

## KOLUMNE

# (Open) Research Data Publications

## Dataverse und GRO.data für Forschungsdatenpublikationen

**Annette Strauch-Davey**

Offener Zugang und Bereitstellung von Informationen ist durch Forschungsdaten möglich, die von allen genutzt, geteilt und weiterverwendet werden dürfen. Die Digitalisierung hat längst neue Aufgabengebiete in Universitätsbibliotheken im Bereich Publizieren geschaffen. Diese Kolumne richtet ihr Augenmerk auf das Publizieren von Forschungsdaten und auf die Verfügbarmachung von Forschung im Forschungsdaten-Repositorium GRO.data,<sup>1</sup> einem universellen Repositorium, welches auf der Software Dataverse basiert und gute Möglichkeiten zur Forschungsdatenpublikation bietet.

Im Dezember 2020 schaltete die DFG das Portal „Wissenschaftliche Integrität“ frei. Hier soll nach und nach ein Referenzwerk zur „Guten Wissenschaftlichen Praxis“ entstehen. Es versucht damit den DFG-Kodex von 2019 praxisnah zu kommentieren. Das Publizieren von Forschungsdaten bezieht sich im Kodex vor allem auf das „Data Sharing“, wobei Forschende beim Teilen ihrer Forschungsdaten für ihre Datenerhebung Anerkennung erhalten sollen. In den Förder Richtlinien der DFG von 2020 heißt es:

*„Gehen Sie auch auf die Möglichkeit der Nachnutzung durch andere Wissenschaftlerinnen und Wissenschaftler ein. Bitte berücksichtigen Sie dabei auch – sofern vorhanden – die in Ihrer Fachdisziplin existierenden Standards und die Angebote existierender Datenrepositorien oder Archive.“<sup>2</sup>*

Aufgrund der Richtlinien mussten wir für das Forschungsdatenmanagement unter anderem über Forschungsdaten-Repositorien nachdenken. Ein einfaches Digital-Asset-Management mit geschlossenen



*Annette Strauch-Davey M.A., Forschungsdatenmanagement (FDM) /Research Data Management (RDM), Stiftung Universität Hildesheim, Universitätsbibliothek (UB)*

Pools reicht für vertrauenswürdige Daten im Sinne von FAIR einfach nicht aus. In den Datenmanagementplänen zeigen wir, dass die Forschungsvorgehensweisen allen Anforderungen der Forschungsförderer entsprechen, auch wo die Daten-Veröffentlichungen geplant sind. Die wissenschaftspolitischen Forderungen konnten somit zur Weiterentwicklung der Dataverse-Repositorien und zu den Zielen der NFDI beitragen.

*„Die nationale Forschungsdateninfrastruktur (NFDI) soll die Datenbestände von Wissenschaft und Forschung systematisch erschließen, nachhaltig sichern und zugänglich machen sowie (inter-)national vernetzen.“<sup>3</sup>*

Zu einer Verknüpfung von Forschungsdaten trägt die Datenveröffentlichung auf GRO.data bei. Dieses offene Forschungsdaten-Repositorium basiert auf dem Open-Source-Programm *Dataverse* der Harvard University.

1 <https://data.goettingen-research-online.de/> [25.Juni 2021].

2 [https://www.dfg.de/formulare/54\\_01/54\\_01\\_de.pdf](https://www.dfg.de/formulare/54_01/54_01_de.pdf) [25.Juni 2021].

3 <https://www.dfg.de/foerderung/programme/nfdi/> [25.Juni 2021].

*“Harvard’s open online repository for sharing, preserving, citing, exploring and analyzing research data.”<sup>4</sup>*

Dataverse ist ein generisches Forschungsdaten-repositorium. Ein Dataverse-Repositorium ist eine Softwareinstallation, die mehrere virtuelle Archive hostet, die als Dataverse-Sammlungen (Dataverse Collection)<sup>5</sup> bezeichnet werden. Jede Dataverse-Sammlung enthält Datensätze, und jeder Datensatz schließt beschreibende Metadaten und Dateien (inklusive Dokumentation und Code, die den Daten beigefügt sind) ein.

Ein Dataverse ist ein Container für Datensätze (Forschungsdaten, Code, Dokumentation und Metadaten) und andere „Dataverses“, die für einzelne Forscherinnen und Forscher, Abteilungen an Universitäten, Zeitschriften und Organisationen eingerichtet werden können.

GRO.data wird von der GWDG gehostet, die sich um Datensicherheit und -schutz, einschließlich Backups und Zugriffskontrolle und um die Verfügbarkeit und Zuverlässigkeit kümmert.<sup>6</sup> Die Academic Cloud, über die man sich einloggt, ist eine gemeinsame Infrastruktur des LANIT/RZ der Hochschulrechenzentren in Niedersachsen. Sie ist für 210.000 Nutzer konzipiert. Das Projekt wird vom Niedersächsischen Ministerium für Wissenschaft und Kultur gefördert. Inzwischen bietet die eResearch-Alliance über das Göttingen Campus sogenannte „Data Labs“ an.<sup>7</sup> In diesen Schulungen werden Zugang, das Anlegen von Datensätzen und Metadaten sowie die Zusammenarbeit mit anderen Forscherinnen und Forschern “hands-on” demonstriert. Für die Kommunikation zu Dataverse weltweit gibt es ein Konsortium, das globale “Dataverse Community Consortium”.<sup>8</sup> In dieser Arbeitsgemeinschaft tauschen sich hauptsächlich die Entwicklerinnen und Entwickler aus. Auch eine deutschsprachige Dataverse-Community befindet sich derzeit im Aufbau. Forschungsdaten aus den Repositorien ermöglichen neue Forschungspotentiale.

Die Datenveröffentlichungen beziehen sich auf originäre wissenschaftliche Forschungsergebnisse, Ursprungsdaten, Metadaten, Quellenmaterial, digitale Darstellungen von Bild- und Graphik-Material und wis-

senchaftliches Material in multimedialer Form.<sup>9</sup> Forschende, die im Open Access Forschungsdaten auf GRO.data publizieren, gewähren das freie, weltweite Zugangsrecht zu ihren Veröffentlichungen.

Wissenschaftliche Bibliotheken orientieren sich am Papier Wissenschaftliche Bibliotheken 2025, das von der Sektion 4 „Wissenschaftliche Universalbibliotheken“ im Deutschen Bibliotheksverband e.V. (dbv) im Januar 2018 beschlossen wurde,<sup>10</sup> um bei der Arbeit mit den Forschenden die Maximierung der Sichtbarkeit und Transparenz zu ermöglichen. In einem Dataverse-Repositorium für Forschungsdaten wie GRO.data können Forscherinnen und Forscher also ihre Daten dauerhaft sichern, sie nachhal-

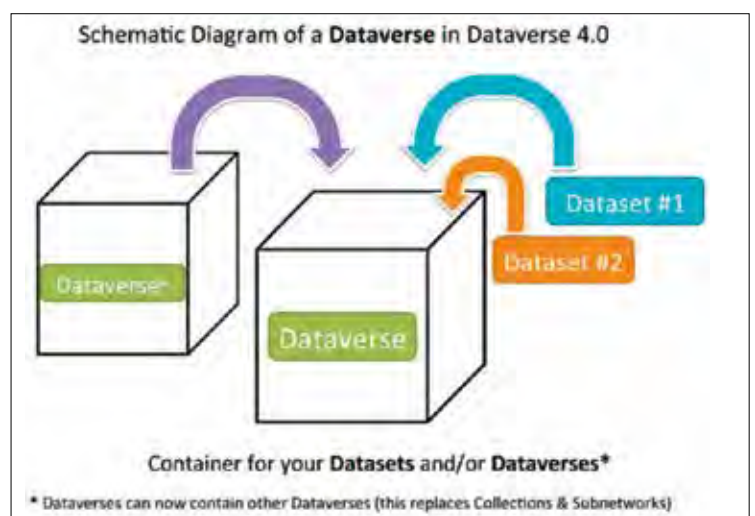


Abbildung 1: Schematisches Diagramm eines Dataverses, <https://guides.dataverse.org/en/latest/user/dataverse-management.html> [25. Juni 2021]

tig und qualitätsgerecht öffentlich im Open Access publizieren. Die Forschungsdaten erhalten bei der Datenpublikation einen persistenten Identifikator (DataCite DOI),<sup>11</sup> der bei DataCite dann automatisch indiziert und zum Harvesting über eine OAI-Schnittstelle angeboten wird. Die DOI-Vergabe geschieht beim Klick auf die Option “Data Publication” zur Datenveröffentlichung.

### Metadatenmanagement

Bibliotheken entscheiden sich für das Dataverse, insbesondere wegen des guten Arbeitens mit verschie-

4 <https://dataverse.harvard.edu/> [25. Juni 2021].

5 <https://dataverse.org/researchers> [25. Juni 2021].

6 <https://info.gwdg.de/news/neue-dienste-in-der-academic-cloud-gwdg-meet-chat-dataverse/> [25. Juni 2021].

7 <https://www.eresearch.uni-goettingen.de/de/news/12-01-2021-eresearch-lab-on-gro-data/> [25. Juni 2021].

8 <http://dataversecommunity.global/> [25. Juni 2021].

9 <https://openaccess.mpg.de/Berliner-Erklärung> [25. Juni 2021].

10 [https://www.bibliotheksverband.de/fileadmin/user\\_upload/Sektionen/Publikationen/WB2025\\_Endfassung\\_endg.pdf](https://www.bibliotheksverband.de/fileadmin/user_upload/Sektionen/Publikationen/WB2025_Endfassung_endg.pdf) [25. Juni 2021].

11 <https://datacite.org/doi.html> [25. Juni 2021]. <https://www.doi.org/factsheets/DOIKeyFacts.html> [25. Juni 2021].



Abbildung 2: Webseite Göttingen Research Online Data (GRO.data), <https://data.goettingen-research-online.de/> [25.Juni 2021]

denen Metadaten schemata. Um Forschungsdaten finden zu können, („Findable“), müssen diese immer mit Metadaten fachgemäß beschrieben werden. Es können disziplinspezifische Metadatenätze ausgewählt werden, und sobald ein Dataset veröffentlicht wurde, können seine Metadaten in eine Vielzahl anderer Metadatenstandards und -formate exportiert werden, wodurch Datasets in anderen Systemen, z.B. anderen Datenrepositorien, besser erkennbar und verwendbar werden.

Auf der Metadatenregisterkarte jeder Dataset-Seite stehen die folgenden Exporte zur Verfügung:

- Dublin Core
- DDI (Data Documentation Initiative Codebook 2.5)
- DDI HTML Codebook (eine besser lesbare HTML-Version des DDI Codebook 2.5-Metadatenexports)
- DataCite 4
- JSON (natives Dataverse-Format)
- OAI\_ORE
- OpenAIRE
- Schema.org JSON-LD

Jeder dieser Metadatenexporte enthält die Metadaten der zuletzt veröffentlichten Version des Datasets. Technisch kann der Service auch an elektronische Laborbücher (ELN) anbinden. Es gibt die Möglichkeit der einfachen Verlinkung veröffentlichter Daten im ORCID-Profil der Forscherinnen und Forscher. Die Dataverse-Software bietet mehrere Methoden zum Hochladen von Dateien in einen Datensatz. Dateien in bestimmten Formaten wie Stata, SPSS, R, Excel (.xlsx), CSV und TSV können als Tabellendaten aufgenommen werden.

Dataverse wendet standardmäßig CC0<sup>12</sup> auf Datensätze an, da es die Wiederverwendung, Erweiterbar-

keit und langfristige Aufbewahrung von Forschungsdaten erleichtert, indem sichergestellt wird, dass die Daten von jedermann sicher verarbeitet werden können, ohne dass potenzielle Urheberrechtsprobleme befürchtet werden müssen. Die Veröffentlichung mit Dataverse muss rechtssicher sein. Alle Fragen, insbesondere zum Datenschutz und zum Urheberrecht müssen vor einem Ingest in das Datenrepositorium geklärt worden sein.<sup>13</sup> Das Dataverse ermöglicht dann somit auch das „R“ von FAIR, und Forschungsdaten können reproduziert werden.

### Ausblick

Das Dataverse-Repositorium mit GRO.data schafft einen weiteren Publikationskanal für Forschungsdaten. Dieser wird die Sichtbarkeit der lokalen Forschung erhöhen können, für alle die, die für eine Forschungsdatenveröffentlichung und Bereitstellung ihrer Daten bereit sind. Die FAIR-Data-Prinzipien (Findable, Accessible, Interoperable, Re-usable) helfen dabei, die Forschungsdaten und die Metadaten weiter aufzubereiten.<sup>14</sup>

### Open Data Impact Award

Der Preis soll Open Data in der Wissenschaft stärken. Forscherinnen und Forscher oder Forschungsgruppen aller Disziplinen an Hochschulen oder außeruniversitären Forschungsinstitutionen mit einem Sitz in Deutschland konnten sich bis zum 15. August 2021 für diese Auszeichnung für offene Daten und Innovationen bewerben. Nachnutzbare Forschungsdaten und Überprüfungsmöglichkeiten der Datensätze anderer waren bei der Bewerbung "Open Data" essentiell, die ihre Forschungsdaten offen gestellt und eine innovative Nachnutzung außerhalb der Wissenschaft ermöglicht haben oder das Potenzial für einen gesellschaftlichen Impact durch Nachnutzung ihrer Daten aufzeigen. Wir sind gespannt auf die offenen Forschungsdaten. Weitere Informationen zum Open Data Impact Award finden Sie unter: <https://www.stifterverband.org/innosci/open-data-impact-award> |

### Literatur

- Klump, J. (2017). Data as Social Capital and the Gift Culture in Research. *Data Science Journal*, 16, doi: <http://doi.org/10.5334/dsj-2017-014>
- Rathmann, T. (2018). Anforderungen der Forschungsförderer an das Forschungsdatenmanagement (Version 1). Zenodo. doi: <http://doi.org/10.5281/zenodo.1464993>

<sup>12</sup> <https://creativecommons.org/share-your-work/public-domain/cc0/> [25.Juni 2021].

<sup>13</sup> <https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/daten-publizieren/entscheidungshilfe-daten-veroeffentlichen/> [25.Juni 2021].

<sup>14</sup> <https://www.ands.org.au/working-with-data/fairdata> [25.Juni 2021].