

4. RADAR-Workshop (25./26. Juni 2018 Karlsruhe): Zusammenfassung und Ausblick

Der 4. RADAR-Workshop fand am 25 und 26. Juni 2018 im Vortragsraum der Bibliothek des Karlsruher Instituts für Technologie (KIT) statt.

Das Forschungsdatenrepositorium RADAR (www.radar-service.eu) bietet akademischen Institutionen und Forschenden ein umfangreiches, disziplinübergreifendes und verlässliches Angebot zur langfristigen Archivierung und Publikation digitaler Forschungsdaten.

Während die ersten drei RADAR-Workshops noch in der Projektphase stattfanden, war dies der erste Workshop nach Ende der Projektphase und Verstetigung der RADAR-Dienstleistungen als Produkt. Die 22 Teilnehmer umfassten Vertreterinnen und Vertreter wissenschaftlicher Forschungseinrichtungen, Bibliotheken, Rechenzentren, Institute und Unternehmen aus ganz Deutschland. Veranstaltet wurde der Workshop von FIZ Karlsruhe unter Beteiligung der Technischen Informationsbibliothek (TIB), des Steinbuch Centre for Computing (SCC/KIT) und des Leibniz-Instituts für Pflanzenbiochemie (IPB).

Der erste Tag richtete sich mit der grundlegenden Einführung in RADAR, einer Live-Demonstration des Systems, der Kurzvorstellung der Programmierschnittstelle (API) sowie Informationen zu Preisen und Verträgen vor allem an Personen, die mit RADAR noch nicht vertraut waren. Der zweite Tag diente dem Ausblick auf zukünftige Optionen sowie dem Austausch mit der Community. Neben Vorstellungen zu den geplanten funktionalen Erweiterungen sowie zu alternativen Einsatzszenarien und Erweiterungen des Geschäftsmodells boten mehrere Teilnehmerbeiträge zum Thema „Integration von RADAR (Repositorien) in Arbeitsprozesse/Workflows“ sowie die abschließende „Fragen- und Antworten-Runde“ Gelegenheit zum gegenseitigen Austausch und zur Diskussion.

Die nachfolgenden Beiträge sind analog der Workshop-Themenblöcke „Preise und Verträge“, „Funktionale Erweiterungen“, „Erweiterungen des Geschäftsmodells/Alternative Szenarien“, „Integration von RADAR (Repositorien) in Arbeitsprozesse/Workflows“ sowie die „Fragen und Antworten-Runde“ unterteilt.

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 Datengeber / Datennutzer sowie RADAR in Bezug auf die DSGVO erläutert.

Aus dem Kreis der Teilnehmer wurde insbesondere das bisher angebotene Preismodell für archivierte Daten hinterfragt. Aus Gründen hochschulinterner Budgetrestriktionen sollten auch für archivierte Daten (analog zu publizierten Daten) Einmalzahlungen möglich sein.

Die Frage, ob es sich bei RADAR um eine Auftragsdatenverarbeitung handelt, wird an interessierten Einrichtungen diskutiert. Die Meinungen der Juristen dazu sind konträr. FIZ Karlsruhe geht davon aus, dass keine Auftragsverarbeitung vorliegt. FIZ Karlsruhe wird hierzu eine juristische Stellungnahme durch seinen Forschungsbereich „Immaterialgüterrechte in verteilten Informationsinfrastrukturen“ erarbeiten lassen und das Ergebnis veröffentlichen.

Funktionale Erweiterungen

Im Vortrag wurden Ideen zu grundlegenden funktionalen Erweiterungen von RADAR vorgestellt. Dazu zählen die Einbindung weiterer Rechenzentren (s.a. nachfolgenden Abschnitt), die Unterstützung fachspezifischer Metadaten (inkl. OAI-PMH Harvesting), die Integration eines Datenmanagementplan-Werkzeugs (z.B. RDMO), die Unterstützung von Versionierung (inkl. Erratum) und Kollektionen (inkl. Zitation / Downloadoption von Einzeldateien) sowie die semantische Erschließung der Metadaten und deren Bereitstellung als Linked Open Data. Des Weiteren wurde in dem Datamanager-Prototyp der Leibniz Universität Hannover und der TIB basierend auf der CKAN-Software eine Schnittstelle zu RADAR geschaffen, die derzeit getestet wird. Weiterhin ist auch die praktische Erprobung von Exit-Szenarien als vertrauensschaffende Maßnahme in Planung.

Zur Umsetzung hat das Projektteam einen Folgeantrag bei der DFG eingereicht, über den noch im Juli entschieden werden soll. Auch im Fall einer Ablehnung des Antrags werden die Erweiterungen angegangen, aber gestreckt über einen längeren Zeitraum.

Mehrere Teilnehmer würden die Integration von RADAR mit z.B. dem DMP-Tool RDMO begrüßen, da letzteres bereits in der Einrichtung verwendet wird und z.B. der Import von Metadaten als Benefit gesehen wird. Es wurde angemerkt, dass trotz der Erweiterung neuer Funktionalität der Fokus – Archivierung und Publikation von Forschungsdaten – nicht vernachlässigt werden sollte und jede neue Funktion im Hinblick daran zu messen sei, inwieweit sie den Fokus stärkt.

Eine weitere funktionale Erweiterung, die in der Runde diskutiert wurde, war die Integration mit dem Forschungsinformationssystem (FIS) VIVO. Zur Zeit der Antragstellung gab es noch keinen Partner, der über eine entsprechende Testumgebung verfügte. Neuerdings hat die TIB ein VIVO-System in der Erprobung. Bei den weiteren im Workshop vertretenen Einrichtungen gibt es aber bisher kaum etablierte FIS.

Erweiterungen des Geschäftsmodells / Alternative Szenarien

Der Vortrag beleuchtete mögliche zukünftige RADAR-Angebote für Konsortien, Einmalpreise für archivierte Daten, die Option der lokalen Datenspeicherung bei der jeweiligen Einrichtung sowie die Option der lokalen Datenspeicherung inkl. des Betriebs der RADAR-Software (mit Wartungsverträgen). Daneben wurde (nicht nur) für baden-württembergische Einrichtungen die Möglichkeit von getrennten Verträgen mit FIZ Karlsruhe und KIT/SCC (bwDataArchiv) reflektiert. Bei allen vorgestellten Möglichkeiten handelt es sich um mögliche zukünftige Szenarien, nicht jedoch bereits um konkret umsetzbare Angebote!

Mehrere Teilnehmer äußerten sich positiv und interessiert zu den Möglichkeiten, die die Anbindung des eigenen Rechenzentrums an RADAR sowie der lokale Betrieb der RADAR-Software bieten könnte. Ideen, ob diese Optionen für die Speicherung personenbezogener Daten verwendet werden könnten oder ob das eigene Rechenzentrum bzw. OwnCloud/Nextcloud als Alternative zum temporären RADAR-Speicher (ggf. bereits während eines noch laufenden Forschungsprojekts) dienen könnte wurden von den Teilnehmern vorgebracht.

Erwähnt wurde auch die Möglichkeit, RADAR für die Implementierung disziplinspezifischer Repositorien zu verwenden. Hier ist insbesondere die Backend-Funktionalität von Interesse, die über das API in eigene Entwicklungen eingebunden werden kann. Damit können sich die Entwickler eines disziplinären Angebots voll auf die fachlichen Anforderungen fokussieren und die Basisfunktionen von RADAR für die verlässliche Speicherung, Archivierung und Publikation nachnutzen.

Integration von RADAR/Repositorien in Arbeitsprozesse/Workflows

Folgende Teilnehmer berichteten aus Ihren eigenen Einrichtungen:

- Dr. Angelina Kraft, Technische Informationsbibliothek (TIB): *“RADAR-Workflow an der TIB”*
- Dr. Hans-Jürgen Goebelbecker (KIT, Bibliothek): *“Forschungsdatenmanagement: Das Serviceteam RDM@KIT”*
- Markus Becker, Leibniz-Institut für Plasmaforschung und Technologie Technologie e.V. (INP): *“InPT-Dat - Forschungsdatenmanagement am Leibniz-Institut für Plasmaforschung und Technologie e.V. (INP)”*
- Claudia Piesche (Universität Bayreuth, IT-Servicezentrum) *„FDM@UBT: Integration von RADAR in Arbeitsprozesse / Workflows“*
- Thomas Schwager (Universität Vechta, Bibliothek): *“Projekt UniV-FDM: Bottom-up-Managementmodell zur Etablierung eines institutionellen FDM”*

Aus den Vorträgen lassen sich einige übergreifende Erkenntnisse bzw. Herausforderungen ableiten. Eine klar definierte Festschreibung von Verantwortlichkeiten im Rahmen einer FDM-Policy sowie institutionsübergreifend abgestimmte Workflows sind eine wichtige Voraussetzung für ein funktionierendes FDM. Selbst wenn an den Einrichtungen schon fachspezifische Repositorien bedient bzw. genutzt werden, kann Bedarf für ein generisches Repository wie RADAR existieren. FDM-Teams und andere am FDM-beteiligte Abteilungen müssen die Forschenden zum Umgang mit Forschungsdaten und zu Publikationsmöglichkeiten beraten und schulen. Fehlende Sensibilität bzw. Motivation der Forschenden (auch aufgrund fehlender Anreizsysteme), Forschungsdaten nachhaltig zu verwahren/publizieren, wirken hemmend. Die Publikationsbereitschaft der „jungen Forschenden-Generation“ ist meist signifikant höher. Auch die Überarbeitung der GWP sowie Förderauflagen werden Einfluss auf das Archivierungs- und Publikationsverhalten haben.

Fragen und Antworten-Runde / anschließende Diskussion

Es wurde u.a. nochmals darauf hingewiesen, dass ein als generisch konzipiertes Repository wie RADAR aufgrund der Vielzahl der dort speicherbaren Datenformate (Schätzungen: >2000) keine funktionale Langzeitarchivierung leisten kann. RADAR gibt jedoch für einige häufig genutzte Datentypen Empfehlungen, welche Dateiformate für die LZA geeignet erscheinen.

Es ist nicht geplant, RADAR zur „Forschungsdaten-Suchmaschine“ auszubauen, die Metadaten werden per OAI-PMH zum Harvesting angeboten und sind somit an anderer Stelle frei recherchierbar. Unter anderem finden sich die Daten aus RADAR im DataCite Metadata Store, in BASE, in B2FIND und z.B. in fachspezifischen Suchportalen wie an der TIB. Auch große Suchmaschinen wie Google indexieren RADAR.

Das deskriptive RADAR Metadatenchema wird kontinuierlich gegen das DataCite Metadata Schema abgeglichen und sinnvolle Änderungen/Ergänzungen werden zeitnah in RADAR nachgezogen. Die DataCite-Metadaten stehen bereits als RDF-XML Metadaten bereit. Wichtig scheint die automatisierte Metadatenerzeugung bereits bei der Generierung der Forschungsdaten. Eine Übernahme von Metadaten nach Maschinenausgabe im XML-Format ist denkbar.

Teilnehmer stellten in Frage, dass klassische Metadaten Schemata zukünftig durch die Möglichkeiten von TextMining, LinkedOpenData etc. hinfällig sein könnten. Dieser Anforderung wird in der RADAR-Weiterentwicklung mit der Etablierung semantischer Vokabulare Rechnung getragen.

Anregungen für Weiterentwicklungen aus dem Teilnehmerkreis:

- Einführung einer weiteren funktionalen Ebene „Subadministrator / Gruppenadministrator“ für Teilbereiche größerer Einrichtungen/Hochschulen (z.B. Fakultäten/Institute) unter denen die eigentlichen Arbeitsbereiche (z.B. für Lehrstühle, Arbeitsgruppen) eingerichtet werden. Hierzu ist eine genauere Analyse des Use Case / Workflows notwendig.
- Einführung einer weiteren Nutzerrolle, die lediglich die Metadaten bearbeiten kann, nicht jedoch die Forschungsdaten selbst. Eine genauere Analyse des Use Case / Workflows ist notwendig.
- Einführung einer Option, die Metadaten archivierter Daten sichtbar zu machen.
- Erstellung einer Funktionalität zur Übergabe bereits archivierter Datenpakete in den Publikationsstatus.
- Umzug von Forschungsdaten im Status „in Bearbeitung“ (temporärer Speicher) von einem Workspace zum anderen.
- Wäre es wünschenswert, alle Co-Autoren zu kontaktieren und um Zustimmung zu bitten, bevor ein Datensatz eingespeichert wird?
- Wäre ein Tool hilfreich, mit dem institutionellen Datengebern Entscheidungen zu „schwierigen“ Metadatenfeldern (z.B. Rechteinhaber, Lizenzen) erleichtert würde (Fragebaum etc.)?
- Detaillierte Abrechnung bzw. die Möglichkeit der Kostenzuordnung durch verbesserte Statistiken/Reports zum in Anspruch genommenen Datenvolumen je Vertrag und je Arbeitsbereich (auf Abruf, wählbarer Abfragezeitraum)
- Einfaches Customizing, z.B. Einfügen des Institutions-Logo