



nationale
Forschungsdaten
Infrastruktur
for CULTURE

RADAR4Culture: Ein Repositorium für Forschungsdaten zu materiellen und immateriellen Kulturgütern/1/

Neben der mittlerweile weit verbreiteten Open-Access-Publikation/2/ von Forschungsergebnissen in Textform entstehen in den Kulturwissenschaften und der GLAM-Community (Galleries, Libraries, Archives, Museums) auch Forschungsdaten, die nicht in Textform vorliegen, sondern andere Formate haben oder die Rohdaten dessen sind, was Forschende in einem Text wissenschaftlich abhandeln. Sowohl die textbasierte Publikation als auch die nicht-textbasierten Forschungsdaten brauchen einen Aufbewahrungsort, ein sogenanntes Repositorium, um sicher und zitierbar gespeichert und online zugänglich zu sein.

Für beide Fälle steht seit kürzerem im Umfeld des Konsortiums für materielle und immaterielle Kulturgüter, NFDI4Culture/3/, ein neuer leistungsstarker Service zur einfachen Publikation von Forschungsdaten zur Verfügung: RADAR4Culture./4/ Dieser Publikationsdienst ist ein wichtiger Baustein beim Unterfangen von NFDI4Culture, eine bedarfsgerechte Infrastruktur für die Forschungsdaten der Disziplinen Architektur-, Kunst- und Musik- bis hin zu Theater-, Tanz-, Film- und Medienwissenschaften bereitzustellen. Da viele Angebote zur Veröffentlichung von Forschungsdaten institutions- und disziplinabhängig sind, ist RADAR4Culture bewusst offengehalten und möchte explizit bestehende Lücken in der Repositorien-Landschaft schließen. RADAR4Culture basiert daher auf dem generischen Forschungsdatenrepositorium RADAR/5/, dem seit Jahren erfolgreichen und disziplinübergreifenden Dienst zum institutionellen Datenmanagement, das von FIZ

Karlsruhe – Leibniz-Institut für Informationsinfrastruktur bereitgestellt wird. FIZ Karlsruhe ist eine mitantragstellende Institution von NFDI4Culture und in vielen Teilbereichen des Konsortiums involviert, so auch im Bereich Datenpublikation und Langzeitarchivierung.

RADAR4Culture ermöglicht bereits seit dem 24. März 2022/6/ die kostenfreie und institutionenungebundene Veröffentlichung beliebiger Datentypen und -formate. Bei RADAR4Culture können Forschungsdaten mit beschreibenden Metadaten und Digital Object Identifiern (DOI) veröffentlicht werden/7/ – eine wichtige Voraussetzung für eine FAIRe/8/ Datenpublikation.

Wie RADAR4Culture als Repositorium zu einer FAIRen Datenhaltung beiträgt, zeigt sich bei einem Blick hinter die Kulissen – auf das Metadatenschema, das dem Dienst zugrunde liegt. Das Metadatenschema von RADAR findet auch bei RADAR4Culture Anwendung und basiert zu großen Teilen auf dem Metadatenschema der etablierten Non-Profit-Organisation DataCite./9/ Sowohl DataCite als auch RADAR bieten bezüglich einer FAIRen Datenhaltung interessante Möglichkeiten: Einige Felder der Schemata sind mit kontrolliertem Vokabular hinterlegt, und dort lassen sich unter anderem auch Normdatensätze bibliothekarischer Herkunft einsetzen. Diese Verbindung von Metadatenschema und Normdaten unter dem Aspekt der FAIR-Prinzipien soll im Folgenden genauer beleuchtet werden.

Zunächst wäre im Kontext der Kulturwissenschaften und GLAM-Einrichtungen auf ihrem Weg zu FAIRen Daten nochmals die Bedeutung von Normdaten hervorzuheben. Normdaten sind mit ihrer eindeutigen ID und zugehöriger URI nichts anderes als persistente Identifier. Ihre URI ist unveränderlich und dauerhaft im Netz erreichbar und dient damit auch in der Forschung als Referenzpunkt für bestimmte Entitäten. Die Auszeichnung von (Meta-)Daten mit PIDs wird beispielweise gleich im ersten FAIR-Prinzip zur Auffindbarkeit von Daten (F1) gefordert. Auf inhaltlicher Ebene entsteht durch die Verknüpfung von Forschungsdaten mit Normdaten ein Datenetz, das Informationen miteinander in Beziehung

setzt. Das macht sie für die Darstellung komplexer Zusammenhänge in den Kultur- und Geisteswissenschaften extrem nützlich. Normdaten wie beispielsweise die der GND bieten für diese Disziplinen, aber auch für den GLAM-Bereich die größte Schnittmenge an Entitäten. Über die GND-ID der jeweiligen Datensätze können musikalische Werke, literarische Werke, Bauwerke, Werke der Bildenden Kunst sowie Personen und Sachschlagwörter, die mit diesen in Zusammenhang stehen, eindeutig und persistent referenziert werden.

Da Normdaten ein kontrolliertes Vokabular darstellen, das auch in RDF (Research Description Framework) ausgespielt werden kann, erhöhen diese Identifikatoren die Maschinenlesbarkeit von Forschungsdaten. Normdaten sind also Knotenpunkte im digitalen Netzwerk von Forschungsdaten. Dieses Netzwerk kann aber nur entstehen, wenn auch die Metadaten zu den Forschungsdaten in Repositorien Normdaten-IDs enthalten und somit eine Vernetzung und eine verbesserte Auffindbarkeit von Daten zum gleichen Thema möglich wird. Daher ist es essenziell, dass Metadaten-schemata von Repositorien dies ermöglichen. Die Frage ist, wo kann man dort Normdaten-URLs hinterlegen?

Das Metadatenchema 4.4 von DataCite ist in drei Bereiche aufgeteilt: Einen Pflichtbereich, einen empfohlenen und einen optionalen Bereich.^{/10/} Insgesamt besteht das Schema aus 20 Properties; davon haben einige eine hierarchische Struktur, d. h. sie umfassen Sub-Properties. Die Properties haben verschiedene Eigenschaften: Manche können wiederholt werden, und manche haben ein kontrolliertes Vokabular hinterlegt.^{/11/} Zu den sechs verpflichtend auszufüllenden Feldern gehören: Identifier, Creator, Title, Publisher, PublicationYear, ResourceType. Die sechs empfohlenen Felder sind: Subject, Contributor, Date, RelatedIdentifier, Description, GeoLocation. Weitere acht

optionale Felder kommen noch hinzu: Language, Alternateldentifier, Size, Format, Version, Rights, FundingReference, RelatedItem.

Im Pflichtbereich des Metadatenchemas von DataCite gibt es die Möglichkeit, URLs bei „Creator“ einzutragen, dort wird angegeben, wer der Urheber von Forschungsdaten ist. Im empfohlenen Bereich bietet sich dieselbe Möglichkeit bei „Contributor“, um für die weiteren Beitragenden ebenfalls URLs einzutragen. Beide Elemente sind ähnlich aufgebaut. Im Beispiel^{/12/} von DataCite ist eine ORCID^{/13/} eingetragen. Sowohl „Creator“ als auch „Contributor“ können eine zugeordnete „Affiliation“ haben, diese wiederum kann ebenfalls mit einem Normdatensatz verknüpft werden. Hier ist es eine ROR-ID.^{/14/}

Normdaten können aber nicht nur in Bezug auf Personen oder Institutionen eingesetzt werden, sondern auch im „Subject“-Bereich von DataCite. Dieser gehört zu den empfohlenen Feldern. In diesem Bereich wird spezifiziert, was der Gegenstand der abgelegten Forschungsdaten ist, welche Personen, Orte, Objekte, Epochen oder Konzepte beispielsweise darin hauptsächlich behandelt werden. Dazu kann man bei „valueURI“ beispielsweise ein GND-Sachschlagwort verlinken, das den Forschungsgegenstand beschreibt. Da „valueURI“ wiederholt werden kann, sind hier mehrere Nennungen möglich, und es ist natürlich empfohlen, so viele Sachschlagworte wie möglich aufzulisten. Diese präzisieren nicht nur die Metadaten, sondern helfen auch bei der Suche innerhalb des Repositoriums. Mit der Möglichkeit, die sich über „valueURI“ bietet, wird den FAIR-Prinzipien F.2 sowie I.2, I.3 und R.1 entsprochen: Diese Prinzipien betreffen die Auffindbarkeit, Interoperabilität und Wiederverwertbarkeit von Daten und besagen, dass die Daten ausführlich mit Metadaten beschrieben sein sollen und kontrolliertes Vokabular enthalten sollen, das wiederum selbst FAIR ist. Im

```
<creators>
  <creator>
    <creatorName nameType="Personal">Miller, Elizabeth</creatorName>
    <givenName>Elizabeth</givenName>
    <familyName>Miller</familyName>
    <nameIdentifier schemeURI="https://orcid.org/" nameIdentifierScheme="ORCID">0000-0001-5000-0007</nameIdentifier>
    <affiliation>DataCite</affiliation>
  </creator>
</creators>
```

Abb. 1: ORCID-Eintrag in einem Beispiel von DataCite

folgenden Beispiel von DataCite sind Sachschlagwörter aus den Library of Congress Subject Headings eingetragen./15/

Da das Metadatenchema von RADAR auf DataCite basiert, bietet es dort ähnliche Möglichkeiten der Verknüpfung mit Normdaten. RADAR handelt aber einige Felder anders und geht auch über DataCite hinaus. Im Metadatenchema 9.1 von RADAR4Culture, das mit dem Release 1.15 verfügbar ist/16/, können von strukturierten Daten bis zu Normdaten verschiedene Arten von PIDs in personenbezogenen Angaben hinterlegt werden. Dies betrifft die Elemente „Creator“ und „Contributor“ sowie deren Unter-Element „Affiliation“. Bei diesen Elementen orientiert sich RADAR an DataCite 4.4, so ist beispielsweise für „Affiliation“ ebenfalls ROR implementiert, und für Personen können ORCIDs eingesetzt werden. Aber auch eine GND-ID wäre möglich.

Beim Feld „Publisher“ ermöglicht RADAR sowohl eine ROR-ID, falls es sich um eine Institution handelt, eine ORCID, falls es sich um eine Person handelt, oder einen Freitext. Damit geht RADAR schon einen Schritt weiter, denn diese Implementierung ist an DataCite 4.5 angelehnt.

Eine besondere Möglichkeit bietet RADAR bei der Vergabe von Sachschlagwörtern. Was bei DataCite über das Feld „valueURI“ im „subject“-Bereich möglich ist, in dem der Forschungsgegenstand beschrieben wird, setzt RADAR im Element „keywords“ um. „Keywords“ wiederum ist auf „subject“ gemappt und umfasst alle Attribute, die DataCite 4.4. für „subject“ enthält. Damit sind auch bei „keywords“ von RADAR dieselben Angaben wie bei DataCite möglich. Konkret können hier wieder Sachschlagwörter aus dem Normdatenbereich im Feld „valueURI“ abgelegt werden. Das Feld ist mehrfach wiederholbar und

sollte auch unbedingt so oft wie möglich belegt werden, d. h. es sollten so viele Sachschlagwörter wie möglich eingetragen werden. Das Besondere an diesem Feld ist aber, dass RADAR hier direkt die GND implementiert hat./17/ Die Nutzenden können im RADAR-Editor für die Metadaten entweder die Eingabe als Freitext wählen oder Vorschlägen der GND folgen. Bei der Wahl der GND wird bei der Eingabe eine Vorschlagsliste aus der GND angezeigt.

Dies ist ein wichtiger Schritt, um Dateneinliefernde dazu zu animieren, ihre Forschungsdaten mit reichen Metadaten auszuzeichnen und dabei auf die Vorteile von Normdaten zu setzen. Denn wie eine aktuelle Untersuchung zeigt, werden gerade die „subject“-Felder (bei RADAR „keywords“) bisher noch nicht erschöpfend genutzt:/18/ Während deutlich mehr als die Hälfte der Dateneinliefernden in den untersuchten Repositorien Angaben zum „subject“ machen, gibt aber quasi niemand davon etwas in „schemeURI“ an, was der Beschreibung des verwendeten Normdaten-Identifiers entspricht. Tatsächlich wird so gut wie kein kontrolliertes Vokabular in diesen Feldern genutzt.

RADARs Schritt, hier die GND anzubieten, ist daher nicht nur unter dem Gesichtspunkt der FAIRness von Metadaten zu Forschungsdaten, sondern auch in Bezug auf Linked Open Data/19/ begrüßenswert.

Zusammenfassend gesagt, kann über die Nutzung der IDs von Normdatenanbietern wie bspw. der GND in Repositorien wie RADAR4Culture vieles erreicht werden: Zum einen kann damit die Suche innerhalb des Systems von RADAR verbessert werden, zum anderen ist die Verwendung von kontrolliertem Vokabular in Metadaten und Forschungsdaten eine wichtige Voraussetzung zur

```
<subjects>
  <subject xml:lang="en" schemeURI="https://lccn.loc.gov/sh85062931" subjectScheme="LCCN">Temperature</subject>
  <subject xml:lang="en" schemeURI="https://lccn.loc.gov/sh85133712" subjectScheme="LCCN">Humidity</subject>
  <subject xml:lang="en" schemeURI="https://lccn.loc.gov/sh88003345" subjectScheme="LCCN">Classrooms</subject>
  <subject xml:lang="en">Ponhook Lake (N.S.)</subject>
</subjects>
```

Abb. 2: Sachschlagwörter aus den Library of Congress Subject Headings in einem Beispiel von DataCite

Erfüllung der FAIR-Prinzipien sowie für die Eignung der Daten, um von Graphtechnologien/20/ im Linked-Open-Data-Kontext erfasst zu werden.

Fragestellungen sowohl innerhalb eines Repositoriums als auch repositorienübergreifend könnten beispielsweise lauten: Wie viele Projekte zu einem bestimmten Gebäude oder Kunstwerk sind schon im Repositorium? Wie viele musikwissenschaftliche oder kunstgeschichtliche Forschungsprojekte sind vertreten? Gibt es Datenkorpora zu ähnlichen Inhalten usw. Über die eindeutigen IDs von Normdaten zu Sachschlagwörtern, Personen und Körperschaften können solche Fragestellungen beantwortet werden. Aber auch abgesehen von inhaltlichen Fragestellungen ermöglichen die entsprechenden IDs beispielsweise eine Abfrage nach der Anzahl von Datenkorpora je nach Institution oder Person. Damit lässt sich also auch der quantitative Aspekt von Forschungsarbeit beleuchten./21/

Die Nutzung von Normdaten in diesem Kontext bietet über die bibliothekarische Verwendung und die Einbindung in Forschungsdaten hinaus also viele Vorteile. Normdatenanbieter wie z. B. die GND stellen ein FAIRes, kontrolliertes, fächerübergreifendes Vokabular dar. Die GND vernetzt ihre Daten auch international mit Mappings zu anderen Systemen, Ontologien oder Thesauri wie z. B. ORCID, VIAF, LCSH oder RAMEAU./22/

Normdaten oder auch strukturierte Daten wie z. B. von ORCID, ROR oder auch geonames dienen als Zugriffs- und Ankerpunkte im Netz, sind maschinenverarbeitbar und verbessern Suchmöglichkeiten. Der qualitative Vorteil von bibliothekarischen Normdaten besteht dabei in den redaktionellen Prozessen und strengen Regeln für das Anlegen von Datensätzen. Außerdem haben diese Normdaten mit leistungsstarken Nationalbibliotheken als Anbieter eine hohe Zukunftssicherheit.

/1/ Dieser Beitrag entstand in Anlehnung an einen online gehaltenen Vortrag von Desiree Mayer und Melanie Größel beim 5. NFDI4Culture-Forum *Datenpublikation und -archivierung* zum Thema „Persistent Identifiers“ am 17.11.2022.

/2/ Frei zugängliche Publikation im Internet.

ROR setzt beim Thema Qualitätsmanagement auf einen zentralisierten, community-basierten Kuratierungsprozess, der es auch der Allgemeinheit ermöglicht, Ergänzungen oder Korrekturen über ein öffentliches Webformular vorzuschlagen. Diese Vorschläge werden von einem sogenannten Curation advisory board geprüft. Dieser Prüfprozess ist wiederum über github einsehbar und transparent. Geprüft wird, ob Vorschläge den Richtlinien für Metadaten von ROR entsprechen./23/ Während ROR also wie bibliothekarische Normdaten einen zentralisierten Redaktionsprozess betreibt, kann man sich bei ORCID nur selbst registrieren und kontrolliert damit auch, welche Daten man über sich als Person hinterlegt. Gemeinsam ist allen Anbietern von strukturierten Daten und Normdaten, dass sie sich zur Persistenz verpflichten und sich ihre IDs daher als URIs in Metadatenschemata von Repositorien eignen. Als PIDs bilden sie in den Metadaten die Basis für die Findbarkeit (F.1), Interoperabilität (I.2, I.3) und Wiederverwertbarkeit (R.1) ihrer zugehörigen Forschungsdaten.

RADAR4Culture ermöglicht mit seiner individuellen Umsetzung des DataCite-Metadaten-schemas die Nutzung von PIDs in den Metadaten und ist daher als neuer Publikationsdienst für Forschungsdaten aus den 4Culture-Fächern eine willkommene Ergänzung in der bisherigen Repositorienlandschaft, da es Forschende und Dateneinliefernde in die Lage versetzt, ihre Forschungsdaten gemäß den FAIR-Prinzipien und mit Anbindung an Linked-Open-Data-Technologien abzuliegen.

Desiree Mayer, Musikwissenschaftlerin, Wissenschaftliche Mitarbeiterin an der SLUB Dresden für NFDI4Culture.

/3/ NFDI4Culture – Konsortium für Forschungsdaten materieller und immaterieller Kulturgüter, <https://nfdi4culture.de/de/index.html> (28.11.2022).

/4/ RADAR4Culture, <https://radar.products.fiz-karlsruhe.de/de/radaraabout/radar4culture> (28.11.2022).

/5/ RADAR wird in Deutschland betrieben und unterliegt somit deutschem Recht. Das Speicherkontingent für RADAR-4Culture wird von FIZ Karlsruhe sowie den kooperierenden Rechenzentren am Karlsruher Institut für Technologie (KIT) und der Technischen Universität (TU) Dresden kostenfrei bereitgestellt; vgl. <https://radar.products.fiz-karlsruhe.de/de/radarsupport/faq> (28.11.2022).

/6/ Vgl. *Die Forschenden im Blick: maßgeschneiderte Forschungsdatenrepositorys für die Fachcommunitys Chemie und Kulturgüter*, www.fiz-karlsruhe.de/de/nachricht/die-forschenden-im-blick-massgeschneiderte-forschungsdatenrepositorys-fuer-die (28.11.2022).

/7/ Vgl.: <https://radar.products.fiz-karlsruhe.de/de/radarsupport/faq>.

/8/ Mark D. Wilkinson et al.: *The FAIR Guiding Principles for scientific data management and stewardship*, <https://www.nature.com/articles/sdata201618> (28.11.2022). FAIR steht für Findable, Accessible, Interoperable und Reusable.

/9/ Zum Zeitpunkt des Verfassens dieses Artikels war dies die aktuelle Version: *DataCite Metadata Schema 4.4, Released 30 Mar 2021*, <https://schema.datacite.org/meta/kernel-4.4/> (28.11.2022).

/10/ DataCite Metadata Working Group (2021). *DataCite Metadata Schema. Documentation for the Publication and Citation of Research Data and Other Research. Version 4.4*. DataCite e.V., https://schema.datacite.org/meta/kernel-4.4/doc/DataCite-MetadataKernel_v4.4.pdf, S. 7 (28.11.2022).

/11/ Vgl. ebd., S. 18–38 (28.11.2022).

/12/ *Full DataCite XML Example. Demonstration of DataCite Properties*, <https://schema.datacite.org/meta/kernel-4.4/example/datacite-example-full-v4.xml> (28.11.2022).

/13/ ORCID ist die Open Researcher and Contributor ID, ein kostenloser Service, bei dem sich Forschende selbst für eine PID registrieren können.

/14/ ROR steht für Research Organization Registry: „The Research Organization Registry (ROR) is a global, community-led

registry of open persistent identifiers for research organizations. ROR makes it easy for anyone or any system to disambiguate institution names and connect research organizations to researchers and research outputs“; vgl. *Research Organization Registry (ROR), About ROR*, <https://ror.org/about/#what-is-ror> (28.11.2022).

/15/ *Full DataCite XML Example. Demonstration of DataCite Properties (subject)*, https://schema.datacite.org/meta/kernel-4.4/example/datacite-example-Box_dateCollected_DataCollector-v4.xml (28.11.2022).

/16/ re3data.org: *RADAR4Culture; editing status 2022-11-24*; re3data.org – Registry of Research Data Repositories, <http://doi.org/10.17616/R31NJNAZ> (28.11.2022).

/17/ Für die GND-Implementierung wird die API von lobid-gnd verwendet; vgl. *lobid-gnd – Rechercheoberfläche & LOD-API für die GND*, <http://lobid.org/gnd> (28.11.2022).

/18/ Vgl. Dorothea Strecker (2021): *Quantitative assessment of metadata collections of research data repositories*. Humboldt-Universität zu Berlin, <https://doi.org/10.18452/22916>, Abb. 11, S. 37 (28.11.2022).

/19/ Linked Open Data sind Daten, die im World Wide Web frei zugänglich, eindeutig identifiziert und vernetzt sind.

/20/ Mittels Graphen können Beziehungen und vernetzte Informationen von Linked Open Data repräsentiert und dargestellt werden.

/21/ Vgl. Paul Vierkant (2022, November 16): ... – *data needs meta* – ..., NFDI4Culture-Forum Datenpublikation und -archivierung #5: Persistent Identifiers, <https://doi.org/10.5281/zenodo.7326660>, Folie 16 und 17 (28.11.2022).

/22/ Vgl. Jakob Voss, Yvonne Jahns (2018): *Übersicht von GND-Konkordanzen*, <https://doi.org/10.5281/zenodo.1689997> (28.11.2022).

/23/ Vgl. Research Organization Registry (ROR), *Registry*, <https://ror.org/registry/> (28.11.2022).

Bis zur letzten Minute spannend und so vielfältig wie der Beruf – Die Jahrestagung 2022 in Düsseldorf

Am letzten Tagungstag leitet Cortina Wuthe die Vorträge mit den Worten „das Programm ist so vielfältig wie unser Beruf“ ein, aber diese Worte können eigentlich als Überschrift für die gesamte Tagung verwendet werden. Vielfalt der Themen und Referent*innen wie auch Vielfalt der Formate

und Veranstaltungsorte zeichnen die diesjährige Jahrestagung der IAML Deutschland vom 13. bis 16. September 2022 aus. Zwei Teilnehmende sind in Düsseldorf zum ersten Mal dabei: Andrea Ciocchetti von der Hochschule für Musik und Tanz Köln und Kristina Pott aus der Stadtbibliothek Essen, die beide in diesem Jahr als Seiteneinsteiger*innen ihre neuen Positionen angetreten haben. Sie berichten an dieser Stelle gemeinsam von ihren Eindrücken.

Ein roter Faden, sogar mehrere Fäden verbinden die vielseitigen Veranstaltungen und Themen, und