

Frank Scholze, Roland Bertelmann, Maxi Kindling,
Heinz Pampel und Paul Vierkant

Open Access und Forschungsdaten

Die Digitalisierung hat der Wissenschaft neue Möglichkeiten des Umgangs mit Information und Wissen eröffnet. Das Potenzial der vernetzten Forschung wurde bereits in der „Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities“ festgehalten.¹ Nicht nur qualitätsgesicherte Aufsätze, sondern auch „raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material“ sollen über das Internet offen zugänglich und nachnutzbar gemacht werden – ein Anliegen, das heute vermehrt unter dem Begriff „Open Science“ diskutiert wird.²

Die beiden zentralen Argumente für den offenen Zugang zu digitalen Forschungsdaten sind zum einen die Möglichkeit der Nachnutzung der Daten in neuen Zusammenhängen und zum anderen die Nachvollziehbarkeit bzw. Reproduzierbarkeit der Daten zur Sicherung der guten wissenschaftlichen Praxis.

Die Forderung nach offenen Forschungsdaten wirkt auf den einzelnen Forschenden und dessen Umgang mit den digitalen Daten, die je nach Disziplin in vielfältigen Formen und Formaten vorliegen. The Royal Society empfiehlt in ihrer lesenswerten Denkschrift *Science as an open enterprise*: „Where data justify it, scientists should make them available in an appropriate data repository.“³

Forschende erkennen zwar mehrheitlich die Potenziale des offenen Zugangs zu wissenschaftlichen Daten, stehen dem „data sharing“ der eigenen Daten dennoch häufig zurückhaltend gegenüber. Auf Basis der Erhebungen von Kuipers 2009 und Tenopir 2011 lassen sich rechtliche, technische und kulturelle Barrieren für einen freien Austausch von Forschungsdaten benennen.⁴ Die genann-

1 Berlin Declaration. 2003.

2 Grundlegend hierzu Nielsen 2012. Vgl. auch Bartling 2014.

3 The Royal Society. 2012.

4 Kuipers 2009 sowie Tenopir 2011 nennen „legal issues“, „misuse of data“, „incompatible data types“ sowie „insufficient time“ und „lack of funding“.

Bei diesem Beitrag handelt es sich um eine aktualisierte und gekürzte Fassung von: Pampel, H., Bertelmann, R., Scholze, F., Kindling, M., Vierkant, P. (2015): Stand und Perspektive des globalen Verzeichnisses von Forschungsdaten-Repositoryen re3data.org. In: Müller, P., Neumair, B., Reiser, H., Rodosek, G. D. (Eds.), 8. DFN-Forum Kommunikationstechnologien: Beiträge der Fachtagung 08.06-09.06.2015 in Lübeck, (GI-Edition: lecture notes in informatics; 243), Bonn: Gesellschaft für Informatik, 13–22.

ten Barrieren machen deutlich, dass die Veröffentlichung von Forschungsdaten eines definierten Rahmens bedarf. Die Konzeption und Umsetzung dieses Rahmens wird unter dem Begriff „Research Data Management“ diskutiert. Ziel des Forschungsdatenmanagements (FDM) ist es, durch organisatorische, technische, rechtliche und finanzielle Maßnahmen die dauerhafte Nachnutzung der Daten auf Basis von vertrauenswürdigen Infrastrukturen zu gewährleisten. Im Fokus steht dabei der Betrieb von digitalen Informationsinfrastrukturen wie Forschungsdaten-Repositoryn (FDR), mit Hilfe derer die dauerhafte Zugänglichmachung der Daten möglich ist.

Die Diskussion über den Umgang mit Forschungsdaten und die damit verbundenen Implikationen für digitale Informationsinfrastrukturen sind längst im wissenschaftspolitischen Raum angekommen: Auf europäischer Ebene sind die Mitgliedsstaaten und deren wissenschaftliche Einrichtungen seit 2012 durch die Europäische Kommission dazu aufgefordert, die dauerhafte Zugänglichkeit von digitalen Forschungsdaten sicherzustellen.⁵ Insbesondere durch die Verankerung des Themas im aktuellen EU-Forschungsrahmenprogramm Horizon 2020 wird die Diskussion gefördert. In mehreren Förderlinien müssen Mittelempfänger im Rahmen des „Open Research Data Pilot“ den möglichst offenen Zugang zu den entstehenden Daten sicherstellen.⁶

Die deutsche Bundesregierung „misst dem Thema des offenen Zugangs zu wissenschaftlichen Informationen eine hohe Bedeutung bei“ und hat angekündigt, das Thema im Rahmen der bereits im Koalitionsvertrag angekündigten „Open-Access-Strategie“ aufzugreifen und im Rahmen der „Digitalen Agenda“ umzusetzen.⁷ Ähnliche Entwicklungen finden auch in anderen europäischen Ländern statt.⁸

⁵ European Commission. 2012.

⁶ European Commission. 2013.

⁷ Bundesregierung. 2014. Darüber hinaus haben einzelne Bundesländer begonnen, sich mit dem Themenfeld zu befassen. Das Land Berlin hat 2014 die Entwicklung einer „Open-Access-Strategie für Berlin“ angekündigt, welche auch die Zugänglichmachung und Nachnutzung von Forschungsdaten beinhalten soll. Das Land Baden-Württemberg hat 2014 ein „Fachkonzept zur Weiterentwicklung der wissenschaftlichen Infrastruktur“ vorgelegt und eine „Forschungsdaten-Strategie“ angekündigt. Seit November 2014 arbeitet der von Bund und Länder eingesetzte Rat für Informationsinfrastrukturen an „disziplin- und institutionsübergreifender Empfehlungen für die weitere Entwicklung und den Ausbau der digitalen Infrastrukturen“. Wissenschaftliche Einrichtungen haben auf diese Entwicklung reagiert und institutionelle Richtlinien erlassen, so z. B. die Universität Bielefeld und die Humboldt-Universität zu Berlin. Vgl. auch die Dokumentation im Wiki Forschungsdaten.org: http://www.forschungsdaten.org/index.php/Data_Policies#Institutionelle_Policies.

⁸ Nicol 2013.

In den USA wurden die öffentlich finanzierten Förderorganisationen 2013 per Regierungsdirektive dazu aufgefordert, die Zugänglichkeit von Forschungsdaten zu verbessern, die im Rahmen von öffentlich geförderten Projekten entstehen. Mittelempfänger müssen in sogenannten „Data Management Plans“ ihre Maßnahmen zur Sicherung der dauerhaften Zugänglichkeit der digitalen Daten beschreiben.⁹

In Deutschland erwartet die Deutsche Forschungsgemeinschaft (DFG) seit 2010 von Antragstellenden Aussagen hinsichtlich des Umgangs mit Forschungsdaten, die im Rahmen eines beantragten Projekts entstehen.¹⁰ Dieses Anliegen wird vermehrt auch über „Data Policies“ von wissenschaftlichen Zeitschriften verfolgt. In der entsprechenden Richtlinie der Nature Publishing Group (NPG) heißt es: „authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications“.¹¹ Angeregt wird die Zugänglichmachung der Daten über öffentliche Repositorien. Einen Schritt weiter geht die Public Library Of Science (PLOS). Publizierende bei PLOS müssen seit 2014 bei der Einreichung ein „Data Availability Statement“ abgeben. In diesem muss angegeben werden, an welchem „Ort“ die Daten, die Grundlage der eingereichten Publikation sind, zugänglich gemacht sind oder zum Zeitpunkt der Veröffentlichung zugänglich sein werden.¹²

1 Forschungsdaten-Repositorien

Um Forschende beim „data sharing“ zu unterstützen, bedarf es digitaler Informationsinfrastrukturen, mittels derer Forschungsdaten zugänglich und nachnutzbar gemacht werden können. Diese Infrastrukturen werden vermehrt unter dem Begriff Forschungsdaten-Repositorien (FDR) zusammengefasst. Anliegen dieser Infrastrukturen, die sich in ihren Funktionalitäten teils stark unterscheiden, ist es, Forschungsdaten möglichst dauerhaft zugänglich zu machen. Die Landschaft

⁹ Office of Science and Technology Policy. 2013. Diese Praxis wird bereits seit 2003 am National Institute of Health (NIH) umgesetzt. Antragsteller, die eine Zuwendung ab 500.000 US-Dollar beantragen, sind dort aufgefordert, Aussagen zum „data sharing“ zu machen. Seit 2011 verfügt auch die National Science Foundation (NSF) über eine entsprechende Regelung. Auch dort müssen Antragsteller Maßnahmen zur Umsetzung der Regelung in einem „Data Management Plan“ spezifizieren.

¹⁰ Deutsche Forschungsgemeinschaft. 2014.

¹¹ Nature Publishing Group. 2013.

¹² Public Library of Science. 2013.

dieser FDR ist noch heterogen und unübersichtlich. Die Europäische Kommission stellte 2009 in ihrem Positionspapier *ICT infrastructures for e-science fest*: „The landscape of data repositories across Europe is fairly heterogeneous, but there is a solid basis to develop a coherent strategy to overcome the fragmentation and enable research communities to better manage, use, share and preserve data.“¹³

Bisher haben sich nur wenige Untersuchungen einer Zustandsbeschreibung der FDR weltweit und fachübergreifend angenommen.¹⁴ Erhellend ist allerdings der Blick auf disziplinäre Untersuchungen. Eine eindrucksvolle FDR-Landschaft bietet beispielsweise die Biomedizin. Die Ausgabe 2014 der Molecular Biology Database Collection des biomedizinischen Journals *Nucleic Acids Research* weist alleine über 1.550 Infrastrukturen nach, auf denen biomedizinische Forschungsdaten gespeichert werden können.¹⁵

2 re3data.org – Registry of Research Data Repositories

Das Ziel von re3data.org – Registry of Research Data ist es, Forschenden eine Orientierung über bestehende FDR zu bieten, um das „data sharing“ in der Wissenschaft zu fördern.¹⁶ Der Dienst kann darüber hinaus von Förderorganisationen, Informationsinfrastruktur-Einrichtungen und Verlagen bei der Identifikation geeigneter Infrastrukturen zur Speicherung und zur Suche von Forschungsdaten genutzt werden. Das web-basierte Verzeichnis weist im August 2015 über 1.300 FDR nach.

Der Aufbau des Dienstes wird von 2012 bis 2015 von der Deutschen Forschungsgemeinschaft (DFG) gefördert. Projektpartner sind die Abteilung Bibliothek und Informationsdienste (LIS) des Deutschen GeoForschungsZentrums GFZ, das Institut für Bibliotheks- und Informationswissenschaft (IBI) der Humboldt-Universität zu Berlin sowie die Bibliothek des Karlsruher Instituts für Technologie (KIT). Vernetzungspartner ist die Deutsche Initiative für Netzwerkinformation (DINI), in deren Kontext die drei Projektpartner seit längerem beim Betrieb von Repositorien zusammenarbeiten.

¹³ European Commission. 2009.

¹⁴ Marcial 2010 sowie Pampel 2013.

¹⁵ Fernández-Suárez 2014.

¹⁶ <http://www.re3data.org/>

Seit 2014 kooperiert re3data.org mit den Purdue University Libraries in West Lafayette, Indiana, USA und dem dort entwickelten Dienst Databib.¹⁷ Zunächst wurden die Daten ausgetauscht und eine Roadmap für die weitere Integration bis Ende 2015 entwickelt. Im Mai 2015 beschloss die Mitgliederversammlung von DataCite, dass der fusionierte Dienst unter dem Namen re3data.org langfristig im Rahmen dieser internationalen Organisation betrieben werden soll.¹⁸ Hierzu wird eine re3data.org-Arbeitsgruppe bei DataCite etabliert, die die bisherigen Projektpartner und Vertreter von DataCite umfasst. Diese wird die künftige Entwicklung des Dienstes vorantreiben. Der Fokus der Weiterentwicklung liegt neben der Optimierung der Arbeitsabläufe auf der Verbesserung der Nachnutzung der Metadaten mithilfe einer Programmierschnittstelle (API).¹⁹ Damit soll die Einbindung von re3data.org in diverse Dienste des Forschungsdatenmanagements erleichtert werden.²⁰

Zur Beschreibung der FDR wurde im intensiven Dialog mit verschiedenen Fachdisziplinen ein Schema entwickelt. Seit April 2015 liegt das Schema in der Version 2.2 vor.²¹ Mit der Entwicklung des Schemas wurde der Grundstein für einen Standard zur Beschreibung von FDR gelegt. Sechs intuitiv gestaltete Icons visualisieren das Metadaten-Schema und helfen insbesondere Forschenden bei der Identifikation eines geeigneten FDR zur Speicherung ihrer Daten. So ist auf einen Blick sichtbar, unter welchen Zugangsbedingungen Forschungsdaten zugänglich und nachnutzbar gemacht werden (vgl. Tab. 1) oder ob ein FDR persistente Identifikatoren vergibt. FDR-Betreibern bietet das Icon-System eine Unterstützung bei der Weiterentwicklung ihrer Infrastrukturen.

17 <https://www.datacite.org/news/merger-databib-and-re3dataorg-first-version-api-available.html>

18 <https://www.datacite.org/news/datacite-manage-and-develop-re3dataorg.html>

19 Eine erste Version der Programmierschnittstelle (API) ging im März 2015 online. Eine Dokumentation findet sich unter: <http://www.re3data.org/api/doc>

20 Beispielsweise in Werkzeuge zur Erstellung von „Data Management Plans“, wie sie die Europäische Kommission im Rahmen ihres „Open Research Data Pilots“ verlangt.

21 Vierkant 2014. Das Schema umfasst 39 Eigenschaften (Properties).

Tab. 1: re3data.org Icons zur Visualisierung des Zugriffs auf FDR.

Icon			
<i>Access (property)</i>	<i>Open Access</i>	<i>Restricted Access</i>	<i>Closed Access</i>
Access to Repository	open	open or restricted	closed
Access to Data	open (embargoed, restricted, closed)	restricted (embargoed, closed)	closed
Data Upload	open or restricted	closed	–

Das Verzeichnis wurde bereits im November 2014 von über 5.000 einzelnen Nutzenden (Unique Visitors, ohne Roboter) pro Monat besucht. Die erstellten Metadaten sind unter der Creative-Commons-Deed Zero (CC0) offen nachnutzbar und werden z. B. von dem EU-Projekt OpenAIRE nachgenutzt. Mehrere Förderorganisationen, so z. B. die Europäische Kommission, empfehlen in ihren Leit- und Richtlinien zum Umgang mit Forschungsdaten die Nutzung von re3data.org. Auch verweisen viele wissenschaftliche Zeitschriften in ihren Autorenrichtlinien auf den Dienst. Beispiele hierfür sind Zeitschriften der Verlage Copernicus Publications, NPG, PeerJ und Springer.

3 Ausblick

Wie oben beschrieben sind in den letzten Jahren bereits zahlreiche Initiativen entstanden, um das kollaborative Arbeiten mit Forschungsdaten zu unterstützen. In vielen Disziplinen wird dem Zögern der Forschenden durch neue Strategien begegnet, um Anreize zu schaffen. Viele dieser Strategien erfordern ein kontinuierliches Zusammenspiel aller Akteure. Sie werden nur dann zum Erfolg führen, wenn auch Fachgesellschaften und wichtige Vertreter der Disziplinen sich beteiligen. Eine möglichst alle Disziplinen umfassende Professionalisierung des Forschungsdatenmanagements, das Forschende in der Weitergabe ihrer Daten unterstützt, ist notwendig. In diesem Zusammenhang besitzt die Strukturierung und Vernetzung von FDR und deren langfristige Finanzierung Vorrang. Mit dem Nachweis und der damit verbundenen Analyse von FDR leistet re3data.org einen zentralen Beitrag, diesbezüglich Transparenz herzustellen und einen Wissenstransfer aus Disziplinen zu ermöglichen, die bereits über gut funktionierende Infrastrukturen und entsprechende Mechanismen der dauerhaften

Zugänglichkeit wissenschaftlicher Forschungsdaten verfügen. Diese gilt es, auf Fächer zu übertragen, die bislang wenige oder nicht nachhaltige Strukturen des Forschungsdatenmanagements aufgebaut haben.

Wichtig ist dabei, Open Access im Sinne des freien Zugriffs auf Publikationen nicht deckungsgleich auf Forschungsdaten anzuwenden, sondern, im Sinne der eingangs zitierten Open Science-Konzeption der Royal Society, für die einzelnen Disziplinen passende Lösungen „intelligenter Offenheit“ zu finden, d. h. den Zugang zu und die Nachnutzung von Forschungsdaten so frei wie möglich und so eingeschränkt wie nötig zu gestalten. Dies ist nichts Geringeres als ein Kulturwandel der Wissenschaft – auf den Ebenen von Offenheit und Kollaboration. Denn auch und gerade für die Fächer, die noch nicht über etablierte Strukturen des Forschungsdatenmanagements verfügen, stellt sich die Frage, wie viele Parallelstrukturen und Dienste aufgebaut werden sollen. Benötigt jede größere Universität etwa ein „Digital Humanities Center“? Ja in dem Sinne, dass es eine lokale Unterstützungs- und Beratungsstruktur des Kulturwandels geben sollte. Nein, in dem Sinne, dass nicht überall technische Speicherstrukturen und FDR aufgebaut werden müssen. Dies stellt einen weiteren Aspekt des Kulturwandels in Punkto Kollaboration dar, bezogen auf die wissenschaftsnahen Infrastruktureinrichtungen. Die genannten Einsichten können nicht verordnet werden, dazu sind zu viele Ressourcen dezentral im Wissenschaftssystem vorhanden. Vielmehr nötig ist die Einsicht in die Komplexität der Bezüge und in das schiere Unvermögen alles selbst leisten zu können, die Hochschulen und ihre Infrastruktureinrichtungen von der Basis her zum Wandel befähigt. Dies muss gegenläufig zur großflächigen strategischen Planung verlaufen, um das Gesamtsystem zum Erfolg zu führen und im Interesse der Wissenschaft das kollaborative Arbeiten mit Forschungsdaten zu unterstützen. Strategisch wird die Hochschulrektorenkonferenz das Thema Forschungsdaten einem Positionspapier adressieren, das für November 2015 angekündigt ist.²² Auch der Rat für Informationsinfrastrukturen²³ wird sich des Themas annehmen. Bleibt zu hoffen, dass die Gegenläufigkeit auch „bottom up“ realisiert wird.

²² Eine erste Empfehlung zum Umgang mit Forschungsdaten erfolgte bereits im Mai 2014 vgl. <http://www.hrk.de/positionen/gesamtliste-beschluesse/position/convention/management-von-forschungsdaten-eine-zentrale-strategische-herausforderung-fuer-hochschulleitungen/>

²³ <http://www.gwk-bonn.de/themen/uebergreifende-wissenschafts-und-forschungspolitische-themen/informationsinfrastruktur/>

Bibliographie

- Bartling, Sönke und Sascha Friesike, Hg. 2014. *Opening Science: The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*. Heidelberg: Springer. <http://doi.org/10.1007/978-3-319-00026-8>
- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. 2003. Berlin. <http://openaccess.mpg.de/Berliner-Erklarung>
- Bundesregierung. 2014. Strategie der Bundesregierung zum Europäischen Forschungsraum (EFR): Leitlinien und nationale Roadmap. Berlin. http://www.bmbf.de/pubRD/EFR-Strategie_deutsch.pdf
- Deutsche Forschungsgemeinschaft (DFG). 2014. Leitfaden für die Antragstellung – Projektanträge. Bonn. http://www.dfg.de/formulare/54_01/
- European Commission. 2009. ICT infrastructures for e-science. Luxemburg. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0108:FIN:EN:PDF>
- European Commission. 2012. Commission Recommendation on access to and preservation of scientific information. Luxemburg. http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf
- European Commission. 2013. Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Luxemburg. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- Fernández-Suárez, Xose M., Daniel J. Rigden und Michael Y. Galperin. 2014. The 2014 Nucleic Acids Research Database Issue and an updated NAR online Molecular Biology Database Collection. *Nucleic Acids Research*, 42 (Database issue), D1–6. <http://doi.org/10.1093/nar/gkt1282>
- Kuipers, Tom und Jeffrey van der Hoeven. 2009. Insight into digital preservation of research output in Europe. *Survey Report*. http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf
- Marcial, Laura Haak und Bradley M. Hemminger. 2010. Scientific data repositories on the Web: An initial survey. *Journal of the American Society for Information Science and Technology*, 61(10), 2029–2048. <http://doi.org/10.1002/asi.21339>
- Nature Publishing Group. 2013. Guide to Publication Policies of the Nature Journals. <http://www.nature.com/authors/gta.pdf>
- Nicol, Aurore, Julie Caruso und Eric Archambault. 2013. Open data access policies and strategies in the European Research Area and beyond. Montréal. http://www.science-metrix.com/pdf/SM_EC_OA_Data.pdf
- Nielsen, Michael. 2012. *Reinventing Discovery*. Princeton: Princeton University Press.
- Office of Science and Technology Policy. 2013. Increasing Access to the Results of Federally Funded Scientific Research. http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
- Pampel, Heinz et al. 2013. Making Research Data Repositories Visible: The re3data.org Registry. *PLOS ONE* 8 (11): e78080, 2013. <http://doi.org/10.1371/journal.pone.0078080>
- Public Library of Science. 2013. Data Access for the Open Access Literature: *PLOS's Data Policy*. <http://www.plos.org/data-access-for-the-open-access-literature-ploss-data-policy/>
- The Royal Society. 2012. Science as an open enterprise. London. <https://royalsociety.org/~media/policy/projects/sape/2012-06-20-saoc.pdf>

Tenopir, Carol et al. 2011. Data sharing by scientists: practices and perceptions. *PLOS ONE* 6 (6): e21101. <http://doi.org/10.1371/journal.pone.0021101>

Vierkant, Paul et al. 2014. Schema for the description of research data repositories. Version 2.2. <http://doi.org/10.2312/re3.006>.