

Forschungsdatenrepositorium mit einem kooperativen Betriebsmodell

Fallstudie

*Gabriel Schneider, Kommunikations-, Informations-, Medienzentrum (KIM), Universität Konstanz
Matthias Landwehr, Kommunikations-, Informations-, Medienzentrum (KIM), Universität Konstanz*

Zusammenfassung

Datenrepositorien gehören inzwischen zum Kernangebot vieler Bibliotheken. In diesem Kontext sind kooperative Betriebsmodelle eine selten genutzte, aber sehr attraktive Möglichkeit. Durch die Arbeitsteilung und die Trennung von technischem Betrieb und inhaltlicher Betreuung kann so auch bei einer begrenzten Ressourcenlage ein angemessenes Ergebnis erlangt werden. Für Konzeption, Implementierung und Betrieb eines Repositoriums bedarf es qualifiziertes Fachpersonal sowohl auf technischer als auch auf bibliothekarischer bzw. kuratierender Seite. Angesichts des gegenwärtigen Fachkräftemangels sollte daher kritisch hinterfragt werden, ob jede Bibliothek „das Rad neu erfinden“ muss. Das Kommunikations-, Informations-, Medienzentrum (KIM) der Universität Konstanz hat sich deshalb für eine kooperative Umsetzung mit FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur (FIZ Karlsruhe) entschieden.

So wurde das Datenrepositorium KonDATA der Universität Konstanz basierend auf dem von FIZ Karlsruhe betriebenen Dienst RADAR realisiert. Mit „RADAR Local“ stellt FIZ Karlsruhe eine anpassbare Repositoriums-Lösung bereit, die lokal an der Universität betrieben wird. Pflege und Wartung der Software werden dabei remote von Mitarbeiter*innen des Leibniz-Instituts erledigt. Dadurch ist vor Ort kein technisches Personal notwendig, sodass sich das KIM auf die bibliothekarischen Kernkompetenzen bei der Kuratierung der Daten fokussieren kann. Dieser Artikel beschreibt den Bedarf der Universität Konstanz, den Auswahlprozess bis zur geeigneten Lösung und wie das gewählte kooperative Betriebsmodell ausgestaltet und umgesetzt wurde. Anschließend werden das Betriebsmodell, seine Vor- und Nachteile sowie seine langfristigen Auswirkungen diskutiert.

Summary

Nowadays, data repositories are a part of the core services of many libraries. In this context, cooperative operating models are a rarely used but very attractive option. Because of the division of labour and the separation of technical operation and content support, an adequate result can be achieved even with limited resources. The design, implementation, and operation of a repository requires qualified staff on both the technical and library/curation side. In the face of the current shortages of qualified staff, it should be critically questioned whether every library has to “reinvent the wheel”. The Communication, Information, Media Center (KIM) of the University of Konstanz therefore opted for a cooperative implementation with the FIZ Karlsruhe – Leibniz Institute for Information Infrastructure.

KonDATA, the data repository of the University of Konstanz, is based on the RADAR service operated by FIZ Karlsruhe. With “RADAR Local”, FIZ Karlsruhe provides a customizable repository solution that is operated locally at the university. Maintenance of the software is done remotely by employees of the Leibniz Institute. This means that no technical staff is required on site, so that the KIM can

focus on its core library competencies in curating the data. The article describes the requirements of the University of Konstanz, the selection process leading up to an appropriate solution, and how the selected cooperative operating model was designed and implemented. The operating model, its advantages and disadvantages, and its long-term effects are also discussed.

Zitierfähiger Link (DOI): <<https://doi.org/10.5282/o-bib/5732>>

Schneider, Gabriel: ORCID: <<https://orcid.org/0000-0001-6573-3115>>;

Landwehr, Matthias: ORCID: <<https://orcid.org/0000-0001-9274-2578>>

Schlagwörter: Forschungsdatenmanagement, Repositorium, RADAR

Dieses Werk steht unter der [Lizenz Creative Commons Namensnennung 4.0 International](#).

1. Einleitung

Das Publizieren von Forschungsdaten ist im Zuge der Open-Science-Bewegung in den letzten Jahren immer üblicher geworden.¹ Vorgaben von Drittmittelgebern, eine wachsende Repositorienlandschaft, eine einfache Zitierbarkeit von Datensätzen und die Anerkennung von Forschungsdaten als wissenschaftliches Leistungsmerkmal führen dazu, dass immer mehr Forscher*innen diese Daten publizieren.² Durch nationale und internationale Initiativen, wie die Nationale Forschungsdaten Infrastruktur (NFDI)³ oder die European Open Science Cloud (EOSC)⁴, rückt das Thema „offene Forschungsdaten“ immer stärker ins Blickfeld. Das Konzept FAIRe Daten⁵ und dessen Prinzipien werden, je nach wissenschaftlicher Fachdisziplin, immer stärker verankert und sind in einigen Disziplinen zum Quasi-Standard erwachsen, wodurch auch die Anforderungen steigen, die an Hochschulen, wissenschaftliche Bibliotheken und Rechenzentren gestellt werden. Entsprechend gehört das Bereitstellen eines institutionellen Forschungsdatenrepositoriums mittlerweile zum Serviceangebot und zum Selbstverständnis vieler wissenschaftlicher Bibliotheken.

Für die Datensätze von Wissenschaftler*innen sollte immer ein geeignetes fachliches Repositorium die erste Wahl sein. Dort werden die Daten in einem thematisch passenden Ökosystem publiziert, die Repositorien sind in ihrer jeweiligen Disziplin anerkannt und die Sichtbarkeit der Daten ist am höchsten. Findet sich kein geeignetes Repositorium oder sind die Aufnahmebedingungen für die

1 Universität Konstanz, [forschungsdaten.info: Forschungsdaten veröffentlichen?](https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/daten-publizieren/entscheidungshilfe-daten-veroeffentlichen/), 05.08.2020, <<https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/daten-publizieren/entscheidungshilfe-daten-veroeffentlichen/>>, Stand: 14.09.2021.

2 Pampel, Heinz; Elger, Kirsten: Publikation und Zitierung von digitalen Forschungsdaten, in: Putnings, Markus; Neuroth, Heike; Neumann, Janna (Hg.): *Praxishandbuch Forschungsdatenmanagement*, Berlin, Boston 2021, S. 521–536, <<https://doi.org/10.1515/9783110657807-028>>.

3 Nationale Forschungsdateninfrastruktur (NFDI) e.V.: Homepage, <<https://www.nfdi.de/>>, Stand: 14.09.2021.

4 European Open Science Cloud. EOSC Portal: Homepage, 2021, <<https://eosc-portal.eu/>>, Stand: 14.09.2021.

5 Universität Konstanz, [forschungsdaten.info: FAIRe Daten](https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/), 30.04.2021, <<https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/>>, Stand: 14.09.2021.

Datensätze nicht erfüllbar, kann im zweiten Schritt auf generische Dienste wie Zenodo⁶ zurückgegriffen werden. Sind auch diese Dienste keine Option, weil z.B. die Datenmenge zu groß ist, der Wunsch nach einer Veröffentlichung unter dem Namen der Heimatinstitution besteht, ein niedrigschwelliges Angebot gewünscht wird oder eine externe Speicherung aus rechtlichen Gründen nicht möglich ist, bieten sich institutionelle Forschungsdatenrepositorien als Publikationsort an.

Für die Institution bringt ein eigenes Repositorium einige Vorteile mit sich. Durch die Kontrolle über die veröffentlichten Datensätze kann man die Qualität der Metadaten vorgeben und sicherstellen, dass die Daten im eigenen Haus bleiben. Auch der Publikationsprozess kann unabhängig von Dritten gestaltet und gesteuert werden. Zudem erhöht es die Sichtbarkeit der Institution und verbessert den Überblick über die publizierten Datensätze einer Einrichtung.

Allerdings verlangt der Betrieb eines Forschungsdatenrepositoriums ausreichende Ressourcen: (1) Das System muss entwickelt, aufgesetzt, gewartet und gepflegt werden. Für diese Aufgaben benötigt die Institution dauerhaft qualifiziertes IT-Fachpersonal. (2) Darüber hinaus muss das System im laufenden Betrieb betreut und Nutzer*innen bei der Datenpublikation und -archivierung unterstützt werden. Dafür bedarf es qualifizierte Datenkurator*innen. (3) An technischen Ressourcen werden passend dimensionierte Server und redundant vorhandener Speicherplatz benötigt. (4) Organisatorisch müssen diese Ressourcen langfristig garantiert werden, um eine dauerhafte Bereitstellung der publizierten Daten zu gewährleisten.

Um einen Teil der Ressourcen einzusparen, kommen Betriebskonzepte wie „Software-as-a-Service“ infrage, also das Einkufen eines Systems, das von einem externen Dienstleister in einer Cloud betrieben wird. Durch das Auslagern der IT kann sich die Institution auf das Betreuen der Nutzer*innen und das Kuratieren konzentrieren. Anzumerken ist allerdings, dass die Daten bei dieser Lösung auf institutionsfremder Infrastruktur gespeichert werden, was zu rechtlichen Problemen führen kann. Wenn Daten physisch das Haus – und bei Cloud-Lösungen ggf. das Land bzw. die Europäische Union – verlassen, dann müssen diese Prozesse bezüglich der Auswirkungen auf Datenschutz und Urheberrecht ausführlich geprüft werden. Hier kann es u.a. Konflikte mit der Datenschutzgrundverordnung geben. Es sei auch auf das vom Europäischen Gerichtshof für ungültig erklärte Privacy-Shield-Abkommen zwischen der EU und den USA hingewiesen.⁷ Vor diesem Hintergrund wird im Folgenden anhand von KonDATA, dem Forschungsdatenrepositorium der Universität Konstanz, eine alternative Lösung vorgestellt, bei der die IT-Dienstleistung zwar eingekauft, aber trotzdem lokal gehostet wird.

6 Zenodo: Homepage, <<https://zenodo.org/>>, Stand: 14.09.2021.

7 Uphues, Steffen, Ins Wasser gefallen, Privacy Shield für Datenübermittlungen über den Atlantik in die USA ungültig, in: Verein zur Förderung eines Deutschen Forschungsnetzes e. V. (Hg.): DFN Infobrief Recht (8) 2020, S. 2–4, <https://www.dfn.de/fileadmin/3Beratung/Recht/1infobriefarchiv/2020/Infobrief_Recht_08-2020.pdf>, Stand: 14.09.2021.

2. Ausgangslage

Die Universität Konstanz ist eine forschungsstarke Universität, die seit 2007 konstant in allen Exzellenz-Förderlinien erfolgreich war. Ein Schwerpunkt der aktuellen Exzellenzstrategie ist das Thema E-Science und der Ausbau der Dienstleistungen rund um Open Science und Forschungsdatenmanagement. Umgesetzt werden die Maßnahmen durch das KIM, dem Zusammenschluss von Bibliothek und Rechenzentrum, in enger Zusammenarbeit mit anderen Einrichtungen der Universität. Zu diesem Zweck wurde im KIM ein Team Open Science institutionell verankert.⁸

Über die letzten Jahre hat sich in Beratungssituationen herausgestellt, dass es an der Universität Konstanz den Bedarf nach einem eigenen institutionellen Forschungsdatenrepositorium gibt. Auch wenn in den meisten Fällen der Bedarf der Wissenschaftler*innen zur Publikation von Forschungsdaten mit einem Verweis auf einschlägige fachliche oder generische Repositorien befriedigt werden kann, wünschen sich nach wie vor manche Wissenschaftler*innen eine Veröffentlichung im eigenen Haus. Der Fokus liegt dabei auf der Publikation der Daten, denn für deren interne und langfristige Archivierung ist bereits ein anderer Dienst vorhanden. Neben dem konkreten Bedarf der Wissenschaft kommt hinzu, dass der Betrieb eines eigenen Forschungsdatenrepositoriums mittlerweile zum Selbstverständnis vieler Hochschulen und der daran angebotenen Informationseinrichtungen gehört. Das KIM betreibt seit Jahren erfolgreich das Publikationsrepositorium KOPS⁹ und möchte mit KonDATA einen vergleichbaren Dienst für Datenpublikationen etablieren, der Mitgliedern der Universität ein niedrigschwelliges Angebot zur Datenpublikation bietet. KonDATA soll das Serviceangebot komplettieren und wurde deshalb nicht als über die Grenzen der Institution hinauswirkender „Leuchtturm“-Dienst konzipiert.

2.1. Anforderungsanalyse

Für die Anforderungsanalyse wurde zuerst im Auftrag der Direktion eine Arbeitsgruppe zusammengestellt, die gemeinsam die Anforderungen festlegte, einen Kriterienkatalog erstellte, anhand von diesem eine Evaluierung durchführte und mithilfe der Ergebnisse eine Entscheidung treffen sollte. Das Team bestand aus den für Forschungsdatenmanagement zuständigen Mitgliedern des Team Open Science, dem Sachgebietsleiter für Software-Entwicklung und der Betreuerin von KOPS. Die Anforderungen wurden aufgeteilt in grundlegende, die erfüllt werden müssen und wünschenswerte, die zusätzlich umgesetzt werden könnten. Die Erfüllung der als grundlegende identifizierten Anforderungen bei gleichzeitig geringem Ressourceneinsatz war dabei richtungweisend.

Die ermittelten Anforderungen wurden nach technischen und organisatorischen Gesichtspunkten kategorisiert: Die Universität Konstanz benötigte ein generisches Datenrepositorium zur Speicherung und Publikation von Datensätzen. Das Repositorium musste über ein abgestuftes Rollen- und Rechtemanagement verfügen und die Metadaten mussten eine Flexibilität aufweisen, die es ermöglicht, bestimmte Metadaten nur für bestimmte Datenarten oder wissenschaftliche Disziplinen verpflichtend

8 Universität Konstanz, Kommunikations-, Informations-, Medienzentrum (KIM): Open Science, <<https://www.kim.uni-konstanz.de/openscience/>>, 2021, Stand: 14.09.2021.

9 Universität Konstanz, KOPS. Das institutionelle Repositorium der Universität Konstanz: Homepage, <<https://kops.uni-konstanz.de/>>, Stand: 14.09.2021.

zu machen. Die Metadatenstruktur sollte sich zudem an gängigen Standards orientieren und erweiterbar sein. Bei der Publikation mussten persistente Identifikatoren, konkret Digital Object Identifier (DOIs), vergeben werden können. Auch eine Embargofrist für publizierte Datensätze musste möglich sein. Der so entstandene Kriterienkatalog wurde auch in einer landesweiten Arbeitsgruppe der Science Data Center in Baden-Württemberg diskutiert.¹⁰

Abseits der funktionalen Anforderungen sollte der Dienst selbstverständlich benutzerfreundlich, barrierefrei und auf dem aktuellen Stand der Softwaretechnik sein. Der Wartungs- und Pflegeaufwand sollte so gering wie möglich ausfallen, gleichzeitig musste eine hinreichende Individualisierung in Bezug auf Corporate Design und Branding möglich sein. Da der Dienst einen Basischarakter haben sollte, wurden keine ausgefallenen Sonderanforderungen aufgestellt. Diskutierte, aber verworfene Punkte waren z.B. eine automatische Anbindung an wissenschaftliche Prozesse mit elektronischen Laborbüchern oder umfangreiche Schnittstellen zu anderen Diensten. Ein anderes Beispiel war eine frei konfigurierbare Oberfläche mit umfangreichen Vorschaumöglichkeiten und basalen Funktionen zur Analyse der Datensätze.

Um den bereits angesprochenen Bedarf nach einer lokalen Speicherung abzudecken, kam die Nutzung eines ausschließlich extern in der Cloud betriebenen Angebots nicht infrage. Der Betrieb auf einer hauseigenen Hardware war ein zwingendes Kriterium.

2.2. Entscheidungsfindung

Die Entscheidungsfindung gliederte sich in zwei Schritte auf. Die naheliegende Lösung im universitären Umfeld, so auch an der Universität Konstanz, ist oft die Eigenentwicklung eines entsprechenden Dienstes. Diese Option verwarf das KIM im Falle von KonDATA. Durch die Entwicklung und den Betrieb von KOPS hat das KIM langjährige Kompetenz mit DSpace als Repositoriums-Software aufgebaut. Eine Integration von KonDATA in KOPS wurde abgelehnt, da die unterschiedlichen Anforderungen – KonDATA als einfacher Dienst mit Basisfunktionalität und KOPS als umfangreicher und etablierter Dienst – eine komplexe technische Lösung mit vielen Kompromissen erfordert hätte. Erschwerend kam hinzu, dass eine mögliche Integration erst nach einem Softwareupdate von KOPS auf die Version DSpace 7 hätte stattfinden können.

Bei den eingesetzten Ressourcen musste der personelle und finanzielle Aufwand berücksichtigt werden und in einem angemessenen Verhältnis zur Bedeutung des Dienstes stehen. Die Arbeitsgruppe kalkulierte für KonDATA mit einem Mindestbedarf von sechs Personenmonaten für die Einführung und vier Personenmonaten pro Jahr für die Weiterentwicklung und den Betrieb. Eine integrierte Lösung mit KOPS hätte den personellen Aufwand deutlich erhöht. Dafür reichen jedoch die vorhandenen Entwicklungsressourcen nicht aus. An dieser Stelle zeigt sich der Fachkräftemangel im IT-Bereich öffentlicher Einrichtungen; die Rekrutierung fähiger Entwicklungskräfte ist sehr schwierig. In Konstanz kommt die Randlage zur Schweiz erschwerend hinzu, denn in unmittelbarer Nähe bieten Privatunternehmen attraktivere Gehälter.

¹⁰ Universität Konstanz, [forschungsdaten.info: Science Data Center](https://www.forschungsdaten.info/fdm-im-deutschsprachigen-raum/baden-wuerttemberg/science-data-center/), 02.03.2021, <<https://www.forschungsdaten.info/fdm-im-deutschsprachigen-raum/baden-wuerttemberg/science-data-center/>>, Stand: 14.09.2021.

Deshalb wurde die grundlegende Entscheidung getroffen, keinen eigenen Dienst zu entwickeln und stattdessen auf der Grundlage des Kriterienkatalogs eine Marktrecherche durchzuführen. Die angebotenen Lösungen von Figshare¹¹ und RADAR Cloud¹² ermöglichen den Betrieb eines eigenen Repositoriums auf der jeweiligen externen Plattform. Diese Angebote schieden allerdings aus, da mit ihnen die Daten nicht an der Universität Konstanz gespeichert worden wären. Dataverse¹³ und Fedora¹⁴ hätten als Repositoriums-Software die Anforderungen erfüllt, allerdings wäre auch hier lokaler Implementations- und Pflegeaufwand angefallen. Zusätzlich hätte damit neben DSpace die Kompetenz für eine zweite Repositorien-Technologie am KIM aufgebaut werden müssen. Am Ende dieses Prozesses fiel die Entscheidung zugunsten der von FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur angebotenen Repositoriums-Lösung RADAR Local.¹⁵ Diese Lösung erfüllte die definierten Anforderungen und die aufgerufenen Kosten. Zudem wurde der benötigte personelle Aufwand auf Seiten des KIM im Vergleich zu einer Eigenentwicklung oder einer anderen Lösung als deutlich günstiger bewertet.

3. RADAR Local

RADAR ist ein Forschungsdatendienst von FIZ Karlsruhe. Ursprünglich als Cloud-Dienst konzipiert, können RADAR-Kunden die All-in-One-Cloud-Lösung nutzen, um ihren Wissenschaftler*innen ein niedrigschwelliges Angebot zur Veröffentlichung von Forschungsdaten zu machen, ohne dafür institutionelle Hardware-Ressourcen bereitstellen zu müssen. RADAR wurde von 2013 bis 2016 in einem DFG-geförderten Projekt entwickelt und wird seit 2017 in der Cloud-Variante angeboten.

Es hat sich jedoch gezeigt, dass eine ausschließliche Fokussierung auf Institutionen ohne ausreichende eigene Hardware als Zielgruppe nicht bedarfsgerecht ist. Aus diesem Grund bietet FIZ Karlsruhe zusätzlich zu RADAR Cloud zwei weitere Betriebsvarianten an, genannt RADAR Hybrid und RADAR Local.¹⁶ Diese Betriebsvarianten ermöglichen es Institutionen, in verschiedenen Graden eigene Hardware für den Betrieb des Repositoriums zu nutzen, ohne sich mit dem Aufsetzen, der Wartung, der Pflege und der Weiterentwicklung der Software beschäftigen zu müssen. So läuft beispielsweise bei RADAR Hybrid die Software immer noch in der ursprünglichen Cloud-Umgebung, die angebotenen Speicherlösungen werden jedoch von der nutzenden Institution selbst bereitgestellt.

RADAR Local geht noch einen Schritt weiter. Anstatt die Cloud-Installation zu verwenden, entsteht bei RADAR Local eine lokale, eigene Instanz der RADAR-Software auf der Infrastruktur der nutzenden Institution. Auch die Speichersysteme werden hierbei von der nutzenden Institution bereitgestellt.

11 Figshare: Homepage, <<https://figshare.com/>>, Stand: 14.09.2021.

12 FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur: RADAR Cloud, 2021, <<https://radar.products.fiz-karlsruhe.de/de/radarvariants/betriebsvarianten#radar+cloud+>>, Stand: 14.09.2021.

13 The Dataverse Project: Homepage, <<https://dataverse.org/>>, Stand: 14.09.2021.

14 Lyrisis: Fedora: Homepage, <<https://duraspace.org/fedora/>>, 2021, Stand: 14.09.2021.

15 FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur: Homepage, 2021, <<https://www.fiz-karlsruhe.de/>>, Stand: 14.09.2021.

16 FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur: Betriebsvarianten, 2021, <<https://radar.products.fiz-karlsruhe.de/de/radarvariants/betriebsvarianten>>, Stand: 14.09.21.

Entsprechend stellt das KIM für KonDATA virtuelle Maschinen (VMs) zur Verfügung, auf denen die Techniker*innen von FIZ Karlsruhe den nötigen Software-Stack für den Betrieb des Repositoriums installieren, vom Betriebssystem bis zur RADAR-Software, deren Anwendungen und Komponenten containerisiert als Docker Swarm vorliegen.¹⁷ Docker Swarm ist ein Werkzeug zur Verwaltung von in Docker-Containern vorliegenden Anwendungen, die auf unterschiedlichen Systemen im Verbund betrieben werden. An diese VMs werden via Network File System-Protokoll (NFS) und SSH File Transfer-Protokoll (SFTP) Speichersysteme angeschlossen, auf denen die Forschungsdaten veröffentlicht, archiviert und langfristig zur Verfügung gestellt werden.

Durch das Einrichten einer eigenen lokalen Instanz bietet RADAR Local ausgeprägtere Individualisierungsmöglichkeiten als die anderen RADAR-Betriebsvarianten. So kann eine eigene Startseite gestaltet werden, die sich im Domänenraum der Institution verorten lässt, die RADAR-Oberfläche selbst kann durch Auswahl einer Schmuckfarbe und die Integration eines Logos personalisiert werden. Das Angebot lässt sich also optisch an die anderen Dienste der Einrichtung anpassen. Weiterhin fügt sich RADAR Local auch auf der Ebene der Authentifizierung in die Systemlandschaft der nutzenden Institution ein, indem auf Shibboleth als Authentifizierungsverfahren zurückgegriffen wird. Shibboleth wickelt die Authentifizierung über das lokale Identitätsmanagement (IDM) der jeweiligen Einrichtung ab. So können Wissenschaftler*innen der nutzenden Institution ihre gewohnten Zugangsdaten zur Authentifizierung verwenden.

4. Betriebsmodell

Zwischen der Universität Konstanz und FIZ Karlsruhe wurde nach der Entscheidung für RADAR Local ein Dienstleistungsvertrag abgeschlossen,¹⁸ der Zuständigkeiten, Pflichten und Kosten festlegt. Die Mindestlaufzeit des Vertrags beträgt drei Jahre, die Universität Konstanz bezahlt FIZ Karlsruhe eine Pauschalvergütung von 5.500 € zzgl. MwSt. jährlich für das Aufsetzen und den technischen Betrieb der RADAR-Local-Instanz, wobei sich die Vergütung aus einem Grundbetrag von 5.000 € und dem Betrieb eines Testsystems für 500 € zusammensetzt.

4.1. Technische und organisatorische Aufgaben

Im Betriebsmodell von RADAR Local lassen sich die Zuständigkeiten für die einzelnen Aufgabenbereiche klar benennen und aufteilen. Im Fall von KonDATA stellt das KIM auf Seiten der Universität als nutzende Institution die nötige Infrastruktur und Hardware bereit. Dazu mussten einige Arbeiten auf technischer und organisatorischer Ebene vorgenommen werden: Die virtuellen Maschinen und Speichersysteme mussten beantragt und den Sicherheitskonzepten des KIM entsprechend in angemessene Netzstrukturen integriert werden. Virtuelle Maschinen werden auf der etablierten virtuellen Infrastruktur der Universität Konstanz erstellt, die auf VMware basiert. Für den Speicher kommt ein räumlich redundant ausgelegtes Storage-System zum Einsatz, das den nötigen Speicherplatz über

¹⁷ Docker Inc.: Swarm mode overview, 2021, <<https://docs.docker.com/engine/swarm/>>, Stand: 14.09.2021.

¹⁸ FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur: Verträge & Preise, 2021, <<https://radar.products.fiz-karlsruhe.de/de/radaragreementsprices/vertraege-preise>>, Stand: 14.09.2021.

die definierten Schnittstellen zur Verfügung stellt. Diese Storage- und VM-Infrastruktur ist nicht exklusiv für KonDATA. Wartung und Betrieb werden durch das KIM geleistet.

Um den Entwickler*innen des FIZ Karlsruhe den Zugriff von außerhalb des Universitätscampus zu ermöglichen, waren entsprechende Konfigurationen der Firewalls und das Einrichten von Virtual-Private-Network-Zugängen (VPN) nötig. Damit konnten die Entwickler*innen von FIZ Karlsruhe Betriebssysteme und Software installieren und diese für die Nutzung konfigurieren. Das RADAR-Team übernimmt weiterhin die Pflege des gesamten Software-Stacks und steht auch zukünftig dem KIM als Ansprechpartner bei technischen Anfragen zur Verfügung.

4.2. Datenkuratierung

Durch diese Aufteilung kann das KIM sich auf seine Kernkompetenzen konzentrieren: die Betreuung der Wissenschaftler*innen und die Kuratierung der Daten. Letztere umfasst dabei sowohl eine technisch-formale wie auch eine inhaltliche Kuratierung. Formal wird zum Beispiel darauf geachtet, ob Metadaten vollständig und im korrekten Format angegeben wurden oder ob die Forschungsdaten in einem lesbaren Dateiformat vorliegen und sich öffnen lassen. Die Kuratierung erfolgt in enger Abstimmung mit der datengebenden Person und bei der inhaltlichen Kuratierung wird dies noch deutlich intensiviert. Sie umfasst den Blick in die Daten, die Überprüfung der Bezeichnung der Variablen, die Korrektheit von angegebenen Wertebereichen oder allgemein die Plausibilität der zu publizierenden Daten. Dies kann immer nur begrenzt geschehen, da das Personal des KIM nicht über die notwendigen Kenntnisse in der Tiefe in jeder Fachdisziplin verfügen kann. Trotzdem ist die Kuratierung eine wertvolle Notwendigkeit, um die Qualität der Daten zu sichern und zu erhöhen.

4.3. Weiterentwicklung von RADAR

RADAR ist eine dynamische Software, die stetig weiterentwickelt wird. Als nutzende Institution ist die Universität Konstanz beim Abschluss des Vertrags Mitglied im RADAR-Nutzerbeirat geworden. Dieser setzt sich aus nutzenden Institutionen der verschiedenen RADAR-Betriebsvarianten zusammen. Der Beirat diskutiert und priorisiert Weiterentwicklungen der Software. Als nutzende Institution besitzt die Universität Konstanz also ein Mitbestimmungsrecht, um an der Weiterentwicklung der Software mitzuwirken und eigene Anforderungen einzubringen. Sollte abseits dieser vom Beirat getroffenen Entscheidung eine Funktion in der Software benötigt werden, besteht zusätzlich die Möglichkeit, diese Entwicklung separat in Auftrag zu geben und zu finanzieren.

4.4. Rechtliche Rahmenbedingungen

Neben den angesprochenen technischen Arbeiten fallen auf Seiten der nutzenden Einrichtung weitere organisatorische Arbeiten beim Aufsetzen einer lokalen RADAR-Instanz an. So mussten im Fall von KonDATA mehrere rechtliche Dokumente erarbeitet werden. Dazu zählt ein Eintrag ins Verzeichnis für Verarbeitungstätigkeiten (VVT) der Universität Konstanz und der Abschluss eines Auftragsverarbeitungsvertrags (AVV) mit FIZ Karlsruhe, da beim Betrieb des Repositoriums personenbezogene Daten gesammelt und verarbeitet werden und somit deren datenschutzkonformer Umgang geregelt sein muss. FIZ Karlsruhe ist Auftragsverarbeiter im Sinne der DS-GVO. Der zusätzlich zum Dienstleistungsvertrag gemäß Art. 28 DS-GVO abzuschließende AVV liegt als Anhang dem Vertragstext bei. Weiterhin stellt die nutzende Institution als Anbieterin von RADAR Local (im Sinne

des Telemediengesetzes) sowie als deren Verantwortliche (im Sinne der DS-GVO) die Nutzungsbedingungen des Diensts bzw. die Datenschutzerklärung bereit. Dadurch entstehen Gestaltungsräume bei der Formulierung der entsprechenden Dokumente. Beide Dokumente wurden im Fall von KonDATA mit dem Justizariat der Universität abgestimmt.

5. Vor- und Nachteile des Betriebsmodells

Die Umsetzung eines Forschungsdatenrepositoriums im vorgestellten Betriebsmodell bringt sowohl Vor- als auch Nachteile mit sich. Sehr positiv ist für die Universität Konstanz, dass sie als nutzende Institution keine eigenen Entwicklungsressourcen bereitstellen muss, da diese Arbeiten von FIZ Karlsruhe übernommen werden. Lediglich das initiale Aufsetzen der Infrastruktur benötigt IT-Kenntnisse im Bereich von VMs, Speicherlösungen, Netzwerken, etc. Der personelle Aufwand ist gering und konnte in die bestehende Betreuung von VM- und Storage-Kunden integriert werden. Zudem ist weniger spezialisiertes Personal notwendig. Die Administration von virtuellen Maschinen und Storage kann ohne tiefe Kenntnis eines Dienstes erfolgen und personelle Abgänge können somit wesentlich leichter aufgefangen werden, als es bei einer Eigenentwicklung der Fall wäre. RADAR Local (oder ein anderes RADAR-Produkt) stellt der nutzenden Institution ein Repositorium zur Verfügung, das auf „State-of-the-Art“-Software basiert und stetig weiterentwickelt wird. Im Gegensatz zu anderen erhältlichen Produkten kann bei RADAR die nutzende Institution als Mitglied des Nutzerbeirats die Weiterentwicklung aktiv beeinflussen.

Das transparente Preismodell ermöglicht von vorneherein eine klare Berechnung der anfallenden Kosten. Die jährlichen Kosten sind dabei sehr attraktiv, vor allem, wenn man sie den Lohnkosten für IT-Personal entgegenstellt. Ein weiterer Vorteil ist auch, dass auf etablierte Authentifizierungsverfahren wie Shibboleth und eine DOI-Vergabe über DataCite zurückgegriffen werden kann. Da das KIM bereits Dienste betreibt, die Datensätze mit DOIs auszeichnen, kann KonDATA mit einem eigenen Präfix in das DataCite-Konto des KIM integriert werden. So lässt sich der Dienst auch auf dieser Ebene in bereits etablierte Prozesse einbinden.

Die Arbeitsteilung ermöglicht es dem KIM, sich auf seine Kernkompetenzen als bibliothekarische Einrichtung zu fokussieren. Diese liegen in der Betreuung und Beratung der Wissenschaftler*innen und in der Kuratierung der Datensätze. Dafür wurden Konzepte und Workflows entworfen, welche diese Prozesse nach definierten Abläufen organisieren.

Auf der anderen Seite muss man konstatieren, dass der Einfluss auf die zukünftige Entwicklung dennoch begrenzt ist. Eigene Anforderungen müssen mit den anderen Beiratsmitgliedern kommuniziert und diskutiert werden und können, je nach Ergebnis, in der Priorisierung der Weiterentwicklung eine untergeordnete Rolle spielen. Die Flexibilität einer Eigenentwicklung kann man zu keiner Zeit erwarten. Auch die begrenzten Individualisierungsmöglichkeiten sind als Nachteil zu sehen, da unklar ist, ob das System bei zukünftigen Anforderungen weiterhin zufriedenstellend sein wird.

Auch wenn RADAR Local kein Open-Source-Produkt ist, ist eine Entwicklung eigener Erweiterungen des Systems unter Nutzung der von RADAR zur Verfügung gestellten Programmierschnittstelle (RADAR API) möglich.¹⁹ Einzelne Komponenten von RADAR können durch eigene Implementierungen ausgetauscht oder parallel betrieben werden. Dies würde jedoch für die Universität Konstanz wieder zusätzlichen Entwicklungsaufwand bedeuten.

Sollte es zu technischen Problemen kommen, hat das KIM selbst keine direkten Eingriffsmöglichkeiten. Stattdessen müssen die Entwickler*innen von FIZ Karlsruhe auf das Problem aufmerksam gemacht werden. Dieser Punkt wird als wenig kritisch gesehen, da eine Hochverfügbarkeit des Systems nicht zwingend notwendig ist. Nichtsdestoweniger gibt sich das KIM generell in ein Abhängigkeitsverhältnis zu FIZ Karlsruhe, was beispielsweise die Exit-Strategie angeht. So müsste bei einer Migration in ein anderes System eine gemeinsame Lösung gefunden und erarbeitet werden. Vertraglich vereinbart ist eine Übergabe des Quellcodes zum Zeitpunkt des Vertragsendes. Damit könnte der Dienst in eingefrorenem Zustand weiterbetrieben werden. Für eine Übertragung der Daten in ein neues System wurde nicht geprüft, inwiefern das über die standardmäßigen Schnittstellen möglich ist.

In der Gesamtschau überwiegen für das KIM der Universität Konstanz aber klar die Vorteile der gewählten Lösung.

6. Fazit und Ausblick

RADAR Local erfüllt die ermittelten Anforderungen der Universität Konstanz an ein Forschungsdatenrepositorium. Die Lösung des kooperativen Betriebsmodells reduziert die für Universitäten und Hochschulen nötigen Ressourcen erheblich und bietet schnell und niedrigschwellig ausreichend Funktionalität und Individualisierungspotenzial. In der Praxis resultiert das kooperative Modell in einer engen Zusammenarbeit zwischen Techniker*innen der Universität Konstanz und von FIZ Karlsruhe. Die gemeinsame Arbeit am Dienst und die Lösung von Problemen funktioniert nach einigen Reibungsverlusten zu Beginn inzwischen sehr zufriedenstellend. Die Entscheidung für eine Umsetzung des Dienstes mit RADAR Local fiel im KIM im Herbst 2020. Mit der Implementierung des Dienstes wurde im Frühjahr 2021 begonnen und das Repositorium startete im dritten Quartal 2021 in den Pilotbetrieb.

Positiv zu bewerten ist der Aspekt der Nachhaltigkeit, da ein etabliertes Softwareprodukt genutzt wird und keine Eigenentwicklung dauerhaft betrieben werden muss. Das gewählte Modell wird von der Universität Konstanz auch als Blaupause für andere Dienste und Anforderungen gesehen. Dadurch lassen sich Situationen vermeiden, in denen aus guten und nachvollziehbaren Gründen auf selbstentwickelte Dienste gesetzt wurde, die sich aber zu einem späteren Zeitpunkt aufgrund von gestiegenen Anforderungen, geänderten technischen Rahmenbedingungen oder personeller Fluktuation nicht mehr wie zu Beginn betreiben lassen.

¹⁹ FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur: RADAR API, 2021, <<https://radar.products.fiz-karlsruhe.de/de/radarfeatures/radar-api>>, Stand: 14.09.2021.

Die Universität Konstanz startete den Produktivbetrieb des Dienstes im vierten Quartal 2021, es werden erste Datensätze lokaler Wissenschaftler*innen kuratiert und über KonDATA veröffentlicht sowie langfristig, mindestens für 10 Jahre, zur Verfügung gestellt. Sobald sich das System etabliert hat, wird über Schnittstellen eine Anbindung an andere Dienste der Universität Konstanz angestrebt. Dazu zählen das Publikationsrepositorium KOPS und das Forschungsinformationssystem. Hier wird es spannend zu beobachten sein, wie sich der Entwicklungsprozess von RADAR und die Zusammenarbeit zwischen der Universität Konstanz und FIZ Karlsruhe langfristig in der Praxis darstellen wird.

Literaturverzeichnis

- Pampel, Heinz; Elger, Kirsten: Publikation und Zitierung von digitalen Forschungsdaten, in: Putnings, Markus; Neuroth, Heike; Neumann, Janna (Hg.): Praxishandbuch Forschungsdatenmanagement, Berlin, Boston 2021, S. 521–536, <<https://doi.org/10.1515/9783110657807-028>>.
- Uphues, Steffen, Ins Wasser gefallen, Privacy Shield für Datenübermittlungen über den Atlantik in die USA ungültig, in: Verein zur Förderung eines Deutschen Forschungsnetzes e. V. (Hg.): DFN Infobrief Recht (8) 2020, S. 2–4, <https://www.dfn.de/fileadmin/3Beratung/Recht/1infobriefearchiv/2020/Infobrief_Recht_08-2020.pdf>, Stand: 14.09.2021.