
Bedarfsgerechte Weiterentwicklung von RADAR als Forschungsdaten-Repository für das KIT

Felix Bach¹, Kerstin Soltau² und Matthias Razum²

¹Karlsruher Institut für Technologie (KIT);

²FIZ Karlsruhe - Leibniz-Institut für Informationsinfrastruktur

Das Karlsruher Institut für Technologie (KIT) baut aktuell ein institutionelles Forschungsdaten Repository auf Basis eigener Infrastruktur und des am Steinbuch Centre for Computing (SCC) betriebenen Speicherdienstes bwDataArchive auf. Als Management-Schicht und Benutzeroberfläche soll für die Forschenden die RADAR-Software unter der Bezeichnung RADAR4KIT zum Einsatz kommen und in das Dienstportfolio für das Forschungsdatenmanagement (FDM) der Hochschule integriert werden. Dabei soll das Management von Forschungsdaten sowohl durch die von der RADAR-Software bereitgestellten Dienstleistungen und -merkmale als auch durch die optimale Integration in die Dienstlandschaft des KIT vereinfacht werden.

Das KIT hat jedoch, neben der effektiven Einbindung der KIT-eigenen Infrastruktur, noch weitere institutionsspezifische Anforderungen an RADAR, die dessen bisheriger Funktionsumfang nicht abdeckt. Der Beitrag beschreibt die bedarfsgetriebene Weiterentwicklung der RADAR-Software und die Integration von RADAR für das KIT. Alle neu entwickelten RADAR-Funktionalitäten sind dabei nicht KIT-spezifisch ausgerichtet, sondern werden auch anderen Institutionen zur Verfügung stehen. Die beschriebenen Anpassungen stellen eine Öffnung RADARs für neue, alternative Einsatzszenarien dar und bedingen eine Weiterentwicklung des RADAR-Geschäftsmodells, der Dienstleistungsverträge und der Leistungsbeschreibung.

1. Einleitung

Das KIT¹ baut aktuell ein institutionelles Forschungsdaten-Repository auf Basis eigener Infrastruktur und des am SCC² betriebenen Speicherdienstes bwDataArchive³ auf. Als Management-Schicht und Benutzeroberfläche soll für die Forschenden die RADAR-Software unter der Bezeichnung RADAR4KIT zum Einsatz kommen und in das FDM-Dienstportfolio der Hochschule integriert werden. Das Ziel dabei ist es, das Management von Forschungsdaten für die Wissenschaftlerinnen und Wissenschaftler am KIT durch optimale Integration in die vorhandene Dienstlandschaft des KIT zu vereinfachen. Hierzu

¹ <https://www.kit.edu>

² <https://www.scc.kit.edu>

³ <https://www.rda.kit.edu>

können Diensteigenschaften von RADAR wie beispielsweise das Rollen- und Rechtekonzept, das für die Langzeitarchivierung am KIT geeignete Datenmanagement, die Möglichkeit der Datenpublikation mit DOI-Vergabe, die Integrationsmöglichkeiten via API und das disziplinübergreifende Metadatenschema einen entscheidenden Beitrag leisten.

Das KIT hat neben der effektiven Einbindung der eigenen Speicherinfrastruktur noch weitere institutionsspezifische Anforderungen an RADAR, wie z.B. die Verwendung eines eigenen DOI-Prefix, die Möglichkeit einer institutionellen Sicht auf eigene Forschungsdatensätze und Anpassungsmöglichkeiten der Plattform an das Corporate Design des KIT.

Diese Anforderungen werden bisher noch nicht vollständig durch den bestehenden Funktionsumfang von RADAR und dessen Betriebsmodell abgedeckt. Die notwendigen Anpassungen und Erweiterungen der RADAR-Software, die zukünftig auch allen anderen Institutionen zur Verfügung stehen, werden in diesem Beitrag beschrieben.

2. Anforderungen des KIT

2.1. Integration eines Forschungsdaten-Repositoriums in die FDM-Dienste des KIT

Das KIT hat 2016 eine Forschungsdaten-Policy⁴ verabschiedet, in der es sich zu einem verantwortungsvollen und nachhaltigen Umgang mit Forschungsdaten verpflichtet. Wissenschaftler und Wissenschaftlerinnen sollen durch geeignete Infrastrukturen und Dienste beim FDM unterstützt werden. Hierzu wurde das Serviceteam RDM@KIT⁵ ins Leben gerufen, das Forschenden Support entlang des Forschungsdaten-Zyklus bietet - insbesondere zu FDM-Diensten, Metadaten, Forschungsdaten (FD)-Archivierung und dem Aufbau von Repositorien und elektronischen Laborjournalen (engl. electronic lab notebooks, ELN).

Die bislang am KIT verfügbaren FDM-relevanten Dienste (siehe Abbildung 1) decken bereits die Bedarfe der Forschenden in den meisten Phasen des FD-Zyklus gut ab. So unterstützt bereits bei der Planung des FDM neuer Projekte der sog. Research Data Management Organizer (RDMO)⁶ und es gibt ein breites Informationsangebot auf den Webseiten des Serviceteams RDM@KIT und auf [forschungsdaten.info](https://www.forschungsdaten.info)⁷. Fachspezifische Labore, ELN und Virtuelle Forschungsumgebungen (VFU) unterstützen das Sammeln und die Analyse von Daten. Speicher- und Archivsysteme wie [bwDataArchive](https://www.bwdataarchive.org/) stehen zur Verfügung, das Teilen und Zusammenarbeiten an Dokumenten wird zumindest für kleinere Dokumente und Programm-Quellcode durch Dienste wie [bwSync&Share](https://bwsyncandshare.kit.edu)⁸ und [GitLab](https://git.scc.kit.edu/KIT)⁹ möglich. Außerdem stehen einige fachspezifische Repositorien mit erweitertem Funktionsumfang bereit (z.B. [Chemotion](https://www.chemotion.net)¹⁰), die in aktuellen Projekten (z.B. Scientific Data Center MoMaF)¹¹

⁴ <https://www.rdm.kit.edu/downloads/KIT-FDM-Policy.pdf>

⁵ <https://www.rdm.kit.edu>

⁶ <https://rdmo.forschungsdaten.info>

⁷ <https://www.forschungsdaten.info>

⁸ <https://bwsyncandshare.kit.edu>

⁹ <https://git.scc.kit.edu/KIT>

¹⁰ <https://www.chemotion.net>

¹¹ https://www.kit.edu/kit/pi_2019_021_molekuel-und-materialforschung-daten-leicht-teilen.php

für eine Nutzbarkeit durch benachbarte Fachbereiche weiterentwickelt werden. Des Weiteren gewährleisten zentrale Nachweis- und Suchsysteme den Zugriff auf archivierte FD, die zusammen mit Text-Publikationen¹² abgelegt werden. Was dem KIT allerdings momentan noch fehlt, ist eine Möglichkeit für Forschende, Forschungsdaten - auch große Datenmengen - mit anderen zu teilen oder zentral zu publizieren, zusammen mit entsprechenden Metadaten und referenzierbar über eine DOI.

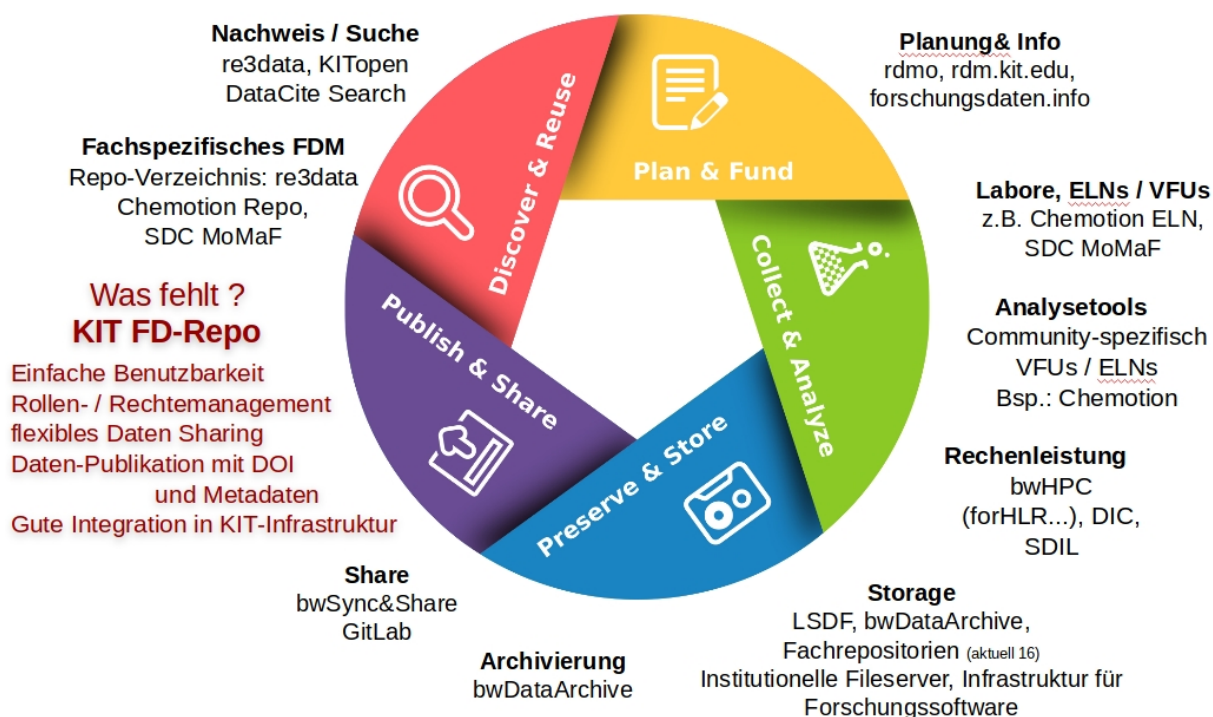


Abbildung 1.: Die FDM-Dienstlandschaft am KIT entlang des FD-Zyklus.

Diese fehlende Funktionalität bietet das generische Forschungsdaten-Repository RADAR¹³. Jedoch benötigt das KIT den Dienst RADAR nicht genau so, wie er aktuell durch FIZ Karlsruhe - Leibniz-Institut für Informationsinfrastruktur angeboten wird (siehe Leistungsbeschreibung¹⁴). Dies hat im Wesentlichen zwei Gründe: Das KIT möchte zum einen ein eigenes, an das Corporate Design angepasstes fachbereichübergreifendes Forschungsdaten-Repository, das nur Datensätze von KIT-Wissenschaftlern enthält und präsentiert. Und zum anderen betreibt das KIT eigene, groß angelegte Speicher-Infrastrukturen und -Dienste, die in einem institutionellen Forschungsdaten-Repository so genutzt und integriert werden sollen, dass einerseits effiziente Datenflüsse ermöglicht werden und andererseits die Daten ausschließlich am KIT gespeichert werden.

¹² <https://www.bibliothek.kit.edu/cms/kitopen.php>

¹³ <https://www.radar-service.eu>

¹⁴ https://www.radar-service.eu/sites/default/files/Dienstbeschreibung_RADAR.pdf

2.2. Anbindung von RADAR4KIT an die KIT-Infrastruktur

RADAR bietet bislang als “All-In-One” Cloud-Dienst für Hochschulen und außeruniversitäre Forschungseinrichtungen eine generische Infrastruktur zur langfristigen Archivierung und Publikation digitaler Forschungsdaten, ohne dass diese eigene Infrastruktur betreiben müssen. Dies ist vor allem für kleinere Einrichtungen von Vorteil, für die sich oft der Betrieb von Repositorien an eigenen Rechenzentren nicht lohnt, passt jedoch nicht optimal für das KIT, das mit dem SCC ein vollständiges Informationstechnologiezentrum betreibt, welches den Ansprüchen der zahlreichen Großgeräte, die am KIT betrieben werden und den großen Datenmengen, die am KIT verarbeitet werden müssen, gerecht wird.

Die RADAR-Software wurde in einem DFG-Projekt¹⁵ (2013-2016) von einem interdisziplinären Projektkonsortium bestehend aus fünf Forschungseinrichtungen entwickelt, das sowohl Community-Vertreter als auch Infrastruktureinrichtungen umfasste. Das KIT war durch das SCC selbst vertreten, wodurch die Bedarfe des KIT schon bei der Entwicklung berücksichtigt werden konnten. Des Weiteren hostet das SCC den größten Teil des RADAR-Dienstes von FIZ Karlsruhe, so dass die primäre IT-Infrastruktur für dessen Betrieb durch das SCC gestellt wird und optimal an die dortigen Speichersysteme angebunden ist. Damit lag das Aufsetzen und Betreiben einer eigenen KIT-Instanz von RADAR (RADAR4KIT) nahe, die ausschließlich eigene technische Infrastruktur verwendet, jedoch in Sachen Nachhaltigkeit davon profitiert, dass eine Weiterentwicklung der RADAR-Software durch FIZ Karlsruhe gewährleistet ist.

Um zu verdeutlichen, wie RADAR4KIT sich in die Speicherinfrastrukturen am KIT integriert, ist in Abbildung 2 die IT-Architektur am SCC zu sehen, auf der u.a. der Dienst bwDataArchive basiert. Der Landesdienst bwDataArchive¹⁶ bietet Wissenschaftlern und Wissenschaftlerinnen Zugang zu einer technischen Infrastruktur zur langfristigen Datenarchivierung, die insbesondere für Universitäten und öffentliche Forschungseinrichtungen aus Baden-Württemberg zur Verfügung gestellt wird. Die Datenarchivierung erfolgt am KIT und umfasst eine verlässliche Speicherung auch großer Datenbestände für einen Zeitraum von zehn oder mehr Jahren. Der Dienst ermöglicht eine qualifizierte Umsetzung der Empfehlungen der Deutschen Forschungsgemeinschaft (DFG) zur Sicherung und Aufbewahrung von Forschungsdaten.

Der Aufbau von bwDataArchive wurde in einem Landesprojekt vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (MWK) gefördert und arbeitete im Rahmen des Helmholtz-Projekts LSDMA eng mit verschiedenen wissenschaftlichen Communities sowie den internationalen Projekten EUDAT, Human Brain Project (HBP) und dem World Wide LHC Computing Grid (WLCG) zusammen.

Im unteren Bereich von Abbildung 2 ist das Speicherbackend zu sehen, welches das High Performance Storage System (HPSS)¹⁷ verwendet, das von der HPSS Collaboration und IBM entwickelt wurde. HPSS ist ein skalierbares, hierarchisches Storage Management (HSM) System, das eine Kombination aus einem schnellen, zentralen Plattenspeicher und langsameren, örtlich verteilten Bandspeichersystemen nutzt. Dieses Backend wird

¹⁵https://www.radar-service.eu/sites/default/files/publications/Abschlussbericht_DFG-Projekt_RADAR_Vero%CC%88ffentlichung.pdf

¹⁶ <https://www.rda.kit.edu>

¹⁷ <http://www.hpss-collaboration.org>

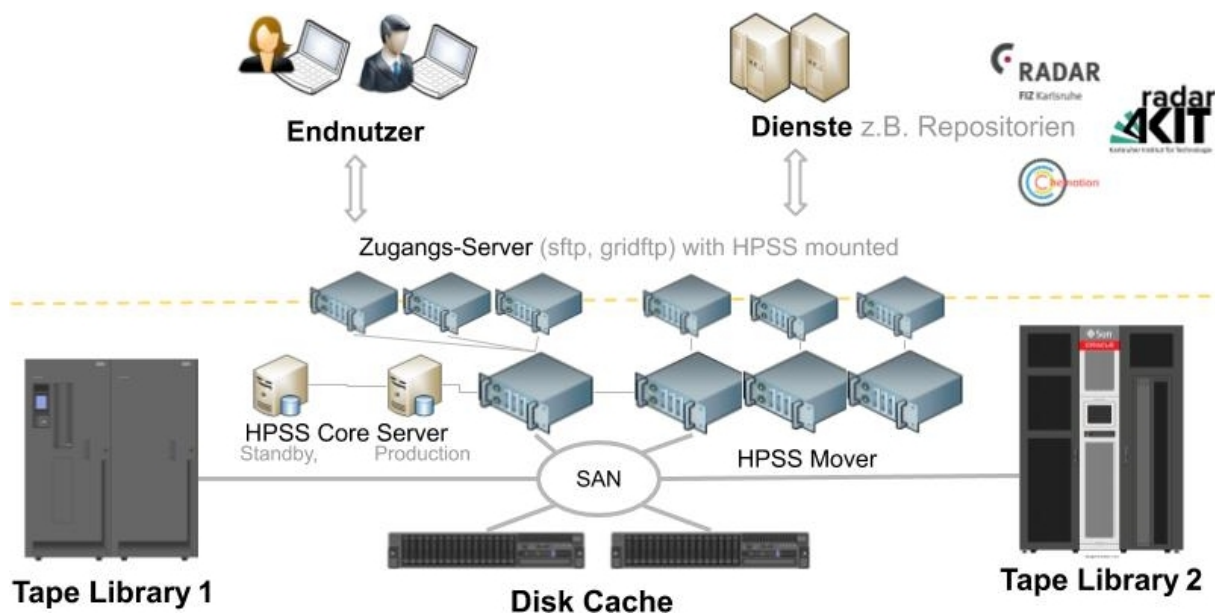


Abbildung 2.: Die bwDataArchive-Infrastruktur und aufsetzende Dienste.

auf mehreren Zugangsservern gemountet, die den Speicher über die Protokolle SFTP und GridFTP zugänglich machen. Hierüber können nun einerseits Endnutzer über den Dienst bwDataArchive zugreifen, als auch eigenständige Dienste wie etwa Repositorien. Momentan zählt zu diesen Diensten neben RADAR auch das Chemie-Repositorium Chemotion und zukünftig auch RADAR4KIT.

3. Forschungsdaten-Repositorium RADAR

Seit 2017 bietet RADAR akademischen Institutionen und Forschenden eine generische Infrastruktur zur langfristigen Archivierung und Publikation digitaler Forschungsdaten. RADAR wird von FIZ Karlsruhe - Leibniz-Institut für Informationsinfrastruktur¹⁸ als disziplinübergreifender "All-in-One" Cloud-Dienst angeboten 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 technische Infrastruktur von RADAR wird von FIZ Karlsruhe, dem SCC und dem Zentrum für Informationsdienste und Hochleistungsrechnen (ZIH) der TU Dresden¹⁹ bereitgestellt. Die Speicherung der Forschungsdaten erfolgt in drei Kopien an geographisch getrennten Standorten in den Rechenzentren des SCC (zwei Kopien) und des ZIH (zusätzliche Kopie). Die TIB Hannover²⁰ registriert DOIs für sämtliche über RADAR publizierte Forschungsdatensätze. Der komplette RADAR-Service und seine Infrastruktur unterliegen damit den rechtlichen Bestimmungen der Bundesrepublik Deutschland.

¹⁸ <https://www.fiz-karlsruhe.de>

¹⁹ <https://tu-dresden.de/zih>

²⁰ <https://www.tib.eu>

3.1. RADAR Dienstleistungen und Geschäftsmodell

Das RADAR-Angebot umfasst drei zentrale Dienstleistungen:

Die **Datenarchivierung** dient der sicheren und formatunabhängigen Aufbewahrung von Forschungsdaten über flexibel wählbare Haltefristen (5, 10, 15 Jahre). Die Forschungsdaten werden dabei in Form paketieter Zusammenstellungen gesichert und erhalten einen eindeutigen Identifier. Sofern von den Datengebern nicht anders vorgesehen, werden archivierte Forschungsdaten und zugehörige Metadaten nicht veröffentlicht. Das Teilen archivierter Datensätze mit anderen Nutzerinnen und Nutzern ist jedoch über eine flexible Zugriffsverwaltung möglich.

Bei der **Datenpublikation** werden die Datensätze für mindestens 25 Jahre gesichert. Jedes publizierte Datenpaket erhält einen persistenten Identifier (DOI), wird 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 der Datensatz eindeutig und dauerhaft identifizierbar, zitierfähig und kann mit wissenschaftlichen Publikationen verknüpft werden. Bei Bedarf können DOIs bereits vor der Datenpublikation reserviert werden. Optional kann die Datenpublikation über eine Embargofrist von bis zu einem Jahr verzögert werden. Für jedes publizierte Datenpaket muss die Datengeberin bzw. der Datengeber eine Lizenz (z.B. Creative Commons 4.0) vergeben, welche dessen Nachnutzung regelt.

Vor der Datenpublikation können im Rahmen eines **Peer-Review-Prozesses** externe Gutachterinnen und Gutachter, beispielsweise von Verlagen, die Forschungsdaten zudem über einen sicheren Link begutachten.

RADAR ist auf langfristigen Betrieb ausgelegt und operiert nicht gewinnorientiert. Bereits während der Projektphase wurde ein nachhaltiges **Geschäftsmodell** erarbeitet, das - nach einer fünfjährigen Anlaufphase - den Betrieb ohne Projektförderung sichern soll. FIZ Karlsruhe als Betreiber sieht die Archivierung und Publikation wissenschaftlicher Forschungsdaten als wichtigen Teil seines öffentlichen Auftrags und in Übereinstimmung mit dem eigenen Leitmotiv „Advancing Science“ - den gesamten wissenschaftlichen Wertschöpfungsprozess in allen Stufen, in denen Daten anfallen und Information und Wissen relevant sind, zu unterstützen. Daher übernimmt FIZ Karlsruhe die Hälfte der operativen Fixkosten aus der eigenen Grundfinanzierung. Der verbleibende Fixkostenanteil und alle variablen Betriebskosten sollen über Nutzungsgebühren eingenommen werden. Die Nutzung als Forschungsdaten-Repository für Institutionen setzt deshalb den Abschluss eines Dienstleistungsvertrags²¹ voraus, für den eine jährliche Grundgebühr anfällt. Die in Anspruch genommenen Dienstleistungen Datenarchivierung und Datenpublikation werden darüber hinaus nutzungsbasiert je nach angefallenem Datenvolumen in Rechnung gestellt²². Während für archivierte Daten innerhalb der gewählten Haltefrist jährlich Kosten berechnet werden, fallen für publizierte Daten nur im Jahr der Datenpublikation Kosten in Form einer Einmalzahlung an. Neben der Preistransparenz gewährt dieses Abrechnungsmodell nutzenden Einrichtungen zum einen die Möglichkeit, archivierte Daten nach Vertragsende flexibel an andere Dienstleister weiterzugeben und zum anderen eine von

²¹ https://www.radar-service.eu/sites/default/files/Dienstvertrag_RADAR.pdf

²² <https://www.radar-service.eu/de/preise>

der Vertragssituation unabhängige garantierte Haltefrist von 25 Jahren bei publizierten Daten.

3.2. RADAR Dienstmerkmale

RADAR unterstützt Einrichtungen beim Forschungsdatenmanagement, indem sich das System flexibel an institutionelle FDM-Workflows anpassen lässt. Zu den in diesem Zusammenhang wichtigsten Dienstmerkmalen zählen das disziplinübergreifende Metadaten-schema, die offene Systemarchitektur sowie das Rollen- und Rechtekonzept, welches die delegierte Administration durch eine nutzende Einrichtung erlaubt.

Das **RADAR Metadatenchema**²³ ist disziplinübergreifend angelegt, kompatibel mit dem DataCite Metadata Schema²⁴ sowie DublinCore²⁵ und fördert die Umsetzung der FAIR Prinzipien.²⁶ Es definiert 10 Pflichtfelder, die für die DOI-Registrierung des Forschungsdatensatzes notwendig sind, und 13 optionale Felder zur Beschreibung der Datenerhebung und -aufbereitung. Das Schema erlaubt die Verwendung von Normdaten für Personen (ORCID iD²⁷) und Förderorganisationen (CrossRef Open Funder Registry). Durch die Kombination aus kontrollierten Vokabularen und Freitext-Einträgen ermöglicht RADAR die Interoperabilität der beschriebenen Forschungsdaten und trägt gleichzeitig der Heterogenität der Daten aus einer Vielzahl von Disziplinen Rechnung. Neben der Beschreibung auf Ebene des Forschungsdatensatzes erlaubt RADAR auch die Beschreibung mit Metadaten auf Datei- und Verzeichnisebene.

Die **RADAR-Systemarchitektur** (Abbildung 3) ist modular aufgebaut und besteht aus dem User Interface (Frontend), der Management-Schicht (Backend) und der Speicherschicht (Archiv), welche das OAIS-konforme Langzeitarchivierungssystem implementieren. 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 beziehungsweise parallel betrieben werden können. Anwendungsfälle umfassen beispielsweise den Betrieb eines institutioneigenen Frontends, das automatisierte Hochladen von Forschungsdaten oder die Übertragung beziehungsweise den Abruf von (Meta-)Daten aus anderen Anwendungen. Über das Frontend können Benutzer und Rollen verwaltet, die institutionseigene RADAR-Umgebung administriert und gestaltet, Arbeitsbereiche eingerichtet, Daten hochgeladen, zu Datenpaketen zusammengestellt und mit Metadaten versehen sowie archiviert bzw. publiziert werden. Die Management-Schicht umfasst die gesamte Geschäftslogik von RADAR. Sie unterteilt sich in den Ingest Service und das Repository Management. Der Ingest Service nimmt einzelne Dateien und gepackte Archive (z.B. ZIP, gZIP, TAR) entgegen, entpackt sie und erzeugt ein neues Datenpaket. Eine eventuell im Dateiarchiv bereits vorhandene Verzeichnisstruktur wird dabei auch im neuen Datenpaket abgebildet. Das Repository Management verwaltet die für RADAR

²³ <https://www.radar-service.eu/de/radar-schema>

²⁴ <https://schema.datacite.org>

²⁵ <http://dublincore.org>

²⁶ <https://www.force11.org/group/fairgroup/fairprinciples>

²⁷ <https://orcid.org>

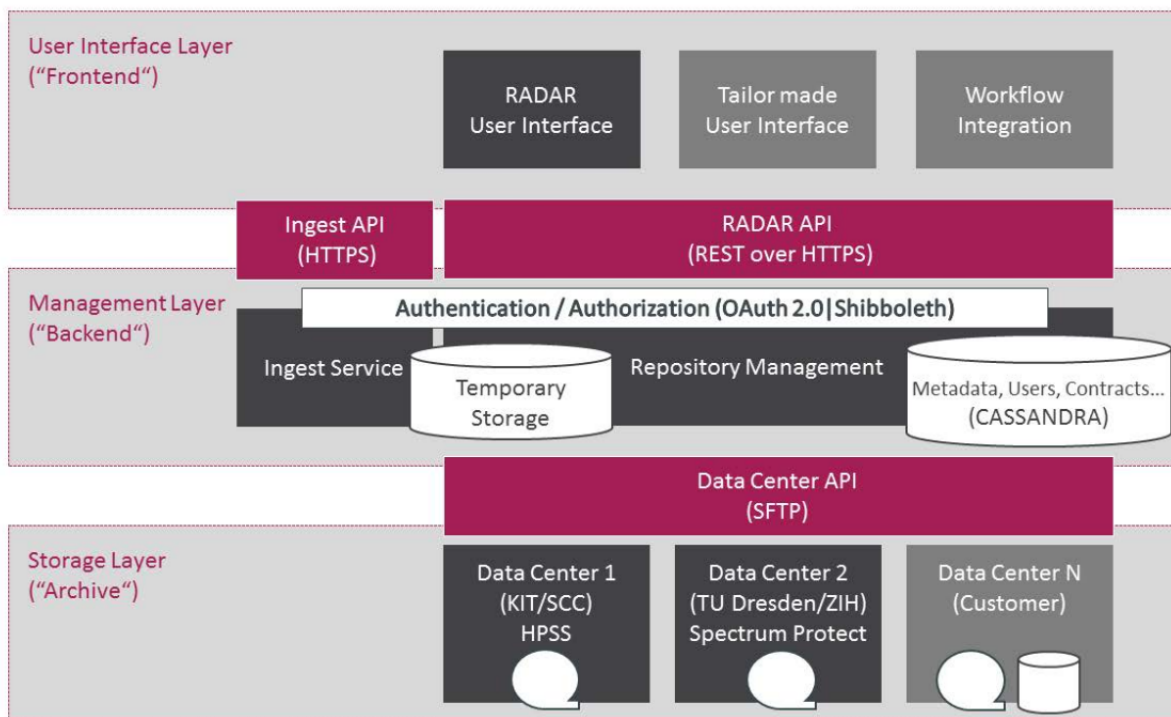


Abbildung 3.: Die RADAR Systemarchitektur.

wichtigen Entitäten wie Verträge, Arbeitsbereiche, Datenpakete, Verzeichnisse und Dateien. Darüber hinaus enthält es Funktionen zum Validieren von Metadaten, zum Archivieren bzw. Publizieren von Datenpaketen, zur Erstellung von Statistiken und zur Überwachung von Quotas. Auch der Lebenszyklus der Daten innerhalb des RADAR-Workflows mit den vier möglichen Stati "in Bearbeitung", "in Begutachtung", "Archiviert" und "Publiziert" wird von der Management-Schicht implementiert. Das **RADAR Rollen- und Rechtemodell** (Abbildung 4) ermöglicht die delegierte Administration durch die nutzende Einrichtung. Entsprechend der institutionellen Bedürfnisse können Arbeitsprozesse strukturiert, Aufgaben verteilt und interne Verantwortlichkeiten definiert werden. Von der Einrichtung eingesetzte Administratorinnen und Administratoren verwalten die RADAR-Arbeitsbereiche, die als zentrale Einstiegspunkte für Forschende eines Projekts oder einer Arbeitsgruppe dienen. Administratorinnen und Administratoren können Kuratorinnen und Kuratoren bestimmen, die 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, jedoch weder archivieren noch publizieren können. Über eine integrierte Nutzerregistrierung wird die Nutzung von RADAR durch Forschende administrativ erleichtert. Die Authentifizierung in RADAR ist sowohl über eine lokale Datenbank als auch über delegierte Verfahren wie z.B. Shibboleth möglich. Auch ein gemischter Betrieb ist denkbar. Sofern die nutzende Einrichtung an DFN-AAI²⁸ teilnimmt, erfolgt die Authentifizierung mit der institutionellen Nutzerkennung.

²⁸ <https://www.aai.dfn.de>

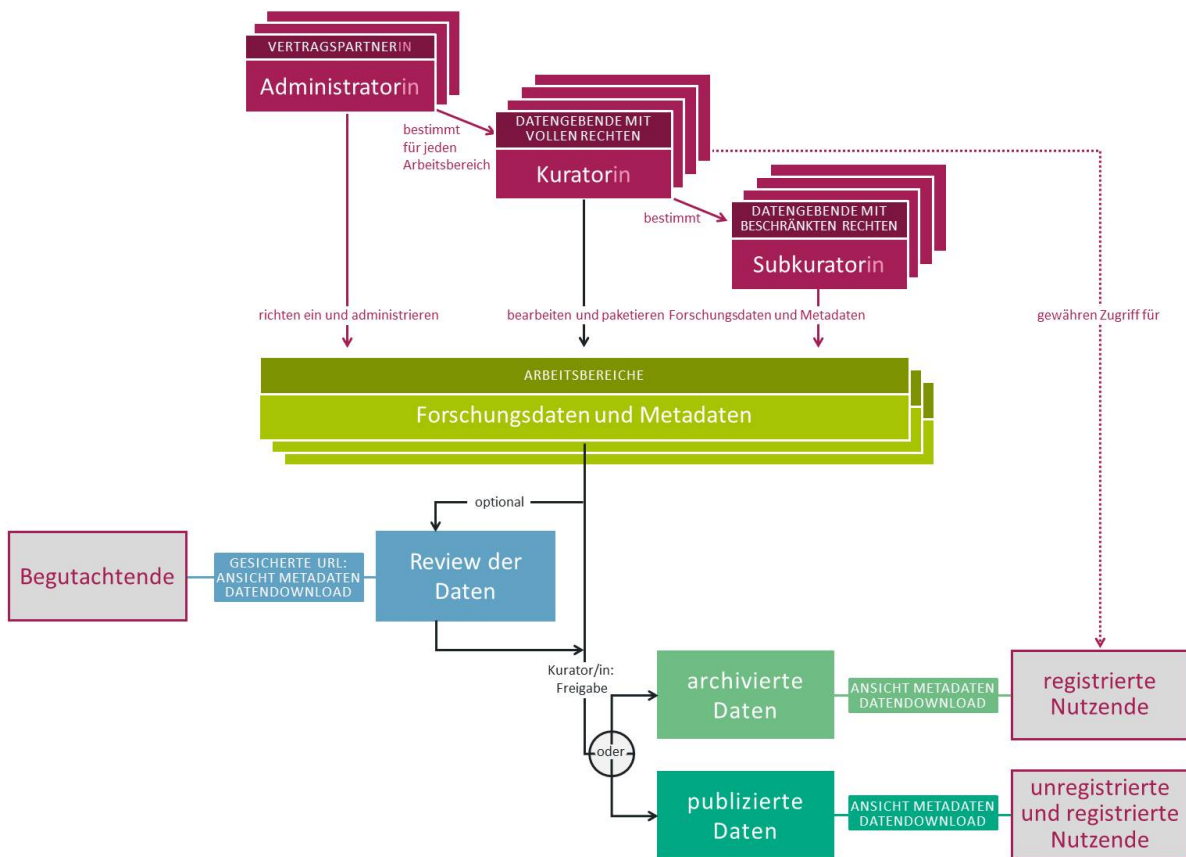


Abbildung 4.: Das RADAR Rollen- und Rechtemodell.

4. Anpassungen für RADAR4KIT

Die genannten Dienstmerkmale sind – neben dem besonderen Verhältnis von KIT und FIZ Karlsruhe als ehemalige RADAR-Projektpartner sowie der Rolle des SCC als Bereitsteller der primären IT-Infrastruktur für RADAR – ausschlaggebend für das KIT, sich beim Aufbau des eigenen Forschungsdaten-Repositoriums für die RADAR-Software zu entscheiden. Die bisherigen Möglichkeiten zur einrichtungsspezifischen Anpassung der Software decken jedoch nicht alle vom vom KIT gewünschten Aspekte ab.

4.1. Neue Anforderungen für einrichtungsspezifische Anpassungen

Aus den oben genannten Gründen sollen weitere institutionsspezifische Anpassungen im engen Austausch miteinander realisiert werden, die ein passgenaues Forschungsdaten-Repositorium für das KIT auf Basis eigener Infrastruktur ermöglichen. Dazu zählen:

- die automatische Vergabe von DOIs für publizierte Datensätze mit einem KIT-eigenen Präfix

- Branding-Optionen, die die Anpassung der Benutzeroberfläche an das KIT-Corporate Design ermöglichen, beispielsweise die Einbindung eines KIT-eigenen Logos sowie die Definition einer alternativen primären Schmuckfarbe
- der Betrieb der RADAR4KIT-Software unter einer eigenen URL auf KIT-eigener Domain
- die Möglichkeit, von der RADAR4KIT-Benutzeroberfläche auf verschiedene KIT-eigene Unterstützungsangebote zu verweisen
- die Möglichkeit einer institutionellen Sicht, über die standardmäßig ausschließlich die vom KIT publizierten Datensätze angezeigt werden
- die Verfügbarkeit der Metadaten als eigenes Set über den OAI-Provider von RADAR zum Harvesting durch Dritte
- die ausschließliche Verwendung KIT-eigener Speicherinfrastruktur, wie in 4.2 beschrieben.

4.2. Anpassungen auf Ebene der RADAR-Speicherschicht

Die technische Infrastruktur der Speicherschicht von RADAR als “All-In-One” Cloud-Dienst wird aktuell von den beiden Rechenzentren bereitgestellt, die derzeit RADAR-Daten hosten. Die Forschungsdaten sind dabei redundant in mehreren Kopien (zwei am SCC, eine weitere am ZIH) an unterschiedlichen, geographisch verteilten Standorten abgelegt, was die Zuverlässigkeit von RADAR erhöht. RADARs Speicherschicht ist über die Data Center API gekapselt und “verbirgt” so die von den Rechenzentren eingesetzten Technologien zur dauerhaften Speicherung der Datenpakete vor der Management-Schicht. Die RADAR-Systemarchitektur garantiert dadurch nicht nur die Unabhängigkeit von einer speziellen Speichertechnologie, sondern schafft gleichzeitig auch die Möglichkeit, weitere Rechenzentren in die technische Infrastruktur einzubinden oder bestehende zu ersetzen. Die Data Center API ermöglicht es somit auch, pro Kunde die zu verwendenden Rechenzentren festzulegen und eine oder alle Datenkopien im eigenen Rechenzentrum (sei es als Tape oder als Disk) zu verwahren.

Für RADAR4KIT wird auf dieser Ebene das Archiv des ZIH als Backup-Speicher deaktiviert. Das KIT wird seine Forschungsdaten somit ausschließlich auf dem vor Ort betriebenen Speicherdienst bwDataArchive in zwei Kopien an unterschiedlichen Standorten verwahren, daneben jedoch das RADAR-Frontend, die RADAR-Managementschicht und die RADAR API regulär nutzen.

4.3. Auswirkungen auf Leistungsbeschreibung, Vertrag und Vergütung

Vor allem die beschriebene Modifikation der RADAR-Systemarchitektur für RADAR4KIT beinhaltet nicht nur technische Herausforderungen. Das neue, alternative Betriebszenario hat auch Auswirkungen auf die Leistungsbeschreibung, den RADAR-Dienstleistungsvortrag inklusive seiner Haftungsregelungen und die Vergütung.

So entfällt beispielsweise im Gegensatz zum gehosteten “All-In-One” Cloud-Dienst bei RADAR4KIT die dritte Kopie der Daten am ZIH, aber auch die Haftung von FIZ Karlsruhe für die gespeicherten Daten. Aus diesem Grund werden Dienstvertrag und Preisgestaltung, basierend auf einer jährlichen Pauschalvergütung, individuell zwischen FIZ Karlsruhe und KIT verhandelt. Die Vergütung bezieht sich dabei auf die Leistungen Softwareentwicklung, den Betrieb und die Pflege der RADAR4KIT-Instanz durch FIZ Karlsruhe sowie das Repository-Management.

5. Fazit und Ausblick

Alle unter 4.1 und 4.2 beschriebenen Anpassungen für RADAR4KIT werden aktuell von FIZ Karlsruhe – in enger Abstimmung mit dem Service-Team RDM@KIT – spezifiziert, entwickelt und umgesetzt. Für die institutionelle Lösung kommen dabei ausschließlich eigene Server und Speicherdienste des KIT zum Einsatz und bwDataArchive wird als Archivierungslösung verwendet. Das Forschungsdaten-Repository RADAR4KIT wird von FIZ Karlsruhe im Auftrag administriert und weiterentwickelt. Gleichzeitig kann das KIT jedoch eine eigene Plattform im Corporate Design anbieten und Datensätze über RADAR4KIT mit einem eigenen DOI Prefix publizieren.

RADAR4KIT wird im Anschluss sukzessive in die Dienste des KIT integriert werden, v.a. an die bereits etablierten Nachweis- und Speichersysteme. Langfristig soll z.B. ein effizientes Management von Datensätzen ermöglicht werden, die bereits in bwDataArchive archiviert wurden. Das RDM@KIT-Team wird außerdem in Schulungen und Support den Nutzerinnen und Nutzern am KIT RADAR4KIT näherbringen.

Alle vom KIT gewünschten funktionalen Anpassungen wurden so in die Software integriert, dass sie zukünftig auch anderen interessierten Einrichtungen zur Verfügung stehen und sich in Teilen auch im “All-In-One” Cloud-Dienst finden. Hiervon können Institutionen profitieren, die weitergehende einrichtungsspezifische Anpassungen oder eine Anbindung eigener Infrastruktur bzw. eine Integration mit Diensten wie bwDataArchive benötigen. Auch für Forschende an Einrichtungen, die bereits bwDataArchive nutzen, könnte sich das Datenmanagement vereinfachen. Die Einbindung des eigenen Rechenzentrums in die RADAR-Speicherschicht kann zukünftig entweder exklusiv oder ergänzend zu einem der bestehenden RADAR-Datenzentren realisiert werden. Die neue Option der Unterstützung lokaler Datenkopien könnte für Institutionen oder Konsortien attraktiv sein, die bereits in eigene Speicherkapazität investiert haben und daher deren Einbindung in Repositorien anstreben. Dies räumt Einrichtungen mehr Kontrolle über die eigenen Forschungsdaten ein und könnte, aufgrund des Wegfalls variabler Gebühren, zu einer Kostensenkung des gesamten Forschungsdatenmanagements führen und dieses somit befördern.