



**Thomas Elsässer, Martin Grötschel, Matthias Scheffler,
Joachim Hermann Ullrich, Friedhelm von Blanckenburg**

Open Research Data in Naturwissenschaften und Mathematik

Empfehlungen der Mathematisch-Naturwissenschaftlichen Klasse der BBAW

Berlin: Berlin-Brandenburgische Akademie der Wissenschaften, 2022
ISBN: 978-3-949455-12-4

(Denkanstöße aus der Akademie : eine Schriftenreihe der Berlin-Brandenburgischen Akademie der
Wissenschaften ; 10 (Juli/2022)

Persistent Identifier: urn:nbn:de:kobv:b4-opus4-37378

Die vorliegende Datei wird Ihnen von der Berlin-Brandenburgischen Akademie der Wissenschaften unter einer
Creative Commons Namensnennung 4.0 International Lizenz zur Verfügung gestellt.



Denkanstöße

aus der Akademie

10
Juli/2022

Eine Schriftenreihe der Berlin-Brandenburgischen
Akademie der Wissenschaften

Thomas Elsässer
Martin Grötschel
Matthias Scheffler
Joachim Hermann Ullrich
Friedhelm von Blanckenburg

OPEN RESEARCH DATA IN NATUR- WISSENSCHAFTEN UND MATHEMATIK

EMPFEHLUNGEN DER MATHEMATISCH-
NATURWISSENSCHAFTLICHEN KLASSE DER BBAW

Informationen zur Publikationsreihe

In der Reihe „Denkanstöße aus der Akademie“ werden Beiträge **von Mitgliedern der Berlin-Brandenburgischen Akademie der Wissenschaften** zu aktuellen forschungspolitischen und wissenschaftlichen Themen veröffentlicht. Die namentlich gekennzeichneten Beiträge geben die Auffassung der Verfasserinnen und Verfasser wieder. Sie repräsentieren nicht notwendigerweise den Standpunkt der Akademie als Institution.

Informationen zu den Autoren

Thomas Elsässer: Direktor am Max-Born-Institut, Professor für Experimentalphysik an der Humboldt-Universität zu Berlin/ **Martin Grötschel:** Präsident der BBAW (2015–2020), Professor für Informationstechnik an der Technischen Universität Berlin (pensioniert), ehem. Präsident des Konrad-Zuse-Zentrums für Informationstechnik Berlin / **Matthias Scheffler:** Direktor des NOMAD Laboratory am Fritz-Haber-Institut der Max-Planck-Gesellschaft und der Humboldt-Universität zu Berlin / **Joachim Hermann Ullrich:** Präsident der Deutschen Physikalischen Gesellschaft / **Friedhelm von Blanckenburg:** Leiter der Sektion „Geochemie der Erdoberfläche“ am Deutschen GeoForschungszentrum (GFZ), Professor der Geochemie der Erdoberfläche, Freie Universität Berlin

Alle Autoren sind Mitglieder der Mathematisch-Naturwissenschaftlichen Klasse der BBAW.

Herausgeber: Der Präsident der Berlin-Brandenburgischen Akademie der Wissenschaften

Redaktion: Ute Tintemann

Grafik: Satz: eckedesign GmbH Berlin; Entwurf: angenehme Gestaltung/Thorsten Probst

Druck: PIEREG Druckcenter Berlin GmbH

@Berlin-Brandenburgische Akademie der Wissenschaften, 2022

Jägerstr. 22-23, 10117 Berlin, www.bbaw.de

Lizenz: CC-BY

ISBN: 978-3-949455-12-4

INHALTSVERZEICHNIS

Einleitung	3
1 Ausgangslage	4
2 Open Research Data und Fächerkultur	6
3 Open Research Data und öffentliche Forschungsförderung	7
4 Umsetzung von Open-Research-Data-Konzepten	8
5 Fazit und Empfehlungen	9

Open Research Data in Naturwissenschaften und Mathematik: Empfehlungen der Mathematisch-Naturwissenschaftlichen Klasse der BBAW

EINLEITUNG

Forschungsdaten haben einen immensen Wert für die Wissenschaft, wie auch für Wirtschaft und Gesellschaft im Allgemeinen. Dieser Wert resultiert nicht nur aus den Kosten ihrer Erzeugung, sondern beruht insbesondere auf ihrem enormen Wertschöpfungspotential. Gegenwärtig wird der „Rohstoff Forschungsdaten“ nur unzureichend genutzt. Ein Umdenken und die Entwicklung breiter Nutzungsmöglichkeiten sind deshalb dringend erforderlich und Gegenstand aktueller Programme wie der Initiative zur Nationalen Forschungsdateninfrastruktur (NFDI). Dieser Prozess muss primär von aktiven Wissenschaftlerinnen und Wissenschaftlern aus komplementären Fachdisziplinen gestaltet werden und zu Lösungen führen, die wissenschaftsgerecht, effizient und unbürokratisch sind. Der interdisziplinäre wissenschaftliche Austausch in der BBAW besitzt ein hohes Potential für eine breite Diskussion der Thematik und die Erarbeitung einer wissenschaftsgeleiteten Forschungsdatenstrategie.

Aus diesen Gründen hat die Mathematisch-Naturwissenschaftliche Klasse Open Research Data (ORD) als Arbeitsthema aufgegriffen und zu ORD in den Naturwissenschaften und der Mathematik am 11.10.2021 einen Workshop durchgeführt. In diesem Rahmen berichteten die Klassenmitglieder Matthias Beller, Matthias Scheffler, Matthias Steinmetz und Bernd Sturmfels über vier laufende NFDI-Konsortien, deren Ziel die Erarbeitung von ORD-Konzepten für unterschiedliche Bereiche der Naturwissenschaften und Mathematik ist. Die Klasse diskutierte über grundsätzliche Aspekte von ORD-Konzepten und vereinbarte die Erarbeitung eines Positionspapiers zu Voraussetzungen und Rahmenbedingungen ihrer wissenschaftsadäquaten Umsetzung. Das vorliegende Papier soll die fächerübergreifende Diskussion der ORD-Thematik in der BBAW unterstützen und zur Entwicklung einer inhaltlichen Stellungnahme der BBAW beitragen. Es richtet sich darüber hinaus an die breitere Fachöffentlichkeit und Förderinstitutionen.

1 AUSGANGSLAGE

In den Naturwissenschaften und der Mathematik liegen Primärdaten traditionell als proprietäre Datenbestände dezentral bei den sie erzeugenden Arbeitsgruppen vor. Sie werden in fachspezifischen – meist durch die experimentellen Infrastrukturen oder IT-Architekturen bestimmten – Formaten gespeichert. Auf diese Weise werden auch Dokumentationspflichten im Kontext guter wissenschaftlicher Praxis erfüllt. Übergreifende Metadaten existieren nur in Ausnahmefällen, was eine spätere Auswertung oder Verwendung der Daten in den Arbeitsgruppen oder durch Dritte stark erschwert und in vielen Fällen sogar verhindert.

Im Gegensatz hierzu werden Datenbestände aus großen experimentellen Infrastrukturen, z.B. internationale Large Scale Facilities im Bereich der Teilchenphysik, Astrophysik, Röntgenstrukturforschung etc., durch die zentrale Einrichtung gespeichert und den Nutzerinnen und Nutzern für Analysen und Interpretationen zur Verfügung gestellt. Large Scale Facilities wie CERN oder SLAC verwalten Datenbestände im Umfang von Hunderten von Petabytes und verwenden bei der Erfassung von Messdaten teilweise automatisierte Algorithmen zur Datenreduktion. Die Einrichtungskosten für derartige Infrastrukturen liegen im Bereich einiger 10^8 €, hinzu kommen hohe laufende Kosten für Betrieb und Personal. Jenseits von Large Scale Facilities existieren ggw. nahezu keine Infrastrukturen zur Bereitstellung großer oder komplexer Datenmengen. Wissenschaftliche Fachverlage und -gesellschaften verfügen in der Regel nicht über die hierfür erforderlichen Ressourcen.

Analyse, Interpretation und Publikation wissenschaftlicher Ergebnisse liegen gegenwärtig meist in der Hand der Arbeitsgruppe, welche die wissenschaftlichen Daten erzeugt hat. Damit werden unter anderem der wissenschaftliche Prioritätsanspruch und die Zurechenbarkeit der Ergebnisse zu individuellen Wissenschaftlerinnen und Wissenschaftlern gesichert. Neben die Publikation in referierten wissenschaftlichen Zeitschriften, Sammelbänden und Monografien tritt die Deposition nicht referierter Preprints in öffentlich zugänglichen Repositorien, z.B. arXiv. Experimentelle Primärdaten, aufbereitete Daten und/oder primäre Simulations- und Berechnungsergebnisse sind in der Regel nicht Bestandteil dieser Publikationen.

Erste Schritte zur Bereitstellung und breiteren Nutzung von Forschungsdaten wurden in einzelnen Bereichen der Naturwissenschaften und der Mathematik umgesetzt. In der **Teilchen- oder der Astrophysik** werden Datenbestände zwischen Einrichtungen und Arbeitsgruppen geteilt, um eine vernetzte Analyse und Interpretation zu erreichen. Die Ergebnisse werden häufig von Konsortien mit einer hohen Zahl von Autoren publiziert.

In der **Strukturforschung** wurden in Einzelfällen durch die originären Experimentatoren aufgearbeitete Rohdaten nach einer Karenzzeit für die Auswertung der wissenschaftlichen Community zur Verfügung gestellt, so z. B. von der LINAC Coherent Light Source (LCLS, SLAC Stanford), um verschiedene Ansätze zur dreidimensionalen Rekonstruktion von Nanopartikeln zu testen. Ein weiteres Beispiel ist die Protein Data Base (PDB), in der Strukturdaten einer Vielzahl biologischer Makromoleküle zusammengefasst und abrufbar sind. Die PDB ist heute eine der wichtigsten Quellen biologischer Strukturdaten.

Materialwissenschaften, Festkörperphysik, chemische Physik: Seit 2013 wird in Berlin das Projekt NOMAD (Novel Materials Discovery) entwickelt (offen seit 2014). Ab 2015 wurde es im Rahmen des Europäischen NOMAD Center of Excellence gefördert und seit Ende 2021 auch im Rahmen der NFDI.¹ Ziel der ersten Phase war es, alle Softwarepakete der *ab initio* Computational Materials Science zu unterstützen. Da INPUT- und OUTPUT-Dateien hochgeladen werden müssen, nicht nur Ergebnisse, können die Daten auf Konsistenz und Vollständigkeit überprüft werden. Anschließend werden sie dann auf eine einheitliche Form gebracht, die auch verwendbar ist, wenn Nutzerinnen und Nutzer mit der zugrundeliegenden Software nicht vertraut sind. NOMAD bietet seine Daten nicht nur FAIR² im herkömmlichen Sinne an, sondern als Findable and AI Ready an (AI: Artificial Intelligence). Wichtige Komponenten zu diesem Zweck sind die NOMAD Encyclopedia und das NOMAD Artificial Intelligence Toolkit.

Mit Milliarden von Daten ist NOMAD derzeit die weltweit größte Datenbank im Feld der Computational Materials Science. In letzter Zeit wird die Datenbasis auf die Daten der Soft Matter Community (force-field codes), auf experimentelle Daten und auf Daten aus dem Bereich des Quantum Computing erweitert. Die Nutzung der Daten beruht auf der CC-BY-Lizenz. Zum Hochladen von Daten ist eine Registrierung nötig. Ein beschränktes Teilen mit ausgewählten Personen ist bis zu drei Jahren möglich (Embargo). Um Daten zitierbar zu machen, können sie, sofern nicht unter Embargo, mit einem DOI versehen werden. Da Daten so behandelt werden wie eine wissenschaftliche Publikation, ist deren Veränderung oder Löschung ausgeschlossen.

In den **Geowissenschaften** werden große Datenmengen räumlich und zeitlich verteilter Geo- und Umweltparameter in bodenbasierten Netzwerken in Geophysik, Geodäsie, Hydrologie aber auch Satelliten- und Ozeandaten in Echtzeit sowie aus Klimamodellen erhoben. Bereits jetzt werden diese und zunehmend auch wissenschaftliche Software i.d.R. in zitierbaren Datenpublikationen (mit DOI) in öffentlichen, idealerweise fachspezifischen Forschungsdatenrepositorien in kuratierter Form und mit Metadaten gemäß den FAIR-Prinzipien veröffentlicht und archiviert. Wichtige geowissenschaftliche Repositorien in Deutschland sind z.B. PANGAEA, GFZ Data Services oder auch das World Data Center for Climate am Deutschen Klimarechenzentrum DKRZ. Grundlage sind standardisierte Metadaten, darunter ein weltweit eindeutiger und persistenter Identifikator physischer Proben (IGSN, www.igsn.org). In einem Abkommen der großen geowissenschaftlichen Gesellschaften mit Verlagen (Coalition for Publishing Data in the Earth and Space Sciences, COPDESS, <https://copdess.org>) wird für eine wissenschaftliche Veröffentlichung in einer Fachzeitschrift die Publikation der zugrundeliegenden Daten in Repositorien nach FAIR Prinzipien zur Bedingung gemacht. Mehrere NFDI-Projekte in den Geowissenschaften sind mit der Integration der verschiedenen Datenrepositorien befasst.

Mathematische Forschungsdaten reichen von Datenbanken für mathematische Objekte und spezielle Funktionen, über Aspekte des Wissenschaftlichen Rechnens (industrielle Anwendungen, Modelle und Algorithmen) bis hin zu statistischen Daten. Durch die Interdisziplinarität der Mathematik werden sie auch in anderen Wissenschaftsgebieten genutzt. Beispiele hierfür sind

¹ M. Scheffler, M. Aeschlimann, M. Albrecht, T. Bereau, H.-J. Bungartz, C. Felser, M. Greiner, A. Groß, C. T. Koch, K. Kremer, W. E. Nagel, M. Scheidgen, C. Wöll, and C. Draxl, FAIR data enabling new horizons for materials research. *Nature* 604, 635 (2022).

² FAIR steht für Findable, Accessible, Interoperable, Reusable. Siehe auch <https://forschungsdatenmanagement.bbaw.de/de>

der ATLAS of Finite Group Representations mit Detailinformationen über endliche Gruppen, polyDB, eine Datenbank für Objekte der diskreten Geometrie und verwandter Gebiete, die On-Line Encyclopedia of Integer Sequences, swMath, ein Informationsservice für mathematische Software und die MIPLIB, eine Sammlung von Daten industrieller Probleme zum Testen von Optimierungsalgorithmen. Mathematische Datensammlungen entstehen meistens auf Initiative kleiner Personengruppen und sind entsprechend prekär. Eine umfassende Datenkultur und entsprechende Infrastrukturen fehlen. Erste Versuche der Koordinierung werden z. B. durch die NFDI-Initiative MaRDI und das Repositorium MathRepo unternommen.

Der Einsatz automatisierter, etwa künstliche Intelligenz nutzender Prozeduren zur Auffindung, Auswahl und Beurteilung von Daten ist Gegenstand aktueller Forschung und bisher nur in Einzelfällen verwirklicht.

2 OPEN RESEARCH DATA UND FÄCHERKULTUR

Die Verwirklichung einer umfassenden ORD-Praxis in den Naturwissenschaften gemäß den FAIR-Prinzipien erfordert in den folgenden Bereichen eine internationale Verständigung über Konzepte, Inhalte und Regeln:

- **Neue Konzepte und Verfahren:** Datenzentrierte Wissenschaft unterscheidet sich in ihrer Methodik und im Umgang mit wissenschaftlicher Information von traditionellen Formen der Forschung. Wesentlich sind neue Konzepte und Verfahren der Datenanalyse und die Möglichkeit, große Datensätze aus unterschiedlichen Quellen zusammenzuführen. Die hierzu erforderlichen Methoden und Werkzeuge sind z.T. selbst Forschungsgegenstand.
- **Qualitätsstandards für Forschungsdaten:** Experimentelle und theoretische Ergebnisse müssen durch Metadaten umfassend beschrieben und charakterisiert sein und den Anforderungen der Interoperabilität und Reproduzierbarkeit so weit wie irgend möglich genügen. Die Anforderungen dieser Charakterisierung müssen durch die Fach-Community definiert werden. Die Veränderung/Verfälschung deponierter Datensätze muss ausgeschlossen werden; eventuelle Veränderungen können durch neue Versionen abgebildet werden.
- **Metadaten:** Vergleichbare Metadatenysteme und -formate wie auch innerhalb bestimmter Communities vereinbarte Terminologien und Ontologien sind unverzichtbar, um Datensätze angemessen zu beschreiben und suchbar zu machen. Metadaten besitzen einen obligatorischen Mindestumfang („Pflichtfelder“) und können optional um Detailinformationen erweitert werden. Ihre Nutzung erfordert leistungsfähige Softwareoberflächen und differenzierte Suchalgorithmen, die abgestufte Recherchen erlauben. Die Definition geeigneter Metadaten kann nur durch die jeweilige internationale Fachgemeinschaft erfolgen.
- **Datennutzung:** Die Entwicklung und breite Bereitstellung neuer Methoden des Data-Mining sind essentiell für das Auffinden und die effiziente Nutzung von Datensätzen. Eine wichtige Rolle kommt dabei automatisierten Suchverfahren zu, die u. a. auch künstliche Intelligenz nutzen.

Über diese inhaltlichen und technischen Anforderungen hinaus sind im Hinblick auf die Fächerkultur folgende Aspekte wichtig:

- **Urheberrechte:** Wissenschaftliche Priorität ist ein zentrales Kriterium bei der Beurteilung von Forschungsleistungen, dem Zugang zu Publikationsmedien im Peer-Review-Verfahren, der Einschätzung individueller Wissenschaftler:innen und schließlich der Besetzung akademischer Positionen. Die Öffnung proprietärer Datenbestände im Sinne von ORD erfordert deshalb Randbedingungen, unter denen Prioritätsansprüche und Urheberrechte der Datenerzeuger sicher gewahrt werden. Sinnvoll ist beispielsweise eine optionale Sperrfrist (Embargo) vor Veröffentlichung von Daten, in der eine primäre Analyse, Verwertung, z. B. des Intellectual Property, und Publikation von Ergebnissen durch die Datenerzeuger ermöglicht werden. Bei der Publikation von Ergebnissen, die auf deponierten ORD beruhen, ist die Nennung/Zitation der Datenquelle zwingend, die Regeln guter wissenschaftlicher Praxis sollten dieses Szenario explizit einschließen.
- **Wissenschaftliche Anerkennung:** Die Erzeugung, Aufbereitung und Publikation von Datensätzen stellen Leistungen dar, die für wissenschaftliche Erkenntnis und Fächerentwicklung gleichermaßen von Bedeutung sind. Die wissenschaftliche Community braucht, über die bestehenden oben genannten Qualitätsmaßstäbe hinaus, Kriterien, nach denen derartige Leistungen bewertet und anerkannt werden können. In diesem Kontext bedarf es auch einer gemeinsamen Begriffsbildung für an der Erhebung/Gewinnung/Erzeugung der Forschungsdaten Beteiligte. Wenn etwa Laborant:innen, Techniker:innen und Ingenieur:innen eine wesentliche Rolle für den Aufbau und Betrieb komplexer Messtechnik spielen und Anteil an der Datengewinnung haben, tragen sie auch einen Anteil der Verantwortung für die gewonnenen und später dann eventuell veröffentlichten Daten.
- **Verankerung in Curricula:** Um Erzeugung und Nutzung von ORD in der internationalen Wissenschaftsgemeinde zu entwickeln und umzusetzen, müssen die Grundlagen der Herstellung, Analyse und Bearbeitung digitaler Daten ein fester Bestandteil akademischer Curricula in den Naturwissenschaften werden. Dies erfordert die Kombination von Lehrinhalten, die traditionell getrennt und in verschiedenen Einzeldisziplinen unterrichtet werden.

3 OPEN RESEARCH DATA UND ÖFFENTLICHE FORSCHUNGSFÖRDERUNG

Die Umsetzung der FAIR-Prinzipien im Bereich der Forschungsdaten erfordert eine Erweiterung und Anpassung der Förderbedingungen und erzeugt gleichzeitig einen erhöhten Finanzierungsbedarf.

- **Wissenschaftsfreiheit:** Im ORD-Kontext muss die Wissenschaftsfreiheit gesichert sein, das heißt die Entscheidung über Art, Umfang und Zeitpunkt der Publikation von Daten wie auch die Auswahl des Datenrepositoriums liegen primär in der Verantwortung der Forschenden. Dabei sind Randbedingungen zu beachten, die durch die Forschungseinrichtung und/oder die Zuwendungsgeber gesetzt wurden.

- **Deckung der ORD-Kosten:** Erarbeitung und Bereitstellung von ORD sind mit einem hohen zusätzlichen personellen und finanziellen Aufwand verbunden, für den Zuwendungsgeber und Projektförderer angemessene Mittel bereitstellen müssen. Dies erfordert die Einführung entsprechender Haushaltstitel und/oder Förderformate, die auch fächerspezifische Projekte zur Aufbereitung, Bereitstellung und Deposition von ORD ermöglichen.
- **Dezentralität:** Bei der Ausgestaltung von ORD-Konzepten ist dem heterogenen und dezentralen Charakter der Forschungslandschaft Rechnung zu tragen und gleichzeitig der übergreifende administrative Aufwand zu begrenzen. Auch wenn aus Kostengründen eine partielle Bündelung von Kapazitäten sinnvoll ist, wird die Dateninfrastruktur dezentral erstellt und betrieben werden müssen. Aus Gründen der Datensicherheit muss das Gesamtsystem Redundanzen aufweisen, was mit zusätzlichen Kosten verbunden ist. Dezentrale Repositorien müssen über nutzerfreundliche Portale international und fachübergreifend koordiniert werden. Ihre Nutzung erfordert aktuelle Verzeichnisse und leistungsfähige Suchmaschinen (ein Beispiel ist <https://www.re3data.org>).

4 UMSETZUNG VON OPEN-RESEARCH-DATA-KONZEPTEN

- **Finanzen:** Die Errichtung einer leistungsfähigen ORD-Infrastruktur für deutsche Universitäten und Forschungseinrichtungen wird Finanzmittel in Höhe einiger Milliarden Euro erfordern. Für ihre Konzeption und Umsetzung wie auch die Definition von Regelwerken ist auf nationaler Ebene ein abgestimmtes Vorgehen von Wissenschaft, Zuwendungsgebern und Förderorganisationen unverzichtbar, hinsichtlich der Finanzierung auch eine Abstimmung zwischen Bund und Ländern. Die Belange der Wissenschaft müssen dabei im Vordergrund stehen, die rechtlichen Rahmenbedingungen international abgestimmt sein und der administrative Aufwand begrenzt werden. Ein zentraler Aspekt ist die internationale Vernetzung von ORD-Infrastrukturen unter vereinheitlichten Nutzungsbedingungen.
- **Public-Private-Partnership:** Kommerzielle technische Partner, d.h. Unternehmen, die große Speichersysteme in der Cloud betreiben oder einschlägige Hard- und Software-Entwicklungen verfolgen, können grundsätzlich wichtige Beiträge zu Einrichtung und Betrieb einer ORD-Infrastruktur leisten. Dabei müssen ein freier Datenzugang und die Einhaltung nationaler Datenschutz- und Datensicherheitsregeln sichergestellt sein. Eine entgeltliche Nutzung von Datenbeständen durch Firmen sollte grundsätzlich möglich sein, wobei Urheberrechte und die einschlägigen Regelungen des europäischen Beihilferechts zu beachten sind.
- **Wissenschaftliche Publikationen:** Zu klären ist die Rolle wissenschaftlicher Zeitschriften und Bücher, die von Fachgesellschaften oder kommerziellen Verlagen unter Peer Review publiziert werden, bei der Bereitstellung von ORD. Die Veröffentlichung großer Datensätze muss in frei zugänglichen, in der Regel öffentlich finanzierten Repositorien erfolgen.
- **Best Practice:** Bei der Erstellung und Umsetzung von ORD-Konzepten sind kompetente Beratung und die Schaffung von Best-Practice-Beispielen entscheidend für eine breite Ver-

ankerung in der Forschungslandschaft. Erfahrungen aus der BBAW-Initiative zum Forschungsdaten-Management zeigen, dass die Definition eines Datenmanagement-Plans im Rahmen eines Forschungsprojekts einen geeigneten Ausgangspunkt für die schrittweise Umsetzung darstellt. Ein dezentrales System von Beratungsstellen, etwa an Universitäten, bei Forschungs- und bei Förderorganisationen kann dieses Vorgehen direkt unterstützen.

5 FAZIT UND EMPFEHLUNGEN

Open Research Data (ORD) besitzen ein hohes Potential, um wissenschaftliche Informationen und Kompetenz zu vernetzen und damit neue Wege zur Schaffung umfassenden Wissens zu erschließen. Durch die Verbindung von Datenerzeugern und -nutzern können sie auch technologische Anwendungen und wirtschaftliche Nutzung naturwissenschaftlicher Erkenntnisse stark fördern. Bei Planung und Umsetzung einer ORD-Strategie sind die folgenden Gesichtspunkte entscheidend:

- Die internationale wissenschaftliche Fachgemeinschaft der jeweiligen Disziplin spielt die zentrale Rolle bei der Entwicklung und Umsetzung von ORD-Strategien. Dabei stehen die Anliegen und Bedarfe der Forschung im Vordergrund, das heißt neben der breiten Verfügbarkeit qualitätsgeprüfter Datensätze und der Definition von Metadaten die Entwicklung geeigneter Analyse- und Auswertewerkzeuge, die Wahrung wissenschaftlicher Prioritätsansprüche und die freie Entscheidung über Umfang und Art der Datendeposition. Nur so kann eine breite Akzeptanz ORD-basierter Forschungsansätze gelingen.
- Datenrepositorien und die zugehörige Infrastruktur müssen fachspezifisch und dezentral organisiert werden und eine Sicherheitsredundanz aufweisen. Eine Beteiligung international tätiger kommerzieller IT-Anbieter im Hinblick auf die technische Weiterentwicklung und Bereitstellung großer Speicherkapazitäten sollte nicht ausgeschlossen werden, muss aber Randbedingungen des Datenschutzes, des freien Zugangs zu den Daten und der Transparenz von Hard- und Software-Lösungen erfüllen. Die Veröffentlichung großer Datensätze muss in frei zugänglichen Repositorien erfolgen.
- Rechtliche und organisatorische Randbedingungen müssen die effiziente und zeitschonende Erzeugung und Nutzung von ORD ermöglichen. Dabei sind komplexe Genehmigungs- und Verwaltungsprozesse zu vermeiden und administrativer und/oder organisatorischer Aufwand auf ein Minimum zu begrenzen. Auch hier sind dezentrale, einrichtungsbezogene Lösungen erforderlich.

DANKSAGUNG

Die Autoren danken Matthias Beller, Matthias Steinmetz und Bernd Sturmfels für ihre Vorträge im Rahmen des ORD-Workshops und der Klasse für die eingehende Diskussion der Thematik. Matthias Scheffler dankt Claudia Draxl für umfangreiche Diskussionen und die Zusammenarbeit im NOMAD Projekt.

In der Reihe „Denkanstöße aus der Akademie“ erschienen bisher

1 / Nov 2015

Franz-Xaver Kaufmann, Hans Günter Hockerts, Stephan Leibfried,
Michael Stolleis, Michael Zürn

**Zur Entwicklung von Forschung und Lehre zur Sozialpolitik an Universitäten
in der Bundesrepublik Deutschland** (nur online)

2 / Dez 2018

Christoph Markschies

Zwei Texte zur Akademie der Wissenschaften im einundzwanzigsten Jahrhundert
(nur online)

3 / März 2020

Carola Lentz, Andrea Noll

**Wissenschaftskooperationen mit dem globalen Süden:
Herausforderungen, Potentiale und Zukunftsvisionen** (nur online)

4 / März 2021

Jochen Gläser, Wolf-Hagen Krauth, Christine Windbichler, Michael Zürn

**Befangenheit und Expertise in Berufungsverfahren:
Ein wissenschaftspolitischer Denkanstoß** (online und gedruckt)

5 / Juni 2021

Andreas Radbruch, Konrad Reinhart (Hrsg.)

Nachhaltige Medizin (online und gedruckt)

6 / Juni 2021

Jutta Allmendinger, Martin Mann, Lukas Haffert, Christoph Markschies
**Junge Wissenschaftler:innen und die Pandemie: Unterstützung und
systematische Verbesserungen – in der Krise und über die Krise hinaus**
(nur online)

7 / Nov 2021

Olaf Dössel, Tobias Schäffter, Gitta Kutyniok, Britta Rutert (Hrsg.)

Apps und Wearables für die Gesundheit (online und gedruckt)

8 / Dez 2021

Detlev Ganten, Max Löhning, Britta Rutert, Britta Siegmund

Gesundheitsregion Berlin-Brandenburg (online und gedruckt)

9 / Juli 2022

Jürgen Gerhards, Astrid Eichhorn, Julia Fischer, Ute Frevert und Christoph Markschies
**Klimaschutz und akademische Dienstreisen. Empfehlungen für ein umweltschonendes
Reiseverhalten in der Wissenschaft** (online und gedruckt)

