

RADAR – Ein Forschungsdaten-Repository als Dienstleistung für die Wissenschaft



Foto: FIZ Karlsruhe, P. Schwarz

Matthias Razum



Foto: Inga Sommer, Hamburg

Janna Neumann



Foto: privat

Matthias Hahn

Träger der Informations-
infrastruktur

Die Nachvollziehbarkeit und Reproduzierbarkeit wissenschaftlicher Erkenntnisse basieren zunehmend auf digitalen Daten. Deren Publikation, Verfügbarkeit und Nachnutzung muss im Rahmen guter wissenschaftlicher Praxis gewährleistet werden. Das Projekt RADAR geht diese Herausforderung durch die Etablierung einer generischen Infrastruktur für die Archivierung und Publikation von Forschungsdaten an. Es richtet sich an Wissenschaftler¹ (z. B. in Projekten) und Institutionen (z. B. Bibliotheken) mit einem zweistufigen Ansatz: ein disziplinübergreifendes Einstiegsangebot zur formatunabhängigen Datenarchivierung mit minimalem Metadatensatz und ein höherwertiges Angebot mit integrierter Datenpublikation. Thematisch legt RADAR den Schwerpunkt auf die »small sciences«, in denen Forschungsdateninfrastrukturen meist noch fehlen. RADAR erlaubt eine temporäre oder – im Falle einer Datenpublikation – zeitlich unbegrenzte Datenarchivierung. Das angestrebte Geschäftsmodell zielt auf einen sich selbst tragenden Betrieb mit einer Kombination aus Einmalzahlungen und institutionellen Angeboten ab. RADAR ist als ein Baustein der internationalen Informationsinfrastruktur geplant, der sich über Schnittstellen auch in Dienste Dritter integrieren lässt.

The transparency and reproducibility of scientific findings are increasingly based on digital data. In compliance with good scientific practice data need to be published, accessible, and re-usable. The RADAR project has accepted this challenge and aims to establish a generic infrastructure, which will provide archiving and publication services for research data. Target groups are scientists (e.g., in projects) and institutions (e.g., libraries). Both groups are offered a two-stage approach with a cross-discipline starter package for format-independent data preservation with a minimum metadata set and a superior package for preserving data with integrated data publication. The thematic emphasis of RADAR is on the »small sciences«, which often lack sufficient research data infrastructure. The RADAR infrastructure will provide the opportunity of temporary or – in case of data publication – long-term preservation of research data. Using a self-supporting business model, RADAR will offer one-off payments and institutional subscription services. As such, RADAR is intended to become an integral part of the international information infrastructure which also allows the integration of third-party services.

MOTIVATION

Die Verfügbarkeit wissenschaftlicher Daten ist einer der Schlüsselfaktoren für die weitere Entwicklung der (empirischen) Wissenschaften. Die Nachvollziehbarkeit und Reproduzierbarkeit von wissenschaftlichen Ergebnissen (*reproducible research*, siehe etwa Stodden (1)) spielt dabei ebenso eine Rolle wie neuartige Ansätze des Erkenntnisgewinns, die Jim Gray als *Fourth Paradigm* bezeichnet (2). Ungeachtet der mittlerweile zahlreichen politischen Vorgaben und Richtlinien (z. B. durch die OECD (3), UNESCO², EU (4) oder die DFG (5)) ist jedoch nur ein geringer Teil der produzierten Daten tatsächlich verfügbar, da eine auf Dauerhaftigkeit angelegte und verlässliche Infrastruktur zur Erschließung, Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten in vielen Fachdisziplinen

fehlt. Gleichzeitig ist aber die verlässliche und dauerhafte Bereitstellung von Forschungsdaten eine wesentliche Voraussetzung für deren Publikations- und Zitierfähigkeit, die die Basis der oben genannten Ziele bildet. Die Veröffentlichung von Forschungsdaten sollte ein integraler Bestandteil des Forschungsprozesses sein (6), jedoch kennt die klassische wissenschaftliche Publikationsweise kaum Standards, wie mit den zugrundeliegenden Forschungsdaten zu verfahren ist. Forschungsdaten entstehen zu einem anderen Zeitpunkt als klassische Publikationen und durchlaufen einen eigenen »Lebenszyklus«. Zur Illustration dieser Prozesse im Forschungsdatenmanagement legen die Autoren das Domänenmodell von Treloar (7) und Klump (8) zugrunde (s. Abb. 1). Demzufolge werden die Forschungsdaten durch den Wissenschaftler (*private Domäne*) erzeugt und analysiert. Zur Diskussion der Ergebnisse mit ausgewählten Kollegen innerhalb und außerhalb seiner Institution macht er diese – meist in bereits bearbeiteter Form – über geeignete Systeme eingeschränkt zugreifbar (*kollaborative Domäne*). Mit der Veröffentlichung der Daten gehen diese in die öffentliche Domäne über, die für die Archivierung und langfristige Erhaltung sorgt. Mit diesem Übergang ist die sogenannte *curation boundary* zu überwinden, bei der insbesondere Selektions- und Erschließungsaufgaben zu erbringen sind – intellektuelle Prozesse, die einen hohen Arbeits- und Kostenaufwand bedeuten. Die vierte Domäne schließlich erlaubt den Zugriff auf die archivierten Daten, z. B. über Fachportale oder virtuelle Forschungsumgebungen.

Die Forschung alleine ist mit dem Aufbau, der Bereitstellung und dem dauerhaften Betrieb einer solchen Infrastruktur überfordert. Sie muss die Ziele und wesentlichen Eigenschaften bestimmen, kann aber keine auf Dauer angelegten Dienstleistungen erbringen. Dies trifft insbesondere auf die langfristige Aufgabe der Archivierung zu. Spätestens hier kommen die Träger der Informationsinfrastruktur ins Spiel: Bibliotheken, Archive, Rechenzentren, Dienstleister und Verlage. Vor diesem Hintergrund haben im Herbst 2012 FIZ Karlsruhe – Leibniz-Zentrum für Informationsinfrastruktur, die Technische Informationsbibliothek Hannover (TIB), das Steinbuch Centre for Computing (SCC) des KIT mit zwei wissenschaftlichen Partnern (dem Department für Chemie der LMU München und dem Leibniz-Institut für Pflanzenbiochemie in Halle) ihre

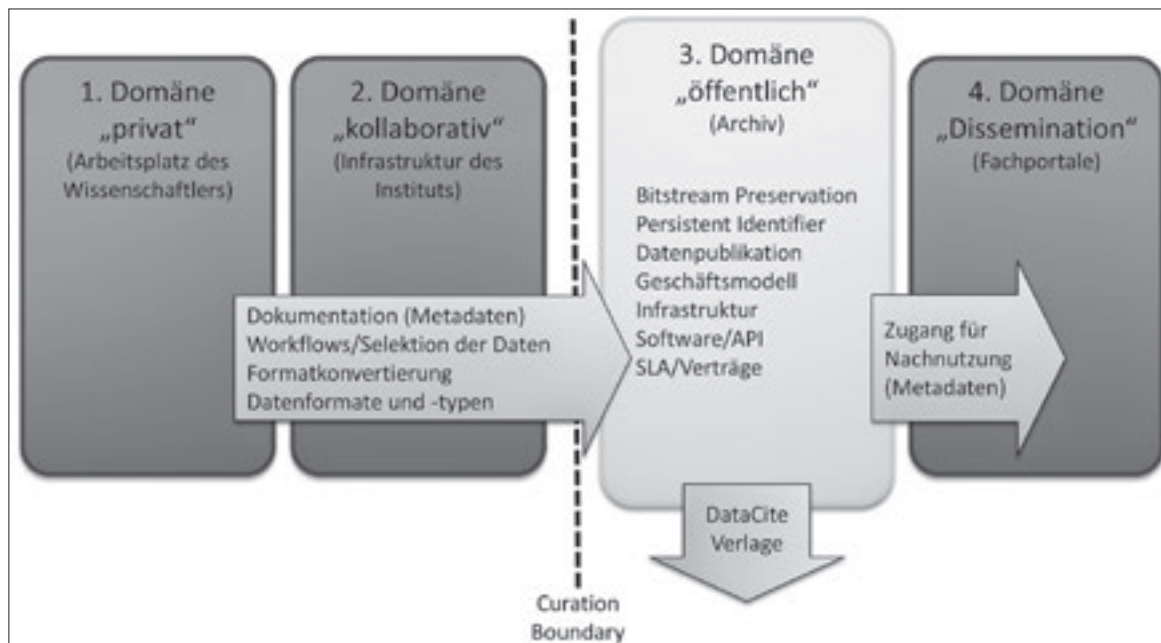


Abb. 1: Lebenszyklus von Forschungsdaten

Kompetenzen gebündelt und im Themenfeld 4 (»Forschungsnahe Informationsinfrastruktur«) des im 2012 veröffentlichten Förderprogramms der DFG zur »Neuausrichtung überregionaler Informationsservices« das Projekt RADAR (Research Data Repository) eingereicht. Ziel ist die Etablierung eines Daten-Repositoriums als Basisdienstleistung für Forscher und wissenschaftliche Institutionen zur Archivierung und Publikation ihrer Forschungsdaten. Die DFG bewilligte das Projekt mit einer Laufzeit von zunächst 24 Monaten, die sich nach einer erfolgreichen Zwischenevaluierung um ein weiteres Jahr verlängert. Die Arbeiten am Projekt wurden im September 2013 aufgenommen.

HERAUSFORDERUNGEN

Für die erfolgreiche Etablierung einer solchen Basisdienstleistung müssen die Projektpartner komplexe Herausforderungen angehen und mit dem geplanten Datenzentrum bestehende Defizite abmildern oder gar aufheben. Die folgenden vier Abschnitte zeigen diese auf.

Herausforderung 1: Infrastrukturbedarf

Sowohl international wie auch national gibt es erfolgreiche Beispiele für Forschungsdaten-Repositorien, etwa das World Data System der International Council of Science (ICSU)³ mit seinen mehr als 50 weltweit verteilten disziplinspezifischen Datenzentren oder GESIS – Leibniz-Institut für Sozialwissenschaften⁴ mit seinen archivierten Studien und empirischen Primärdaten aus den Sozialwissenschaften. Auch Fachdatenbanken wie z. B. die Protein Data Bank (PDB)⁵ nehmen

Forschungsdaten auf. Gerade letztere zeigt aber auch, dass diese existierenden Angebote oft nur einen Ausschnitt der insgesamt anfallenden Forschungsdaten erfassen. Um Nachvollziehbarkeit und Reproduzierbarkeit von Forschungsergebnissen zu gewährleisten, sollte ihr Entstehungsprozess nachvollziehbar erhalten bleiben, inklusive der Rohdaten und Zwischenergebnisse (9). Diese Daten nehmen die meisten disziplin-spezifischen Datenzentren aber nicht an. RADAR unterstützt bewusst die Archivierung und Bereitstellung solcher Daten über sein zweistufiges Angebot u. a. mit angepassten Qualitätsansprüchen an deren Erschließung. Noch problematischer aber ist, dass in manchen Disziplinen entsprechende Repositorien vollständig fehlen. Dieses Defizit korreliert oftmals mit kleineren Datenmengen, aber auch mit der in manchen Bereichen geringeren gesellschaftlichen Wahrnehmung dieser fachspezifischen Daten.

Disziplinen mit großen Datenmengen wie Astronomie, Hochenergiephysik oder Erdwissenschaften mussten sich frühzeitig mit der Etablierung von Dateninfrastrukturen und der Kompetenzentwicklung zu deren Aufbau und Betrieb auseinandersetzen. Dieser Druck fehlte bei anderen Disziplinen auch aufgrund der bisher noch handhabbaren Datenmengen – man spricht deswegen auch von den *small sciences* oder *long tail of science* (10). Dem Mangel an einer entsprechenden Forschungsdateninfrastruktur stehen die enormen Zuwächse an digitalen Forschungsdaten gegenüber; dies betrifft sowohl ihr Volumen als auch Art, Verwendung und Nachnutzungswert der Daten. Was letztlich relevante, d. h. erhaltenswerte Forschungsda-

RADAR = Research Data Repository

small sciences

ten sind, ist dabei häufig abhängig vom Kontext, der Zeit und nachfolgenden zukünftigen Erkenntnissen und Entdeckungen, kann also nicht notwendigerweise bereits zum Zeitpunkt ihres Entstehens durch den Forscher bewertet werden.

Herausforderung 2: Heterogenität

unterschiedliche Forschungsformen

In seinen Empfehlungen zur Weiterentwicklung der Informationsinfrastruktur in Deutschland bis 2020 unterscheidet der Wissenschaftsrat sechs Forschungsformen: experimentierende, beobachtende, auf Simulation beruhende, hermeneutisch-interpretierende, begrifflich-theoretische und gestaltende (11). Nicht für alle Forschungsformen besitzen (digitale) Forschungsdaten schon heute den gleichen hohen Stellenwert. Experimentelle Daten entstehen z.B. in Beschleunigern, chemischen, biologischen oder medizinischen Laboren und sind oft gekennzeichnet durch große Datenmengen und vielfach proprietäre Datenformate. Simulationsdaten entstehen durch Berechnungen mit Hilfe von Computern. Hier stellen eher die Ausgangsdaten sowie die verwendeten Simulationsalgorithmen und deren Parametrisierung die Herausforderung für die Archivierung dar. Beobachtungsdaten entstehen z.B. durch Teleskope, Mikroskope und Satelliten, aber auch durch Interviews oder durch Audio- und Videoaufzeichnungen und können oftmals nur schwer reproduziert werden, da es sich teilweise um singuläre Ereignisse handelt. Darüber hinaus sind als weitere Datengruppe Referenzdaten zu nennen, wie z.B. Genomdatenbanken, die disziplinübergreifend von Wissenschaftlern genutzt werden. Hinzu kommt begleitendes Material, wie Kalibrierungsdaten, Parameter des Experimentes, Laborbücher etc., die die eigentlichen Daten in einen Kontext setzen und damit erst verständlich machen. Die enorme disziplinabhängige Vielfalt der Datenformen und -typen ist kaum einzuschränken: sie reicht von proprietären Formaten über Audio- und Videostreams, Textkorpora bis hin zu Webarchiven⁶ und Datenbanken. Zu der großen Vielfalt an Datentypen und den zu ihrer Speicherung eingesetzten Datenformaten kommen noch auf die disziplinspezifischen Anforderungen hin abgestimmte, beschreibende Metadaten hinzu. Diese stellen die Vielzahl der Datenobjekte oft erst in einen Zusammenhang, machen ihre Genese nachvollziehbar und sie überhaupt erst auffindbar.

Wissenschaftler stellen auf ihrem jeweiligen Fachgebiet individuelle und höchstspezifische Anforderungen an das Forschungsdatenmanagement. Großforschungseinrichtungen wie z.B. CERN oder *Big Data*-Disziplinen wie beispielsweise die Klimaforschung pflegen einen anderen Umgang mit ihren Forschungs-

daten und sehen sich mit anderen Herausforderungen konfrontiert (12) als z.B. kleine Forschergruppen in der Chemie. Dementsprechend unterschiedlich ist der Grad der Standardisierung von Datenformaten und Metadatenformaten. Gleichwohl können vier Ansätze (13) in Hinblick auf den Umgang der unterschiedlichen Fachdisziplinen mit der Formatheterogenität unterschieden werden:

1. die Formate können durch Richtlinien der Institution oder
2. durch Fakten (z.B. Standards, Geräteabhängigkeiten) eingeschränkt sein,
3. es hat u.U. noch keine Standardisierung in dieser Fachdisziplin stattgefunden,
4. die Formate können nicht eingeschränkt werden, da dies z.B. innovationshemmend wäre.

Die enorme Vielzahl der Daten- und Metadatenformate stellt für die langfristige Nutzbarkeit der Daten, insbesondere für die funktionale Langzeitarchivierung, eine große Herausforderung dar.⁷ Die Beobachtung der sich entwickelnden Technologie der digitalen Langzeitarchivierung und der Anforderungen der relevanten Zielgruppen haben einen großen Einfluss auf notwendige *Preservation Policies* und Datenmanagementpläne. Viele dieser Fragestellungen sind zurzeit noch Gegenstand der Forschung (14). Um das Projekt nicht zu überfrachten, verzichtet RADAR vorerst bewusst auf eine funktionale Langzeitarchivierung der Daten, beobachtet aber relevante Projekte⁸ und Initiativen⁹ bzw. plant entsprechende Schnittstellen oder Kooperationsmöglichkeiten ein.

Herausforderung 3: Datenpublikation

Sollen Teile von Forschungsdaten Bestandteil einer wissenschaftlichen Publikation werden, sei es als Ergänzung zu einer klassischen Publikation (*supporting material*) oder als eigenständige Datenpublikation, so müssen sie zwei elementare Anforderungen erfüllen: eindeutige und persistente Identifizierbarkeit und verlässlicher dauerhafter Zugriff. Das Domänenmodell sieht hierfür den Übergang in die *öffentliche Domäne* vor. Doch was im Domänenmodell der logisch nächste Schritt ist, stellt Wissenschaftler und Verlage in vielen Disziplinen vor große Probleme. Wer übernimmt für sie die Identifizierung und Speicherung der Daten weit über die Dauer ihrer Forschungsprojekte hinaus? Wissenschaftliche Verlage widmen sich in jüngster Zeit selbst diesem Thema und suchen neue Wege der Publikation von Forschungsergebnissen und der zugrundeliegenden Daten. Darüber hinaus wurde eine Kooperation zwischen STM-Verlagen¹⁰ und DataCite vor Kurzem beschlossen.¹¹

disziplinabhängige Vielfalt der Datenformen

Herausforderung 4: Vertrauenswürdigkeit und Nachhaltigkeit

Ein verlässlicher und dauerhafter Zugriff auf Forschungsdaten muss von den Datenzentren garantiert werden (15). In der Vergangenheit wurden hierfür unterschiedliche Kriterienkataloge entwickelt und Zertifizierungsstellen geschaffen (z. B. World Data System¹², TRAC (16), nestor (17), DCC¹³, DINI (18) und Data Seal of Approval (19)). Dadurch erlangen Archive das Siegel der Vertrauenswürdigkeit. Die Harmonisierung der Kriterienkataloge und die Klärung von Zuständigkeiten sind jedoch noch offen.

Neben den technischen, organisatorischen und juristischen Aspekten des Forschungsdatenmanagements ist die wirtschaftliche Sicht ebenfalls wichtig. Dem ökonomischen Aspekt des Datenmanagements widmete sich zum ersten Mal umfassend der BRTF Final Report (20). Für den Bereich des Forschungsdatenmanagements spricht der Report fünf Empfehlungen für Institutionen und Geldgeber aus. Weitere Publikationen auf diesem Gebiet sind »Keeping Research Data Safe« (21) und »Aligning National Approaches to Digital Preservation« (22). Bis dato fehlen aber belastbare Zahlen und Geschäftsmodelle trotz vieler Untersuchungen und entsprechender Modelle.¹⁴ Für einen verlässlichen und dauerhaften Betrieb eines Datenzentrums ist trotzdem ein funktionierendes und möglichst selbsttragendes Geschäftsmodell unumgänglich, dass einerseits kompatibel mit Forschungsprojekten und deren zeitlich befristeter Finanzierung, andererseits aber unabhängig von der Notwendigkeit einer institutionellen Förderung und deren Unwägbarkeiten ist.

ZIELE VON RADAR

In einem Zeitraum von drei Jahren wollen die Projektpartner eine Infrastruktur für die Datenarchivierung und -publikation (Datenzentrum) mit einem nachhaltigen und selbsttragenden Geschäftsmodell aufbauen. Der Aufbau und die Etablierung einer solchen Infrastruktur umfasst Werkzeuge und Prozesse, um Forschungsdaten:

- systematisch zu erschließen,
- dauerhaft in einem Datenarchiv zu bewahren und der Öffentlichkeit zugänglich zu machen,
- durch DOI-Vergabe verfügbar, zitierfähig und verlinkbar zu machen sowie
- für die Nachnutzung qualitätsgesichert bereitzustellen.

RADAR geht die Herausforderungen in vielfältiger Weise an, versucht aber nicht, alle Probleme gleichzeitig zu lösen. Vielmehr beschränkt sich das Projekt bewusst auf einige Kernaufgaben und fügt sich ansonsten in

die breitere (inter-)nationale Informationsinfrastruktur ein.

Je mehr man sich dem Arbeitsplatz der Wissenschaftler annähert (s. Abb. 1, kollaborative bzw. private Domäne), desto offensichtlicher wird die Vielfalt der Forschungsprozesse und die Heterogenität der damit verbundenen Datentypen, Formate und Metadaten. Daher ist es zielführend, den Schwerpunkt von RADAR zuerst auf die dritte, öffentliche Domäne zu legen, da dies die einzige Stelle ist, an der sich eine generische, disziplinübergreifende Dienstleistung etablieren lässt. Gleichzeitig wirkt ein verlässliches und auf Dauerhaftigkeit ausgelegtes Datenarchiv dem Defizit einer fehlenden Infrastruktur entgegen. Das Datenarchiv bildet eine wesentliche Grundlage für die Nachnutzung und Publikation von Forschungsdaten und bringt allein dadurch bereits einen erheblichen Nutzen für Forschende (auch in der forschenden Industrie), Wissenschaft und Gesellschaft.

Die enge Zusammenarbeit mit Wissenschaftlern, Fachgesellschaften und Verlagen ermöglicht die bedarfsgerechte Entwicklung der Infrastruktur. Deswegen besteht das Konsortium hinter RADAR nicht nur aus Infrastruktureinrichtungen, sondern umfasst auch zwei wissenschaftliche Partner, die ihre disziplinspezifischen Anforderungen einbringen und die adäquate Umsetzung kontinuierlich evaluieren. Hier wurde mit der (pflanzlichen Bio-)Chemie eine Pilotdisziplin ausgewählt, in der einerseits der Bedarf an einem Datenarchiv besteht, andererseits aber sehr gut auf andere Disziplinen übertragbare Datentypen und -formate anfallen (NMR-Spektroskopie und 2D/DIGE-Bilder¹⁵). Mit dem Thieme Verlag und dem GDCh / VCH-Wiley Verlagsbeirat sowie der Gesellschaft Deutscher Chemiker (GDCh) konnten bereits wichtige Kooperationspartner für diese Disziplin gefunden werden.

RADAR zielt auf eine disziplin-übergreifende Dienstleistung: neben der Chemie wurden bereits zur Antragstellung zwei weitere Kooperationspartner aus anderen Disziplinen gewonnen,¹⁶ die ihr Interesse an RADAR und den geplanten Dienstleistungen jeweils durch einen »Letter of Intent« zum Ausdruck brachten. Die erfolgreiche Übertragung der Arbeitsergebnisse aus den ersten zwei Jahren des Projekts auf andere Disziplinen ist ein wichtiges Kriterium für die kommende Zwischenevaluierung und der damit verbundenen eventuellen Weiterförderung von RADAR im dritten Jahr.

ZWEISTUFIGES ANGEBOT

RADAR verfolgt für sein vorgesehenes Dienstleistungsangebot eines disziplinübergreifend nutzbaren Forschungsdatenarchivs einen zweistufigen Ansatz

Schwerpunkt auf der öffentlichen Domäne

selbsttragendes Geschäftsmodell unumgänglich

Übertragbarkeit der Arbeitsergebnisse

mit einem Einstiegsangebot zur Archivierung von Forschungsdaten und einem höherwertigen Angebot mit integrierter Datenpublikation.

Das **Einstiegsangebot** zeichnet sich durch format-unabhängige Archivierung, Bitstream Preservation sowie einen minimalen Metadatensatz mit einem standardisierten Kern aus. Es erlaubt eine Verknüpfung von Daten mit Metadaten, die Verwaltung von Zugriffsrechten und gewährleistet den öffentlichen Zugang zu den Daten. Dieses Angebot richtet sich an Kunden, die primär an der Einhaltung von empfohlenen Haltefristen (5) interessiert sind. Darüber hinaus eignet es sich aber auch für andere Daten wie z. B. Negativdaten¹⁷, die man nicht als Teil einer Publikation verwenden möchte, die aber für andere von hohem Interesse sein können. Das Einstiegsangebot unterscheidet sich von kommerziellen Diensten wie z. B. Dropbox¹⁸ oder figshare¹⁹ durch das hinter ihm stehende Betreiberkonsortium aus wissenschaftsnahen Gedächtnisorganisationen und Infrastruktureinrichtungen, die deutschem Recht unterliegen und keine kommerziellen Interessen verfolgen, sowie in dem gegenüber diesen Diensten gebotenen Mehrwert wie der Überwachung von Haltefristen, der Vergabe von persistenten Identifikatoren (Handles) und der zugesicherten Bitstream Preservation. Darüber hinaus ist bei entsprechender Aufbereitung der Daten ein Übergang aus dieser Stufe in eine höherwertige Angebotsstufe, wie im nächsten Absatz beschrieben, jederzeit möglich.

Die zweite, **höherwertige Angebotsstufe** eignet sich für die dauerhafte Datenarchivierung mit (optionaler) Datenpublikation. Neben den Diensten des Einstiegsangebots ist hier die Vergabe von format- und disziplinspezifischen Metadaten sowie von dauerhaften DOI-Namen in den Publikationsprozess implementiert. Bei der Ausgestaltung der Metadatenprofile kann über eine Kooperation mit dem Projekt Large Scale Data Management & Analysis (LSDMA)²⁰ auf umfangreiche Vorarbeiten zurückgegriffen werden. Der Fokus von RADAR wird vorerst auf der Publikation von Forschungsdaten liegen, die im Zusammenhang mit einer (traditionellen) wissenschaftlichen Publikation stehen (23). Dafür sollen Workflows für den Archivierungs- und Publikationsprozess von Forschungsdaten erarbeitet und die Wissenschaftler bei der Speicherung und Identifizierung ihrer Daten auch über den Zeitraum eines Forschungsprojekts hinaus unterstützt werden.

RADAR will bewusst nicht in Konkurrenz zu existierenden Datenzentren und Fachdatenbanken treten, sondern diese ergänzen und komplettieren. Es deckt auch das Vorhalten nicht publizierter Daten für eine

spezifizierte Haltefrist ab (etwa zehn Jahre nach Projektende gemäß DFG-Empfehlung) und hilft damit das Dilemma aufzulösen, dass Forschungsförderer einerseits ein Vorhalten der Daten fordern, andererseits aber keine verlässliche Infrastruktur dafür bereitstellt. Das geplante Datenzentrum hilft, die bisherige, oft wenig organisierte Speicherung der Daten (im Extremfall auf dem Speicherstick des Wissenschaftlers) zu überwinden.

GESCHÄFTSMODELL

Eines der zentralen Ziele von RADAR ist die Erarbeitung eines sich selbst tragenden, nachhaltigen Geschäftsmodells für den Betrieb eines Datenzentrums. Hierzu gab es bereits zur Antragsstellung erste Überlegungen, die nun im Projektverlauf zu konkretisieren und zu validieren sind. Gegenüber der Antragstellung zeigte sich bereits, dass die Zielgruppen für die angebotenen Dienste erweitert werden sollten. Neben Forschungsprojekten (d. h. Wissenschaftlern) spielen auch institutionelle Nutzer eine wichtige Rolle. Das können z. B. Bibliotheken sein, die das Forschungsdatenmanagement in ihrer Institution selbst übernehmen, die Aufgabe der Bitstream Preservation aber an einen Dienstleister auslagern möchten. Auch Archive und Museen, die zunehmend ihre Bestände digitalisieren, stehen vor der Aufgabe, dafür entsprechende Infrastrukturen aufzubauen und vorzuhalten. Diese Daten sind für die Forschung interessant und benötigen die von RADAR vorgesehenen Basisdienstleistungen.

Insbesondere Wissenschaftler, die Forschungsanträge stellen, können dort zwar Mittel zur Archivierung ihrer Daten vorsehen, sind aber nicht in der Lage, insbesondere nach Ende ihres Projekts dauerhaft für die Archivierung ihrer Daten zu bezahlen. Entsprechend baut das vorgesehene Geschäftsmodell auf Einmalzahlungen für die Archivierung der Forschungsdaten in Abhängigkeit von Datenumfang und Haltedauer auf. Damit können Wissenschaftler die zu erwartenden Kosten schon während der Antragstellung ihrer Forschungsprojekte abschätzen und mit beantragen. Wissenschaftliche Ergebnisse bleiben sowohl im Rahmen der »Regeln guter wissenschaftlicher Praxis« der DFG überprüfbar als auch für nachfolgende Generationen auffindbar und für andere Forschungsfragen verfügbar. Das Datenzentrum übernimmt die dauerhafte und zuverlässige Speicherung der Daten und garantiert dies gegenüber den Wissenschaftlern mittels einer geeigneten Zertifizierung.

Ob sich diese Herangehensweise auch für institutionelle Nutzer eignet, muss im Verlauf des Projekts evaluiert werden. Vorteilhaft sind die sehr einfach zu



Abb. 2: Unterschiedliche Kostenstrukturen bei Bitstream Preservation und funktionaler LZA

kalkulierenden Kosten für die Nutzer, nachteilig kann sich die dauerhafte Festlegung auf einen Dienstleister und die gegenüber einer jährlichen Gebühr deutlich höheren Einmalzahlungen auswirken. Es ist anzunehmen, dass RADAR beide Zahlungsweisen dem Nutzer parallel zur Auswahl anbieten wird.

Das Angebot einer langfristigen (Haltefristen) oder sogar dauerhaften Archivierung auf Basis einer Einmalzahlung stellt ein scheinbar unkalkulierbares Risiko dar, insbesondere im Lichte der bisher fehlenden Kostenmodelle für die Langzeitarchivierung (s. Herausforderung 4). Das trifft allerdings nicht gleichermaßen für alle Aufgaben zu, wie in Abbildung 2 zu sehen ist.

Die notwendigen Aufgaben für eine effektive Bitstream Preservation sind weitgehend verstanden und gut zu kalkulieren. Auf Basis einer retrospektiven Betrachtung der Kosten für Speichersystem²¹ von FIZ Karlsruhe über die letzten zehn Jahre lässt sich feststellen, dass die Hardware-Kosten für notwendige Speicher- und Rechenkapazität pro Speichereinheit (also z. B. pro Terabyte) kontinuierlich sinken, die Administrationskosten für diese Systeme weitgehend stabil blieben und die Kosten für Wartungsverträge und Software-Lizenzen moderat steigen. Die regelmäßige Überprüfung der Datenintegrität und das zugehörige Reporting kann weitgehend automatisiert werden. Es bleibt das relative Risiko steigender Energiekosten, das jedoch pro Speichereinheit über die Jahre hinweg aufgrund von Effizienzgewinnen bei neueren Hardwaregenerationen abnimmt. Der teuerste Schritt ist die Überwindung der von Treloar et al. beschriebenen »Curation Boundary« (7) bei der Aufnahme neuer Daten in das Archiv. Hierbei handelt es sich jedoch um einmalige Kosten. Damit korrelieren die Kosten der Bit-

stream Preservation weitgehend mit der gespeicherten Datenmenge. Eine retrospektive Betrachtung der Kosten lässt durch Extrapolation für die Zukunft kontinuierlich sinkende Kosten pro Terabyte erwarten. Dadurch lassen sich die Kosten für einen Zeitraum von zehn oder zwanzig Jahren relativ zuverlässig abschätzen.²² Bei einer darüber hinausgehenden Speicherdauer werden die Kosten weitgehend marginal sein. Die benötigte Rechenzentrumsleistung für jetzt eingestellte Datenmengen wird in dreißig bis fünfzig Jahren nur noch geringe, im Vergleich zu heute kaum mehr ins Gewicht fallende Kosten verursachen.

Gegensätzlich stellt sich die Situation bei der funktionalen Langzeitarchivierung dar, insbesondere im Forschungsdatenmanagement. Aufgrund der Heterogenität der Daten und oftmals proprietären Dateiformaten sind die Anforderungen an die dauerhafte Gewährleistung einer adäquaten Performanz (24) extrem komplex. Allerdings lässt sich feststellen, dass hier Kosten ganz überwiegend pro Format bzw. Datentyp und weniger pro gespeicherte Datenmenge entstehen. Auch treten sie nur im Fall einer drohenden Obsoleszenz auf. Die dann notwendigen Maßnahmen (sei es im Rahmen von Migration oder Emulation) entstehen einmalig, die erarbeiteten Lösungen können dabei von vielen Archiven mit gleichen Daten nachgenutzt werden. Die entstehenden Kosten lassen sich zwar kaum im Voraus kalkulieren, korrelieren aber nur schwach mit der Speichermenge und sind dadurch einfacher im Rahmen einer punktuellen Förderung, etwa eines Projekts, zu stemmen. Allerdings kollidiert dies mit dem Ziel von RADAR, ein sich selbst tragendes Geschäftsmodell zu entwickeln, weswegen vorerst bewusst auf die funktionale Langzeitarchivierung verzichtet wird.

Kostenkalkulation

Verzicht auf funktionale LZA

DATENPUBLIKATION

Vergabe von DOI-Namen

Die an das RADAR Forschungsdatenarchiv angeschlossene DOI-Registrierung ermöglicht die Vergabe von eindeutigen und persistenten Identifikatoren (DOI-Namen) für Forschungsdaten und damit eine eindeutige Referenzierbarkeit. Die Vergabe von DOI-Namen für Forschungsdaten gilt zudem als Qualitätsmerkmal für die Referenzierung vor allem auch durch die Verknüpfung der Daten mit dem dazu gehörigen wissenschaftlichen Artikel. Durch die zusätzliche Publikation und Referenzierbarkeit der Daten ergibt sich weiterhin die Möglichkeit, die wissenschaftliche Reputation der (Daten-)Autoren in ihrer jeweiligen Fachcommunity zu erhöhen. Die enorme Relevanz der Themen Referenzierung und Reputation spiegeln sich auch bei dem Aufbau des Data Citation Index in Thomson Reuters Web of Science wider, der das Auffinden von Forschungsdaten und ihrer Urheber deutlich verbessert.²³ Die Möglichkeit zur Zitierung von Forschungsdaten erhöht somit nicht nur die Anerkennung von Datenproduzenten, sondern auch die Sichtbarkeit und damit die Verfügbarkeit ihrer Forschungsergebnisse (25).

Erarbeitung von Workflows

In Zusammenarbeit mit dem Thieme Verlag und dem GDCh / Wiley-VCH Verlagsbeirat werden exemplarisch für den naturwissenschaftlichen Sektor mögliche Workflows entwickelt, die den Wissenschaftler bei der Publikation von Forschungsdaten unterstützen. Dabei müssen sowohl Autorenrichtlinien, Zitierstandards und Datenmanagementpläne definiert als auch Modelle für mögliche Kosten erstellt werden. Die erarbeiteten Workflows sollen an zwei Beispielen, der NMR-Spektroskopie und 2D/DIGE-Bildern, getestet, evaluiert und die Ergebnisse anschließend mithilfe eines im Projekt entwickelten Leitfadens auf andere Disziplinen übertragen werden. Dadurch wird RADAR den Anforderungen an eine generische Infrastruktur gerecht und gibt anderen wissenschaftlichen Disziplinen die Möglichkeit, durch entsprechende Aufbereitung ihrer Daten das höherwertige Dienstleistungsangebot zur Publikation ihrer Daten in Anspruch zu nehmen. Dabei ist eine Analyse bestehender und die Identifizierung neuer, auch fachspezifischer Datenformate sowie die Erarbeitung datentechnischer und fachspezifischer Metadatenschemata für den Umgang mit der vorhandenen Heterogenität in diesem Bereich unabdingbar.

offene Systemarchitektur

Perspektivisch sieht RADAR auch eine eigenständige Datenpublikation ohne deren Verknüpfung mit einem wissenschaftlichen Artikel vor. Hierfür können sowohl die publizierten als auch weitere im Rahmen der Forschungsansätze dokumentierte Daten in Betracht gezogen werden, die somit den gesamten Forschungsprozess nachvollziehbar machen. Die Fach-

wissenschaftler müssen dazu in Zusammenarbeit mit Fachgesellschaften beispielhaft (Qualitäts-)Standards etablieren und Kriterien für die Auswahl publikations- und damit auch archivierungswürdiger Daten entwickeln.²⁴ Besonderes Augenmerk muss hierbei auf den gegenwärtigen, aber auch zukünftigen Nutzen der zu archivierenden Daten für die jeweilige Fachcommunity gelegt werden. Auch aus diesem Grund ist es wichtig, die Auswahlkriterien und -prozesse in Kooperation mit übergeordneten Organisationen wie den Fachgesellschaften als jeweilige Interessenvertretung einer Fachwissenschaft zu erarbeiten (20).

Für die Akzeptanz des Datenarchivs in den Fachcommunities ist eine eindeutige und transparente Definition der Verantwortlichkeiten der verschiedenen Akteure unabdingbar, um die Qualität der archivierten Daten und damit auch die des Datenzentrums zu sichern. Dies beinhaltet die kooperative Entwicklung von Policies zur Datenerhaltung zwischen den Akteuren der Datenproduktion und Datenarchivierung, um archivierungswürdige, qualitativ hochwertige Daten als nationales Kulturgut langfristig zu erhalten und nachhaltig auch über institutionelle Grenzen hinweg verfügbar zu machen (20).

ARCHITEKTUR

Die Archivierung ist ein zentraler Baustein im übergreifenden Forschungsdatenmanagement und unabdingbar für die Publikation von Forschungsdaten sowie deren persistente Identifizierung und Referenzierung. Sie stellt gleichzeitig einen generalisierbaren Baustein einer übergreifenden Informationsinfrastruktur dar. Vor diesem Hintergrund bildet ein Archiv perspektivisch den Nukleus für weitere Dienste:

- fachspezifische Datenportale zur Dissemination und Nachnutzung der Daten,
- die Verknüpfung mit institutionellen Systemen des Forschungsdatenmanagements über eine weitergehende Automatisierung der (Meta-)Datenübernahme und
- die möglichst zeitnahe Erfassung der Forschungsdaten während oder nach ihrer Entstehung in einer Dateninfrastruktur und (teil-)automatische Erfassung kontextueller Informationen (als Metadaten).

Die vorgesehene, offene Systemarchitektur des Datenzentrums und seine offenen Programmierschnittstellen (*application programming interfaces*, API) erlauben es den Systembetreibern und Dritten, über die im Antrag beschriebenen Grundfunktionen von RADAR hinaus derartige Dienste aufzubauen. Damit kann RADAR perspektivisch als »Backend« für weitere disziplinspezifische Angebote dienen.

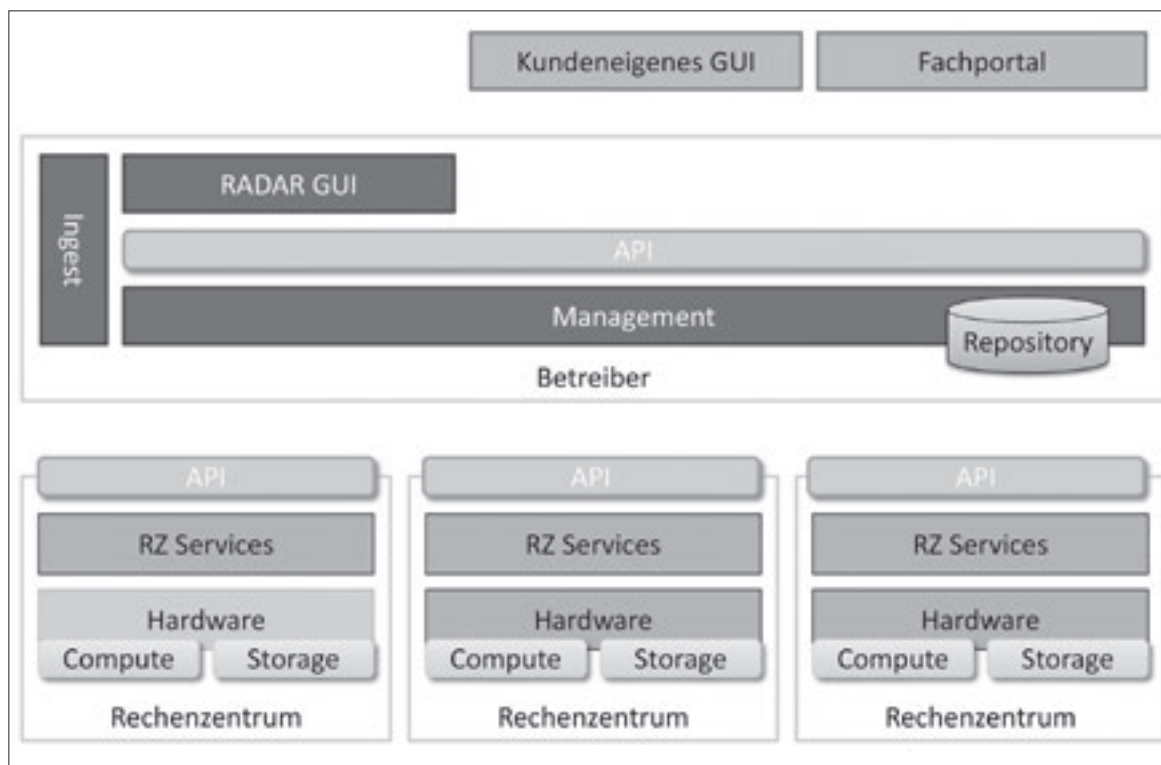


Abb. 3: Schematische Architektur von RADAR mit verteilter Datenhaltung und Schnittstellen

RADAR umfasst bereits in seinem Einstiegsangebot eine Reihe von Dienstleistungen, die folgende Themenbereiche abdecken:

- verlässlicher Speicherplatz für Digitalisate / Forschungsdaten
- Datenaufbereitung (Pre-Ingest)
- Erzeugung technischer Metadaten
- Übernahme beschreibender Metadaten
- Erfassung von Lizenz-Metadaten
- Verwaltung von Zugriffsrechten
- Bereitstellung von Zugriffskopien.

Betrachtet man z. B. den Themenbereich der Bereitstellung eines zuverlässigen Speicherplatzes für Forschungsdaten genauer, sieht man eine Vielzahl an zu berücksichtigenden Aspekten, die sich letztlich auch in der Architektur des Systems niederschlagen müssen:

- mehrfach redundante Speicherung an geografisch verteilten Orten
- langsamer Zugriff für archivierte Daten
- lange Wiederherstellungszeiten im Katastrophenfall
- regelmäßige Überprüfung der Datenkonsistenz inkl. Reporting an datengebende Einrichtung
- Zugriff nur für datengebende Einrichtung bzw. autorisierte Partei und Betreiber möglich
- Vergabe persistenter Identifier (Handles oder DOI-Namen)

- Verwaltung von Speicherplatz-Quotas
- Versionierung.

Die redundante Speicherung an geografisch verteilten Orten führt notwendigerweise zu einer Abstrahierung der von den beteiligten Rechenzentren bereitgestellten Leistungen und zur Spezifikation einer entsprechenden Schnittstelle. Damit löst sich RADAR mittelfristig von der Anbindung an ein konkretes Rechenzentrum und kann perspektivisch neben akademischen auch kommerzielle Rechenzentren bzw. Cloud-Dienste als »Storage-Backend« einbinden. Eine – je nach Anwendungsfall – interessante Alternative kann z. B. Amazon Glacier zur Vorhaltung einer Sicherheitskopie der Daten darstellen. Abbildung 3 zeigt den daraus resultierenden, aktuellen Arbeitsstand der Systemarchitektur für RADAR schematisch auf.

Jedes beteiligte Rechenzentrum als selbstständiger (redundanter) Speicherknoten müsste seine eigene Schicht an Services, die das API bedienen, implementieren. Eine Herausforderung im Projektverlauf wird es sein, das API derart zu gestalten, dass der Aufwand für die Rechenzentren überschaubar bleibt. Eine unterschiedliche technische Auslegung der Rechenzentren ist erwünscht, denn so lassen sich eventuell auftretende systematische Fehler über alle Speicherknoten hinweg vermeiden. Darüber liegt eine Management-Schicht, die die Verteilung der Daten auf die

akademische und kommerzielle Rechenzentren

Speicherknotten und notwendige Verwaltungsaufgaben übernimmt und wiederum über ein API ansprechbar ist. So kann man die Dienste von RADAR nicht nur über die vom Projekt zu entwickelnde Benutzungsoberfläche (GUI), sondern auch über Fachportale oder eigene Benutzungsoberflächen (etwa für institutionelle Kunden) verwenden.

ZUSAMMENFASSUNG UND AUSBLICK

RADAR deckt in seiner ersten Ausbaustufe zunächst nur einen – gleichwohl essentiellen – Teilaspekt des überregionalen Forschungsdatenmanagements ab, bietet jedoch einen evolutionären Ansatz zur Erweiterung. Die Projektpartner haben sich bewusst inhaltlich beschränkt und den pragmatischen Ansatz gewählt, einen verlässlichen Baustein für die breitere (inter-)nationale Informationsinfrastruktur zu entwickeln und als nachhaltigen Dienst zu etablieren. Dazu haben sich Partner mit sehr spezifischen Qualifikationen und Erfahrungen zu einem neuen Informationsverbund zusammengefunden.

RADAR wird sich – in enger Abstimmung mit den Nutzern und verwandten Projekten kontinuierlich weiterentwickeln. Gleichzeitig erlaubt die offene Architektur des Systems, dass Dritte – seien es Wissenschaftler oder Institutionen – in eigenen Projekten ihr fach- und gruppenspezifisches Datenmanagement an die hier geschaffene generische Infrastruktur andocken und so ihre eigenen Systeme auf die von RADAR erbrachte Basisdienstleistung aufsetzen können. Damit kann RADAR seinen Nutzern helfen, Ressourcen zu sparen, über Skaleneffekte die Kosten für die Datenarchivierung zu verringern und die umfangreichen und komplexen Herausforderungen des Forschungsdatenmanagements anzugehen.

LITERATUR

- (1) **Stodden, Victoria; LeVeque, Randall J. und Mitchell, Ian M.** Reproducible research for scientific computing: Tools and strategies for changing the culture. In: Computing in Science and Engineering. July/August 2012, Bd. 14, 4, S. 13–17.
- (2) **Hey, Tony; Tansley, Stewart und Tolle, Kristin [Hrsg.]**. The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, Washington: Microsoft Research, 2009. <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>
- (3) **Pilat, Dirk und Fukasaku, Yukiko**. OECD Principles and Guidelines for Access to Research Data from Public Funding. In: Data Science Journal. 29. June 2007, Bd. 6, S. OD4-OD11.
- (4) **Giarretta, David, et al.** Riding the wave – How Europe can gain from the rising tide of scientific data.

Final report of the high level expert group on scientific data. [Online] October 2010. <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>

- (5) **Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten**. Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme, Deutsche Forschungsgemeinschaft. 2009. www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf
- (6) **EU GRDI2020 Roadmap Report. Global Scientific Data Infrastructures: The Big Data Challenges**. 2011.
- (7) **Treloar, Andrew und Harboe-Ree, Cathrine**. Data management and the curation continuum. How the Monash experience is informing repository relationships. Melbourne, Australia: 2008. 14th Victorian Association for Library Automation, Conference and Exhibition. <http://arrow.monash.edu.au/hdl/1959.1/43940>
- (8) **Klump, Jens**. Managing the Data Continuum. [Online] 30.11.2009. http://oa.helmholtz.de/fileadmin/user_upload/Data_Continuum/klump.pdf
- (9) **Miksa, Tomasz und Rauber, Andreas**. Increasing preservability of research by process management plans. New York, NY, USA: ACM, 2013. DPRMA '13 Proceedings of the 1st International Workshop on Digital Preservation of Research Methods and Artefacts, S. 20. <http://doi.acm.org/10.1145/2499583.2499591>
- (10) **Borgman, Christine L**. The Conundrum of Sharing Research Data. In: Journal of the American Society for Information Science and Technology. April 2012, Bd. 63, 6, S. 1059–1078.
- (11) **Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastruktur in Deutschland bis 2020**. Berlin: Wissenschaftsrat, 2012. www.wissenschaftsrat.de/download/archiv/2359-12.pdf
- (12) **Bicarregui, J., et al.** DMP Planning for Big Science Projects. 2012. <http://arxiv.org/abs/1208.3754v1>
- (13) **Neuroth, H., et al. [Hrsg.]**. Langzeitarchivierung von Forschungsdaten – Eine Bestandsaufnahme. Göttingen: 2012. <http://nbn-resolving.de/urn:nbn:de:0008-2012031401>
- (14) **Becker, Christoph**. Vertrauenswürdige Planung in der digitalen Langzeitarchivierung. In: Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare. 2011, Bd. 64, 2, S. 233–246.
- (15) **Klump, Jens**. Criteria for the Trustworthiness of Data Centres. In: D-Lib Magazine. 2011, Bd. 17, 1–2.
- (16) **Ambacher, Bruce und Dale, Robin L**. Trustworthy Repository Audit & Certification: Criteria and Checklist. 2011. www.crl.edu/PDF/trac.pdf
- (17) **Schoger, Astrid und Dobratz, Susanne [Hrsg.]**. Catalogue of Criteria for Trusted Digital Repositories. Frankfurt/Main: nestor, 2009. <http://nbn-resolving.de/urn:nbn:de:0008-2010030806>

bewusste Wahl eines pragmatischen Ansatzes

kontinuierliche und kooperative Weiterentwicklung

(18) **DINI Arbeitsgruppe »Elektronisches Publizieren«.** In: DINI-Zertifikat Dokumenten- und Publikations-service 2010. 2011. <http://nbn-resolving.de/urn:nbn:de:kobv:11-100182794>

(19) **Data Seal of Approval. Guidelines Version 2.** Den Haag: Data Archiving and Networked Services (DANS), 2013. http://datasealofapproval.org/media/filer_public/2013/09/27/guidelines_2014-2015.pdf

(20) **Sustainable Economics for a Digital Planet – Ensuring Long-Term Access to Digital Information. Final Report.** Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2010. http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

(21) **Beagrie, Neil; Chruszcz, Julia und Lavoie, Brian.** Keeping Research Data Safe – A Cost Model and Guidance for UK Universities. Final Report. s. l.: JISC, 2008. www.jisc.ac.uk/media/documents/publications/keeping_researchdatasafe0408.pdf

(22) **McGovern, N. Y. [Hrsg.].** Aligning National Approaches to Digital Preservation. Atlanta, GA, USA: Educopia Institute Publications, 2012. http://educopia.org/sites/educopia.org/files/ANADP_Educopia_2012.pdf

(23) **FIZ Chemie; TIB Hannover; Universität Paderborn.** Konzeptstudie Vernetzte Primärdaten-Infrastruktur für den Wissenschaftler-Arbeitsplatz in der Chemie. [Online] September 2010. www.tib-hannover.de/fileadmin/projekte/primaer-chemie/Konzeptstudie_Forschungsdaten_Chemie.pdf

(24) **Leitfaden zur Digitalen Bestandserhaltung. Vorgehensmodell und Umsetzung. Version 2.0.** Frankfurt/Main: nestor-Arbeitsgruppe »Digitale Bestandserhaltung«, 2012. <http://nbn-resolving.de/urn:nbn:de:0008-2012092400>

(25) **Kotarski, R., et al.** Report on Best Practices for Citability of Data and Evolving Roles in Scholarly Communication. 2012. www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/08/ODE-ReportBestPracticesCitabilityDataEvolvingRolesScholarlyCommunication.pdf

¹ Die Verwendung der männlichen oder weiblichen Sprachform schließt die jeweilig andere nicht aus und soll lediglich der besseren Lesbarkeit dienen.

² Charta zur Bewahrung des digitalen Kulturerbes: www.unesco.de/444.html

³ www.icsu-wds.org/

⁴ GESIS – Leibniz-Institut für Sozialwissenschaften, www.gesis.org/

⁵ Die PDB wird vom Research Collaboratory for Structural Bio-informatics (RCSB) unterhalten, www.rcsb.org

⁶ Die Library of Congress archiviert die Daten des Microblogdienstes Twitter u. a. zu Forschungszwecken (<http://blogs.loc.gov/loc/2010/04/the-library-and-twitter-an-faq/>)

⁷ Seeing Standards: A Visualization of the Metadata Universe, www.dlib.indiana.edu/~jenlrile/metadatamap/

⁸ Z.B. SCAPE – Scalable Preservation Environments (www.scape-project.eu/) oder bwFLA (<http://bw-fla.uni-freiburg.de/>).

⁹ Z.B. NESTOR, www.langzeitarchivierung.de/

¹⁰ Unter dieser Abkürzung werden Verlage aus dem Bereich Science, Technology und Medicine zusammengefasst.

¹¹ <http://datacite.org/node/65> und www.stm-assoc.org/industry-news/crossref-joins-stm-datacite-statement/

¹² http://icsu-wds.org/images/files/WDS_Certification_Summary_11_June_2012.pdf

¹³ DRAMBORA interactive: Audit Method Based on Risk Assessment (DRAMBORA), www.repositoryaudit.eu/

¹⁴ Siehe z.B. <http://wiki.opf-labs.org/display/CDP/Home>

¹⁵ 2D-/DIGE steht für »Differential In-Gel Electrophoresis«, einem bildbasierten Verfahren aus den Lebenswissenschaften zur Proteinanalyse.

¹⁶ Georg-Eckert-Institut für internationale Schulbuchforschung, www.gei.de/nc/georg-eckert-institut.html und Hornemann Institut, www.hornemann-institut.de/

¹⁷ Daten, die zu keinem Ergebnis geführt haben.

¹⁸ <https://www.dropbox.com/>

¹⁹ <http://figshare.com/>

²⁰ www.helmholtz-lsdma.de/

²¹ Dies betrifft sowohl Disk- als auch Tape-basierte Speichersysteme. Eine Differenzierung nach Serviceklassen (etwa Verlässlichkeit, Geschwindigkeit o. ä.) fand dabei nicht statt.

²² Vgl. auch (21), S. 4–6.

²³ Vgl. auch http://thomsonreuters.com/content/press_room/science/686112

²⁴ Vgl. hierzu auch (11), S. 14.

DIE VERFASSER

Matthias Razum, Leiter e-Science, IT, Entwicklung und angewandte Forschung FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, E-Mail: Matthias.Razum@fiz-karlsruhe.de, www.fiz-karlsruhe.de

Dr. Janna Neumann, Leitung DOI-Service, Abt. Forschung und Entwicklung, Technische Informationsbibliothek (TIB), Welfengarten 1B, 30167 Hannover, E-Mail: Janna.Neumann@tib.uni-hannover.de, www.tib-hannover.de

Dr. Matthias Hahn, Projektleiter, IT, Entwicklung und angewandte Forschung, FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, E-Mail: Matthias.Hahn@fiz-karlsruhe.de, www.fiz-karlsruhe.de