forschungsdatendienste. (Deutsche Forschungsgemeinschaft. Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme, 2019; Harbeck & Kaun, 2019) Disziplinübergreifende Forschungsdatenrepositorien verzeichneten in den letzten Jahren einen deutlichen Anstieg der abgelegten Datensätze. (Assante, Candels, Castelli & Tani, 2016) Allerdings gibt es bisher wenige disziplin- und einrichtungsübergreifende Repositorien, die unter deutscher Jurisdiktion stehen.

Auf Landesebene haben sich Initiativen gebildet, die einrichtungsübergreifende Ansätze für das FDM entwickeln, beispielsweise HeFDI in Hessen oder FoDaKo und

die Landesinitiative NFDI der Digitalen Hochschule in Nordrhein-Westfalen. (Krause et al., 2018; Hess, von Rekowski, Roller, & Walger, 2019; Grasse, López, & Winter, 2018) Parallel soll mit der Nationalen Forschungsdateninfrastruktur (NFDI) ein übergreifendes FDM für das deutsche Wissenschaftssystem entstehen. (Gemeinsame Wissenschaftskonferenz, 2018). Dabei wird die wissenschaftsgeleitete Etablierung disziplinärer Konsortien angeregt, die Fächerkompetenz mit der Etablierung geeigneter Infrastrukturlösungen verknüpfen. Daneben gibt es auf internationaler Ebene Bestrebungen zu einer intensiveren Vernetzung bestehender Infrastrukturkomponenten im Rahmen der European Open Science Cloud (Europäische Kommission, 2018).

2016 wurden mit den FAIR Principles Ziele für Forschungsdaten formuliert: "findable, accessible, interoperable, reusable". (Wilkinson et al., 2016) Indem immer mehr Interessensgruppen Bezug auf die FAIR Principles nehmen, werden sie zunehmend zu einem Quasi-Standard für den Umgang mit Forschungsdaten. (Rat für Informationsinfrastrukturen, 2019; Deutsche Bundesregierung, 2018)

Mit der Bewusstseinschärfung für die Notwendigkeit von FDM wächst auch das Verständnis für damit verbundene Kosten. Neben der Datenspeicherung verursachen diese insbesondere personalintensive Arbeitsschritte wie Übernahme in das Datenarchiv, Erstellung von Metadaten und Maßnahmen zur digitalen Langzeitarchivierung. (Beagrie, Chruszcz & Lavoie, 2008) (Wollschläger & Dickmann, 2016) Eine aktuelle Studie, die von der Europäischen Kommission in Auftrag gegeben wurde, macht jedoch auch auf Opportunitätskosten in Höhe von mindestens 10,2 Milliarden Euro pro Jahr aufmerksam, die durch das Fehlen von FAIR Data in Europa entstehen. (Europäische Kommission. Generaldirektion Forschung und Innovation, 2019, S. 26) Vor diesem Hintergrund ist die Entwicklung nachhaltiger Geschäftsmodelle nach wie vor eine zentrale Herausforderung für Forschungsdatenprogramme. Das zeigt beispielsweise die Bewertung des Förderprogramms „Informationsinfrastrukturen für Forschungsdaten“ durch die Deutsche Forschungsgemeinschaft - demnach wird der langfristige Betrieb von Infrastrukturen, die im Rahmen von Drittmittelprojekten entstehen, insbesondere durch das Auslaufen drittmittelfinanzierter, befristeter Stellen erschwert. (Deutsche Forschungsgemeinschaft, 2019a, S.56)

RADAR im Produktivbetrieb

Mit dem Abschluss mehrerer Dienstleistungsverträge innerhalb weniger Monate nach Aufnahme des Produktivbetriebs wurde die Geschäftsplanung im ersten Betriebsjahr übertroffen. Auch in den Folgejahren konnten weitere Einrichtungen für RADAR gewonnen werden. Funktionale Erweiterungen, wie beispielsweise die Sicherstellung der Barrierefreiheit, die Einführung einer optionalen Datenintegritätsprüfung oder erste Branding-Optionen wurden im Rahmen der laufenden Softwarepflege umgesetzt. Die RADAR-Website³ wird kontinuierlich um Informationsmaterialien, Dokumente und News erweitert. Darüber hinaus wird ein eigener Twitter-Account betrieben. Die Fachöffentlichkeit wird regelmäßig über den aktuellen Funktionsumfang informiert und durch Workshops und Präsentationen in die Weiterentwicklung des Dienstes eingebunden. Die jährlichen Workshops richten sich sowohl an bestehende Nutzende als auch an Einrichtungen, die Interesse am Einsatz von RADAR haben. Um den Austausch mit den Teilnehmenden zu fördern und deren Bedarfe festzustellen, werden interaktiven Elementen wie Fragerunden, Beiträgen der Teilnehmenden und Diskussionen viel Raum gegeben. Das RADAR-Testsystem steht weiterhin zur Verfügung und bietet interessierten Einrichtungen die Möglichkeit, den Dienst unverbindlich kennenzulernen und Rückmeldung zu geben. Zudem werden in Vertragsverhandlungen Bedarfe und Anforderungen geäußert, die in Entscheidungen über die Weiterentwicklung des Dienstes einfließen. Trotz des breiten Interesses überrascht die Zahl über RADAR publizierter Datensätze, die derzeit im niedrigen zweistelligen Bereich liegt. Andererseits deckt sich diese Beobachtung mit Umfrageergebnissen, die auf eine bislang eingeschränkte Bereitschaft von Forschenden zum öffentlichen Teilen von Forschungsdaten hinweisen. (Fecher, 2018) Gleichzeitig zeigt sich, dass zwischen Vertragsabschluss und der Publikation erster Datensätze oftmals mehrere Monate liegen, in denen interne Abläufe in den nutzenden Einrichtungen etabliert und Qualitätssicherungsprozesse durchlaufen werden.

³ <https://www.radar-service.eu/de>

Positionierung von RADAR und geplante Erweiterungen

Disziplinübergreifende Forschungsdatenrepositorien spielen eine wichtige Rolle bei der Datenpublikation in Disziplinen, die bisher keine eigenen Forschungsdateninfrastrukturen und -dienste anbieten. Allerdings müssen sie auch Lösungen für vielfältige Datenformate und -typen, Forschungsgemeinschaften und -praktiken anbieten. (Assante, Candela, Castelli & Tani, 2016)

RADAR wurde als disziplin- und formatunabhängiges Repository konzipiert und ermöglicht Einrichtungen darum, Angebote für eine Vielzahl von Datentypen zu entwickeln. Durch den modularen und offenen Aufbau kann RADAR flexibel an lokale Bedürfnisse angepasst werden. Die oben beschriebenen Entwicklungen des wissenschaftspolitischen Umfelds machen allerdings deutlich, dass disziplinübergreifende Angebote zukünftig stärker auf disziplinspezifische Besonderheiten eingehen und die Interoperabilität befördern müssen.

Forschungsdateninfrastrukturen erzeugen ein "Ökosystem" digitaler Dienste und Ressourcen und wirken als Ermöglicher bestimmter Forschungspraktiken. (Bowker, Baker, Millerand, & Ribes, 2010) Als Schnittstelle zwischen Datengeberinnen bzw. Datengebern und Nutzenden fördern sie den Austausch innerhalb der "Designated Community" und damit die Akzeptanz von Forschungsdaten als eigenständige wissenschaftliche Objekte. RADAR verpflichtet sich der Umsetzung der FAIR Principles und bietet eine Reihe von Maßnahmen zu ihrer Implementierung an (Tab. 1). Zukünftige Weiterentwicklungen des Dienstes sollen daher auch die Auffindbarkeit, Zugänglichkeit, Interoperabilität und Nachnutzbarkeit von Forschungsdaten verbessern.

Findable	F1: (meta)data are assigned a globally unique and eternally persistent identifier	<ul style="list-style-type: none"> • Registrierung von DOIs für publizierte Datenpakete, RADAR-IDs für archivierte Datenpakete
	F2: data are described with rich metadata.	<ul style="list-style-type: none"> • MetadatenSchema basierend auf dem DataCite Metadata Schema • Überprüfung der Metadaten auf Vollständigkeit
	F3: (meta)data are registered or indexed in a searchable resource.	<ul style="list-style-type: none"> • Indexierung bei DataCite, Google und B2FIND • Harvesting über OAI-PMH
	F4: metadata specify the data identifier.	<ul style="list-style-type: none"> • Feld <identifierType> wird bei Ausstellung des Identifiers automatisch ausgefüllt
Accessible	A1: (meta)data are retrievable by their identifier using a standardized communications protocol.	<ul style="list-style-type: none"> • (Meta-)Daten sind über eine API und OAI-PMH zugänglich
	A1.1: the protocol is open, free, and universally implementable.	<ul style="list-style-type: none"> • Verwendete Protokolle und Schnittstellen sind weit verbreitet und gut dokumentiert
	A1.2: the protocol allows for an authentication and authorization procedure, where necessary.	<ul style="list-style-type: none"> • Rollen- und Rechtemodell erlaubt Datengeberinnen bzw. Datengebern Kontrolle über Zugriff auf Datensätze
	A2: metadata are accessible, even when the data are no longer available.	<ul style="list-style-type: none"> • Landing page bleibt nach der Löschung eines Datensatzes erhalten
Interoperable	I1: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	<ul style="list-style-type: none"> • MetadatenSchema basierend auf XML
	I2: (meta)data use vocabularies that follow FAIR principles.	<ul style="list-style-type: none"> • Personenidentifikation über ORCID iDs • Angaben zu Förderorganisationen über Crossref Funder Registry
	I3: (meta)data include qualified references to other (meta)data.	<ul style="list-style-type: none"> • Referenzierung anderer digitaler Ressourcen über persistente Identifier
Re-Usable	R1: meta(data) have a plurality of accurate and relevant attributes.	<ul style="list-style-type: none"> • Umfassende Beschreibung der Eigenschaften von Forschungsdaten über das MetadatenSchema
	R1.1: (meta)data are released with a clear and accessible data usage license.	<ul style="list-style-type: none"> • Für jedes Datenpaket muss eine Lizenz (z.B. Creative Commons 4.0) vergeben werden
	R1.2: (meta)data are associated with their provenance.	<ul style="list-style-type: none"> • MetadatenSchema erlaubt Angaben zur Provenienz von Forschungsdaten
	R1.3: (meta)data meet domain-relevant community standards.	<ul style="list-style-type: none"> • MetadatenSchema wurde auf Anwendbarkeit in verschiedenen Disziplinen getestet

Tabelle 1: Umsetzung der FAIR Principles mit RADAR

Einer der Hauptmerkmale von RADAR war bisher die entfallende Notwendigkeit, eine eigene Dateninfrastruktur aufzubauen und zu betreiben. Die Kommunikation mit interessierten Einrichtungen zeigte, dass dieses Modell nicht alle Anwendungsfälle adäquat abdecken kann. Eine Öffnung des Betriebs- und Geschäftsmodells soll darum die flexible Anpassung an alternative Einsatzszenarien ermöglichen.

Im Folgenden werden Ansätze der bedarfsgesteuerten Weiterentwicklung des Dienstes geschildert.

Einbindung eigener Rechenzentren

Bisher ist RADAR ausschließlich als Cloud-Dienst verfügbar, der die dauerhafte Archivierung und Verfügbarkeit der Forschungsdaten gewährleistet. In Gesprächen mit größeren Hochschulen, außeruniversitären Forschungseinrichtungen und Konsortien wurde der Wunsch nach der Einbindung eigener Rechenzentren in die Storage Layer geäußert, entweder exklusiv oder ergänzend zu bestehenden Datenzentren (SCC und ZIH). Diese Option würde nutzenden Einrichtungen mehr Kontrolle über Forschungsdaten einräumen und könnte gleichzeitig Kosten senken. Derzeit werden technische Anpassungen vorgenommen, die die Umsetzung institutionsspezifischer Vorgaben für die Speicherung von Archivkopien unterstützen. Diese Anpassungen erfordern auch eine Weiterentwicklung des Geschäftsmodells (beispielsweise eine Erweiterung um Pauschalpreise) sowie der Verträge und Haftungsregelungen. Die RADAR-Software wird in diesem Modell nutzenden Einrichtungen weiterhin als gehosteter Dienst zur Verfügung gestellt, dabei aber entweder eine oder alle Kopien auf lokaler Speicherinfrastruktur abgelegt werden. Darüber hinaus ist ein Szenario vorgesehen, in dem die nutzende Einrichtung die RADAR-Software selbst auf eigener Hardware betreibt, aber über einen Wartungsvertrag dabei unterstützt wird und regelmäßig Software-Updates erhält.

Unterstützung großer Datenvolumen

Mit der Adressierung von Konsortien und größeren Hochschulen als Zielgruppe gehen Anforderungen hinsichtlich der Verarbeitung großer Datenmengen (im zwei- bis dreistelligen Terabyte-Bereich) einher, die effizientere Übertragungs- und Verarbeitungsverfahren erfordern. Für solche Datenmengen wurde RADAR als Repositorium für den "Long Tail" nicht konzipiert. Um große Datenvolumina besser zu unterstützen, soll zukünftig ein Datentransfer auch per GridFTP möglich sein. Das bereits in 2018 eingeführte Rabattpreismodell für Datenvolumen über 50 TB greift diesen Aspekt bereits auf.⁴

⁴ <https://www.radar-service.eu/de/preise>

Customizing-Optionen

Neben der Einbindung von eigener Speicherinfrastruktur ermöglicht RADAR in Zukunft weitere Anpassungen für nutzende Einrichtungen. Dazu zählen die Vergabe von DOIs mit einem institutionseigenen Präfix, Anpassungen an das Corporate Design, Verweise auf institutionseigene Unterstützungsangebote sowie die Möglichkeit einer institutionellen Sicht, über die nur die von der jeweiligen Institution publizierten Datensätze angezeigt werden. Diese Anpassungen befinden sich aktuell in der Umsetzung, initiiert durch die Anforderungen des Karlsruher Institut für Technologie (KIT), das als Projektpartner aktuell ein auf RADAR basierendes eigenes Forschungsdatenrepositorium aufbaut. Diese Anpassungen sollen auch anderen nutzenden Institutionen noch 2019 zur Verfügung stehen.

Disziplinspezifische Anpassungen

FIZ Karlsruhe bringt sich mit RADAR in mehreren sich formierenden NFDI-Konsortien als Infrastrukturpartner für die Bereitstellung eines bedarfsgerechten Forschungsdatenrepositoriums ein. RADAR kann sich hier als generischer Baustein für disziplinspezifische Angebote etablieren. Dazu wird RADAR über entsprechende Erweiterungen an disziplinspezifische Bedarfe angepasst. Dies betrifft z. B. die optionale Beschreibung von Datenpaketen mit zusätzlichen, fachspezifischen Metadaten in Form von XML-Dateien. Sofern ein geeignetes Schema referenziert wird, kann die mitgelieferte XML-Datei automatisch validiert werden. Mittels Harvesting von Metadaten über OAI-PMH und einer entsprechenden Indexierung lassen sich fachspezifische Suchen aufbauen. In diesem Zusammenhang könnten NFDI-Konsortien wertvolle Impulse für die Verwendung geeigneter Metadatenschemata geben, beispielsweise unter Anwendung des Research Data Alliance Metadata Directory.⁵

Verknüpfung mit anderen Forschungsdatendiensten

Weiterhin prüft RADAR die Kooperation mit anderen Forschungsdatendiensten, um Forschenden bereits früher im Lebenszyklus von Forschungsdaten sinnvolle Dienste anzubieten. Hier ist insbesondere RDMO zu nennen: ein Werkzeug, das Forschende bei der Erstellung und Umsetzung von Datenmanagementplänen (DMP) unterstützt

⁵ <https://rd-alliance.github.io/metadata-directory/standards/>

und das bereits von vielen Einrichtungen genutzt wird.⁶ (Neuroth et al., 2018) Durch die Integration von RDMO könnte RADAR den gesamten Lebenszyklus von Forschungsdaten begleiten, vom Projektstart über die Datenerhebung und -aufbereitung bis zur Datenpublikation beziehungsweise -archivierung. Durch den Import von Metadaten aus RDMO könnten Doppelseingaben vermieden und die Nutzerfreundlichkeit von RADAR verbessert werden. Perspektivisch wäre auch eine stärkere Integration von RDMO und RADAR im Sinne von "machine actionable Data Management Plans" möglich. (Miksa, Simms, Mietchen & Jones, 2019)

Personenbezogene Daten

Gemäß dem RADAR-Dienstleistungsvertrag ist die Archivierung beziehungsweise Publikation personenbezogener, nicht anonymisierter Daten ohne Einwilligung der Betroffenen nicht zulässig. Bei Beratungsgesprächen mit interessierten Einrichtungen wird dieses Thema regelmäßig adressiert. Aus diesem Grund wird sich RADAR zukünftig stärker mit möglichen Angeboten für personenbezogene Daten befassen. Basierend auf Fallbeispielen soll aus juristischer und technischer Perspektive untersucht werden, ob unter Anwendung bestehender Möglichkeiten der Pseudonymisierung eine Erweiterung des Angebots möglich ist.

Zertifizierung

Da Forschungsdatenrepositorien die Verantwortung für die langfristige Verfügbarkeit von und den Zugriff auf Forschungsdaten übernehmen, müssen sie Qualitätsansprüchen aller beteiligter Interessensgruppen gerecht werden. Viele der am FDM beteiligten Akteure, darunter Forschende, Verlage und Forschungsförderer, stellen ähnliche Ansprüche an Forschungsdatenrepositorien. (Husen, de Wilde, de Ward & Cousijn, 2017) Zertifikate wie das CoreTrustSeal⁷ formalisieren diese Anforderungen und zielen darauf ab, die Vertrauenswürdigkeit von Forschungsdatenrepositorien zu bewerten und sichtbar zu machen. (Dillo & Leeuw, 2018) Dadurch fördern sie Transparenz und können Datengeberinnen und Datengebern Orientierung bei der Entscheidung für geeignete

⁶ <https://rdmorganiser.github.io/kooperationen/>

⁷ <https://www.coretrustseal.org/>

Forschungsdateninfrastrukturen geben. Bisher ist die Zertifizierung von Forschungsdatenrepositorien allerdings noch nicht weit verbreitet.

(Kindling et al., 2017) RADAR lässt sich derzeit nach dem CoreTrustSeal zertifizieren. Damit möchte RADAR einerseits das Vertrauen (potenzieller) Vertragspartner sicherstellen beziehungsweise gewinnen. Andererseits können Vertragspartner durch den zertifizierten Dienst die Vertrauenswürdigkeit an Datengeberinnen und Datengeber "weiterreichen".

Zusammenfassung und Ausblick

Die Betrachtung des nationalen und internationalen Umfelds macht deutlich, wie dynamisch sich die wissenschaftspolitischen Rahmenbedingungen des Forschungsdatenmanagements verändern. Infrastrukturanbieter müssen darum in der Lage sein, flexibel auf veränderte Anforderungen zu reagieren und gleichzeitig ein Modell für die langfristige Verfügbarkeit ihrer Dienstleistung implementieren.

Seit der Überführung von RADAR in den Produktivbetrieb wird der Dienst bedarfsorientiert weiterentwickelt. Bisherige Erfolge bei der Zusammenarbeit mit Forschungseinrichtungen und das breite Interesse von Einrichtungen schlagen sich bisher noch nicht in einer intensiven Nutzung durch Forschende nieder. Als Dienstleister für wissenschaftliche Einrichtungen kann RADAR auf die Nutzung durch Forschende nur indirekt einwirken. Die hier vorgestellten Anpassungen und Erweiterungen zielen darauf ab, Forschungsdatenmanagement und die Publikation von Forschungsdaten für Forschende einfacher und attraktiver zu gestalten. Darum setzt RADAR zukünftig verstärkt auf disziplinspezifische Angebote und den Kontakt zu forschungsnahen Diensten, beispielsweise über Konsortien im Rahmen der NFDI.

Zu diesem Zweck öffnet sich RADAR neuen Kooperationspartnern und -formen. In diesen Kooperationen können sich RADAR und FIZ Karlsruhe mit umfassender technischer, juristischer und informationswissenschaftlicher Expertise einbringen.

Der Ausbau von Customizing-Optionen, der auch die Einbindung institutionseigener Speicherkapazität einschließt, ermöglicht den Einsatz von RADAR für ein breiteres Anwendungsspektrum. Zusätzlich soll der Funktionsumfang erweitert werden, beispielsweise durch die Unterstützung großer Datenvolumen und die Verknüpfungen mit anderen Forschungsdatendiensten.

Durch die Zertifizierung nach dem CoreTrustSeal sollen die Qualität und Vertrauenswürdigkeit des Dienstes gesichert und an die Fachcommunity weitergereicht werden.

Literatur

Assante, M., Candela, L., Castelli, D., & Tani, A. (2016). Are Scientific Data Repositories Coping with Research Data Publishing? *Data Science Journal*, 15, 6. <https://doi.org/10.5334/dsj-2016-006>

Beagrie, N., Chruszcz, J., & Lavoie, B. (2008). *Keeping Research Data Safe. A Cost Model and Guidance for UK Universities*. Abgerufen von <https://www.webarchive.org.uk/wayback/archive/20140613220103/http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf>

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2010). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In J. Hunsinger, L. Klastrop, & M. Allen (Hrsg.), *International Handbook of Internet Research* (S. 97–117). https://doi.org/10.1007/978-1-4020-9789-8_5

Deutsche Bundesregierung. (2018). *Die Hightech-Strategie 2025 – Forschung und Innovation für die Menschen*. Abgerufen von <http://dip21.bundestag.de/dip21/btd/19/041/1904100.pdf>

Deutsche Forschungsgemeinschaft. (2019). *Bewertung des Förderprogramms „Informationsinfrastrukturen für Forschungsdaten“*. http://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/studien/studie_forschung_sdaten.pdf

Deutsche Forschungsgemeinschaft. Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme. (2019). *Fachinformationsdienste für die Wissenschaft. Von den Sondersammelgebieten zu den Fachinformationsdiensten: Zwischenbilanz der Umstrukturierung der Förderung*. Abgerufen von https://www.dfg.de/download/pdf/foerderung/programme/lis/fid_zwischenbilanz_umstrukturierung_foerderung_sondersammelgebiete.pdf

Dillo, I., & Leeuw, L. de. (2018). CoreTrustSeal. *Mitteilungen Der Vereinigung Österreichischer Bibliothekarinnen Und Bibliothekare*, 71(1), 162–170. <https://doi.org/10.31263/voebm.v71i1.1981>

Europäische Kommission. (2018). *Implementation Roadmap for the European Open Science Cloud*. Abgerufen von https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf

Europäische Kommission. Generaldirektion Forschung und Innovation. (2019). *Cost of not having FAIR research data. Cost-benefit analysis for FAIR research data*. Abgerufen von <https://doi.org/10.2777/02999>

Fecher, B. (2018). *Eine Reputationsökonomie*. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-20895-0>

Gemeinsame Wissenschaftskonferenz. (2018). *Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur*. Abgerufen von <https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf>

Grasse, M., López, A., & Winter, N. (2018). Landesinitiative NFDI – a Central Point of Contact for RDM for Higher Education Institutions in the German State of North Rhine-Westphalia. *Data Science Journal*, 17, 25. <https://doi.org/10.5334/dsj-2018-025>

-
- Hahn, M., Neumann, J., & Razum, M. (2014). RADAR – Ein Forschungsdaten-Repository als Dienstleistung für die Wissenschaft. *Zeitschrift Für Bibliothekswesen Und Bibliographie*, 61(1), 18–27. <https://doi.org/10.3196/186429501461150>
- Harbeck, M., & Kaun, M. (2019). Forschungsdaten und Fachinformationsdienste – eine Bestandsaufnahme. *Bibliothek Forschung und Praxis*, 43(1), 35–41. <https://doi.org/10.1515/bfp-2019-2015>
- Hess, V., von Rekowski, T., Roller, S., & Walger, N. (2019). Synergieeffekte durch Kooperation: Hintergründe, Aufgaben und Potentiale des Projekts FoDaKo. *Bibliothek Forschung und Praxis*, 43(1), 98–104. <https://doi.org/10.1515/bfp-2019-2009>
- Husen, S. E., de Wilde, Z. G., de Waard, A., & Cousijn, H. (2017). Recommended versus Certified Repositories: Mind the Gap. *Data Science Journal*, 16, 42. <https://doi.org/10.5334/dsj-2017-042>
- Kindling, M., Pampel, H., van de Sandt, S., Rücknagel, J., Vierkant, P., Kloska, G., ... Scholze, F. (2017). The Landscape of Research Data Repositories in 2015: A re3data Analysis. *D-Lib Magazine*, 23(3/4). <https://doi.org/10.1045/march2017-kindling>
- Krause, E., Brand, O., Deppe, A., Krähwinkel, E., Jagusch, G., Müllerleile, T., ... Wolff-Wölk, A. (2018). FDM vernetzt und kooperativ. *o-bib. Das offene Bibliotheksjournal*, 5(4), 220–236. <https://doi.org/10.5282/o-bib/2018H4S220-236>
- Miksa, T., Simms, S., Mietchen, D., & Jones, S. (2019). Ten principles for machine-actionable data management plans. *PLOS Computational Biology*, 15(3), e1006750. <https://doi.org/10.1371/journal.pcbi.1006750>
- Neuroth, H., Engelhardt, C., Klar, J., Ludwig, J., & Enke, H. (2018). Aktives Forschungsdatenmanagement. *ABI Technik*, 38(1), 55–64. <https://doi.org/10.1515/abitech-2018-0008>
- RADAR. (2016). *Abschlussbericht an die Deutsche Forschungsgemeinschaft*. Abgerufen von https://www.radar-projekt.org/download/attachments/753675/Abschlussbericht_DFG-Projekt_RADAR_Ver%C3%B6ffentlichung.pdf?version=1&modificationDate=1485414994000&api=v2
- Rat für Informationsinfrastrukturen. (2019). *Stellungnahme des Rates für Informationsinfrastrukturen (Rfii) zu den aktuellen Entwicklungen rund um Open Data und Open Access*. Abgerufen von <http://www.rfii.de/?p=3748>
- Wollschläger, T., & Dickmann, F. (2016). Kosten. In H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, & K. Huth (Hrsg.), *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. (Version 2.3, Kapitel 14.2). <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010071949>