

Brandenburg. Geowiss. Beitr.	Cottbus	Bd. 31/2024	S. 77–80	1 Abb., 7 Lit.
------------------------------	---------	-------------	----------	----------------

FID GEO Services: Daten und Softwarepublikationen mit GFZ Data Services

KIRSTEN ELGER, MELANIE LORENZ, INKE ACHTERBERG & MALTE SEMMLER

Die Geowissenschaften im Allgemeinen und die Geologie im Besonderen, sind zunehmend datengetrieben. Die fortschreitende Digitalisierung in Verbindung mit neuen Methoden (z. B. Künstliche Intelligenz, Maschine Learning) bietet immer mehr Möglichkeiten im Umgang mit digitalen Forschungsdaten. Damit verbunden sind aber auch neue Erwartungen an die Datenurheber und Forschende. Dazu gehört in erster Linie die Forderung nach freier Zugänglichkeit, d. h. nach der Veröffentlichung aller Forschungsergebnisse, bei der Forschungsdatenrepositorien eine gute Unterstützung sein können.

Offene Wissenschaft - Hintergründe

Seit die G8-Wissenschaftsminister 2013 die *Open Data Charter* unterzeichnet und damit ein erstes deutliches Bekenntnis zu Open Science von ministerieller Seite abgegeben haben, wird weltweit verstärkt an Möglichkeiten und Richtlinien gearbeitet, die „mit Steuergeldern erhobenen“ Daten, auf denen wissenschaftliche Ergebnisse basieren, auch „dem Steuerzahlenden“ zur Verfügung zu stellen. Diese Forderung ist inzwischen unter dem Konzept der Datenveröffentlichung „so offen wie möglich, so geschlossen wie nötig“ (Council of the European Commission 2016) zu einem zentralen Bestandteil der Wissenschaftspolitik in Europa geworden. Mit der wissenschaftspolitischen Verankerung der FAIR-Prinzipien in der europäischen Forschungsförderung (European Commission 2018) gewinnt die Forderung nach der Veröffentlichung von Forschungsdaten weiter an Bedeutung und ist heute fester Bestandteil der Data Policies von Forschungsförderern, wissenschaftlichen Einrichtungen und Zeitschriften.

Die *FAIR-Prinzipien* für den Umgang mit Forschungsdaten wurden zum ersten Mal von WILKINSON et al. (2016) beschrieben. FAIR setzt sich aus Auffindbarkeit (Findable), Zugänglichkeit (Accessible), Interoperabilität (Interoperable) und Wiederverwendbarkeit (Reusable) zusammen und definiert damit in kompakter Form den adäquaten Umgang mit Forschungsdaten. Zusammenfassend lässt sich sagen, dass die Prinzipien darauf abzielen, die Weitergabe und Nachnutzung von Daten zu verbessern, um sowohl die Reproduzierbarkeit von Forschungsergebnissen als auch die Verarbeitung

und ggf. Neuinterpretation der Daten durch Menschen und Maschinen (Computer) zu ermöglichen. Die FAIR-Prinzipien können auf die Daten selbst und auf die Daten, die sie beschreiben (deren Metadaten), angewendet werden. Metadaten sind der Schlüssel zur Umsetzung der FAIR-Prinzipien. Nur wenn beispielsweise die vollständige Provenienz eines verfügbaren Datensatzes in den Metadaten beschrieben ist, kann später abgeschätzt werden, welche Datensätze für welche Weiterverarbeitungsansätze geeignet sind.

Für die Geowissenschaften sind die beiden Selbstverpflichtungserklärungen der „Coalition for Publishing Data in the Earth and Space Sciences“ (COPDESS, <https://copdess.org/>) wegweisend für die Veröffentlichung von Forschungsergebnissen. In beiden Selbstverpflichtungserklärungen haben sich, unter anderen, führende Wissenschaftsverlage (u. a. Springer Nature, Science, Elsevier, AGU, EGU) verpflichtet, die Bereitstellung und Veröffentlichung der Daten, die den in ihren Artikeln beschriebenen wissenschaftlichen Ergebnissen zugrunde liegen, zur Bedingung für die Veröffentlichung der Artikel zu machen. Dies soll nicht mehr in Form von Datensupplementen zu den Artikeln geschehen, sondern als eigenständige Datenpublikation, idealerweise über fachspezifische Forschungsdatenrepositorien, welche die FAIR-Prinzipien umsetzen. Die COPDESS-Erklärung betont zudem den Wert von Kuratation durch Fachrepositorien und der Verknüpfung von Ressourcen mit Hilfe von Metadaten.

Im Unterschied zu institutionellen oder allgemeinen Repositorien, bieten *fachspezifische Repositorien* oft die Kuratation von Daten und Metadaten durch Fachwissenschaftler an, welche die Forschenden bei der Publikation und Dokumentation ihrer Forschungsdaten unterstützen und sie anleiten. Der fachspezifische Fokus bietet zudem die Möglichkeit, spezifischere Metadatenmodelle zu nutzen als dies bei allgemeinen Repositorien möglich ist.

FID GEO – ein bibliotheksgetriebener Informationsdienst für die offene Wissenschaft

Der *Fachinformationsdienst Geowissenschaften (FID GEO, <https://fidgeo.de>)* ist ein bibliotheksgetragenes Ser-

viceprojekt, das mit seinem Dienstleistungsportfolio einen Beitrag zur Bewältigung der bestehenden Herausforderungen für Open Science leisten will. Gefördert von der Deutschen Forschungsgemeinschaft (DFG) bietet der FID GEO verschiedene Dienstleistungen zur Veröffentlichung von Forschungsergebnissen und zur Unterstützung der geowissenschaftlichen Community in Deutschland an. Über das an der Niedersächsischen Staats- und Universitätsbibliothek Göttingen angesiedelte Fachrepositorium für Texte und geologische Karten GEO-LEOe-docs (<https://e-docs.geo-leo.de/>) können wissenschaftliche Beiträge in Form von Fachartikeln, Tagungsbänden usw. veröffentlicht werden. Die Digitalisierung und Veröffentlichung älterer Forschungsergebnisse und Publikationen wird hier explizit unterstützt. Die damit verbundenen Angebote werden in einem weiteren Beitrag näher vorgestellt (SEMMLER et al. dieser Band, S. 81–83). Forschungsdaten und wissenschaftlicher Software können über das assoziierte geowissenschaftliche Fachrepositorium GFZ Data Services am Deutschen GeoForschungsZentrum in Potsdam veröffentlicht werden. Weitere zentrale Dienstleistungen des FID GEO sind Beratungen und Schulungen, in denen vor allem vermittelt wird, wie Forschungsergebnisse optimal publiziert und verlinkt werden können.

GFZ Data Services – fachspezifisches Forschungsdatenrepositorium für die Geowissenschaften

Ziel dieses Beitrags ist es, das Angebot und die Möglichkeiten zitierfähiger Daten- und Softwarepublikationen vorzustellen und wichtige Zusammenhänge zu erläutern. Am Beispiel von GFZ Data Services soll gezeigt werden, wie das Zusammenspiel von Forschenden und Repositorien zur Veröffentlichung gut beschriebener und für Mensch und Maschine nachnutzbarer Forschungsdaten führen kann und wie darüber hinaus die Zusammenarbeit und der Austausch mit Geoportalen funktionieren kann.

GFZ Data Services ist das Fachrepositorium für geowissenschaftliche Daten am Deutschen GeoForschungsZentrum GFZ (<https://dataservices.gfz-potsdam.de/>). *Fachrepositorien* sind digitale Archive, die auf kuratierte Forschungsdaten (und/oder Software) aus bestimmten wissenschaftlichen Fachgebieten spezialisiert und beschränkt sind. Die veröffentlichten Objekte werden durch die Vergabe eines „Digital Object Identifier“ (DOI) zitierbar gemacht und mit standardisierten Metadaten angereichert. Metadaten enthalten einerseits bibliographische Informationen, die für die Zitation und Nachnutzung der Publikationen notwendig sind. Zum anderen wird die Auffindbarkeit der Daten durch weitere (fachspezifische) Metadaten, wie geographische Koordinaten, Beschreibung der Daten durch kontrollierte, geowissenschaftliche Vokabularien, unterstützt. Auch die Verlinkung der publizierten Daten mit zugehörigen wissenschaftlichen Artikeln oder anderen Datensätzen ist Teil dieser Metadaten. Über

GFZ Data Services (<https://dataservices.gfz-potsdam.de>) werden Daten, Datensammlungen, Datenprodukte und auch wissenschaftliche Software mit DOI veröffentlicht und durch umfangreiche Metadaten und standardisierte Datenbeschreibungen ergänzt. Der Fokus liegt zum einen auf dem sogenannten „Long Tail“ der Forschungsdaten (HEIDORN 2008), den vielen kleinen und sehr unterschiedlichen Datensätzen, die von einzelnen Forschenden oder kleineren Arbeitsgruppen erzeugt werden. Andererseits werden zunehmend Publikationsservices für internationale Projekte und Dienste entwickelt (z. B. für die World Stress Map, für das International Continental Scientific Drilling Programme ICDP, verschiedene internationale geodätische Services u. v. m.). Wie oft in fachspezifischen Repositorien, wird ein besonderer Wert auf beschreibende Metadaten gelegt. Die Verwendung von internationalen Metadatenstandards für Data Discovery (DataCite, ISO19115/INSPIRE) und die Verschlagwortung der Datensätzen mit Hilfe von kontrollierten linked-data Vokabularien der Geowissenschaften [z. B. von der NASA Global Change Master Directory (GCMD), der Commission for the Management and Application of Geoscience Information der International Union for Geological Sciences (IUGS CGI) oder auch der Infrastructure for Spatial Information in the European Community (INSPIRE) Direktive] ermöglicht einerseits eine strukturierte Suche und erhöht andererseits die Auffindbarkeit der Daten. Die Metadaten werden über standardisierte Schnittstellen angeboten und die Implementierung von Schema.org erhöht zusätzlich die Auffindbarkeit unserer Datenpublikationen für Internet-Suchmaschinen, wie Google oder Google Dataset Search. Die Daten selbst werden unter offenen Lizenzen (z. B. Creative Commons) per Downloadlink oder Kontaktformular angeboten (siehe auch ELGER et al. 2018, PAMPPEL & ELGER 2021).

Um Forschende bei der Erfassung von Metadaten und der Beschreibung bzw. Dokumentation ihrer Daten zu unterstützen, bietet GFZ Data Services eine Reihe von Werkzeugen an. Ein Online-Metadaten-Editor übersetzt die von den Forschenden eingegebenen Informationen in maschinenlesbare XML Metadaten (Abb. 1 links). Der Metadaten-Editor dient der möglichst umfassenden, fachspezifischen Beschreibung der zu veröffentlichenden Daten durch maschinenlesbare Metadaten, die von den Nutzern selbständig ausgefüllt werden. Im Metadaten-Editor sind zudem standardisierte und kontrollierte Fachvokabulare hinterlegt, die über ein Auswahlménü ausgewählt werden können. Die Erfassung der räumlichen Ausdehnung des Untersuchungsgebietes erfolgt über eine interaktive Karte (Abb. 1). Dies ist gleichzeitig eine wichtige Qualitätskontrolle der geographischen Koordinaten. Die im Metadaten-Editor aufgenommenen Metadaten werden direkt in die Datenbank des Repositoriums eingelesen und dort von Datenkuratoren geprüft. Um darüber hinaus die zusätzliche technische Beschreibung der Daten zu standardisieren und die Kunden bei deren Erstellung zu unterstützen, haben wir Datenbeschreibungstemplates entwickelt, die in kommentierter



The screenshot shows the 'Online-Metadaten-Editor' interface. It includes a top navigation bar with 'Data/Cite Metadata', 'ISO19115 Metadata', 'Files', and 'Related Publications'. The main content area is divided into several sections:

- Resource Information:** Fields for DOI (10.5880/GFZ.1.4.2016.001), Publisher (GFZ Data Services), Year (2016), Resource Type (Dataset), Title (Supplement to: The New World Atlas of Artificial Night Sky Brightness), and Language of dataset (eng).
- Licenses and Rights:** License (Attribution-NonCommercial 4.0 International) and Rights URI (https://creativecommons.org/licenses/by-nc/4.0/).
- Authors (Persons and/or Institutions):** A table listing authors with columns for Lastname, Firstname, Role, Author ID Type, Author Identif..., Affiliation, Affiliation2, and Affiliation3. Authors listed include Fakchi, Cinzano, Duriscoe, Kyba, Elvige, Beigh, Pustan, Rybnikova, and Furgon.
- Temporal and Spatial Coverage:** A table for 'Temporal and Spatial Coverage (The EDIT-symbol to the left provides vi)' with columns for Latitude (Min, Max) and Longitude (Min, Max). A map below shows a selected region in Europe.

Online-Metadaten-Editor

This document will accompany your data as a PDF description.

Descriptive Title of Dataset
(https://doi.org/will be provided)

Provide a descriptive title that addresses the content of the dataset. Titles like "Supplement to: title of the paper" are not recommended.

Author-1¹, Author-2² ...

1. Affiliation1, City, Country
2. Affiliation2, City, Country

1. Licence

Creative Commons Attribution 4.0 International License (CC BY 4.0)



2. Citation

When using the data please cite:

Derrien, Allan (2019): Rare optical DSLR camera- and unique drone footage of the extremely remote Mount Michael Volcano, Saunders Island (South Sandwich Islands). GFZ Data Services. <http://doi.org/10.5880/ridgeo.2019.020>

The data are supplementary material to: (if applicable)

Enter reference with DOI to the article or URL PhD thesis here, you may also add references to more than one article. If the article is not yet published, please add a "submitted" or simply a note "citation of article XX when available". Ideally highlighted in yellow

Table of contents (optional, recommended for more than 2 pages)

1. Licence
2. Citation
3. Data Description
- 3.1. Sampling method

Datenbeschreibungstemplate

Abb. 1: Werkzeuge für eine gute Datenbeschreibung/-dokumentation bei GFZ Data Services. Links: Ausschnitt aus dem Online-Metadaten-Editor zur Erfassung standardisierter, maschinenlesbarer Metadaten Rechts: Auszug aus dem kommentierten Datenbeschreibungstemplate zur Erfassung zusätzlicher, umfassenderer und standardisierter technischer Datenbeschreibungen

und unkommentierter Version angeboten werden (Abb. 1 rechts). Die zusätzliche Datenbeschreibung ergänzt die im Metadaten-Editor erfassten Metadaten durch umfassende technische Datenbeschreibungen. Alternativ dazu, besonders für größere Projekte, können spezielle Datenreports mit DOI veröffentlicht werden. Die Templates helfen den Forschenden bei ihrer Datenbeschreibung und reduzieren gleichzeitig den Kurationsaufwand. Darüber hinaus bieten wir auf unserer Website umfangreiches Beratungsmaterial zu verschiedenen Aspekten rund um das Publizieren von Forschungsdaten und wissenschaftlicher Software an.

Als Partnerrepositorium des FID GEO bieten wir unseren Datenpublikations- und Kurationservice der gesamten geowissenschaftlichen Gemeinschaft aus Forschung und staatlichen Diensten/Landesämtern sowie weiteren Interessierten an.

Freie Zugänglichkeit von Forschungsergebnissen bedeutet für GFZ Data Services nicht nur Daten zu publizieren, sondern alle Komponenten von Forschungsergebnissen gemeinsam zu betrachten und miteinander zu verknüpfen. Dabei spielen global eindeutige, persistente Identifikatoren (PIDs) eine wichtige Rolle, da sie Objekte eindeutig identifizieren, über einen ausführbaren Link erreichbar sind und somit in maschinenlesbare Metadaten integriert werden

können. GFZ Data Services verwendet derzeit vor allem vier PIDs:

- (1) Digital Object Identifier (DOI, <https://doi.org>) für die zu publizierenden Objekte selbst, aber auch für Zitationen von Veröffentlichungen, Daten oder Software, die mit den veröffentlichten Objekten im Zusammenhang stehen.
- (2) Die Open Researcher and Contributor ID (ORCID, <https://orcid.org>), um Autoren und andere Beitragende Menschen eindeutig zu identifizieren.
- (3) Die noch relativ neue Research Organizations Registry ID (ROR, <https://ror.org>) für die Identifikation von Institutionen und
- (4) die International Generic Sample Number (IGSN, <https://igsn.org>), einen PID für physische Proben). Letzterer wird dafür genutzt, um die publizierten Daten mit den originären Proben digital zu verbinden.

Literatur:

COUNCIL OF THE EUROPEAN COMMISSION (2016): The transition towards an Open Science system. – Council conclusions (adopted on 27/05/2016), <https://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf>, letzter Zugriff am 28. Februar 2024

ELGER, K., BERTELMANN, R., ULBRICHT, D., HABERLAND, C., LAUTERJUNG, J., & B. RADOSAVLJEVIC (2018): FAIR am GFZ. – *System Erde* **8**, S. 52–59, <https://doi.org/10.2312/gfz.syserde.08.01.8>

EUROPEAN COMMISSION. DIRECTORATE GENERAL FOR RESEARCH AND INNOVATION (2018): Turning FAIR into reality: final report and action plan from the European Commission expert group on FAIR data. – <https://doi.org/10.2777/1524>

HEIDORN, B. P. (2008): Shedding Light on the Dark Data in the Long Tail of Science. – *Library Trends* **57**, 2, pp. 280–299, <https://doi.org/10.1353/lib.0.0036>

PAMPEL, H., & K. ELGER (2021): 5.6 Publikation und Zitierung von digitalen Forschungsdaten. – *Praxishandbuch Forschungsdatenmanagement*, S. 521–536, <https://doi.org/10.1515/9783110657807-028>

SEMMLER, M., ACHTERBERG, I., LORENZ, M. & K. ELGER (2024): Publikation geowissenschaftlicher Reihen auf GEO-LEOe-docs. – Vortrag 83. Jahrestagung Arbeitsgemeinschaft Norddeutscher Geologen Bad Muskau, Brandenburg. *Geowiss. Beitr.* **31**, S. 81–83

WILKINSON, M., DUMONTIER, M., AALBERSBERG, I. et al (2016): The FAIR Guiding Principles for scientific data management and stewardship. – *Sci Data* **3**, 160018, <https://doi.org/10.1038/sdata.2016.18>

Anschriften der Autoren:

Kirsten Elger
Melanie Lorenz
Helmholtz-Zentrum Potsdam
Deutsches GeoForschungsZentrum GFZ
Wissenschaftspark Albert Einstein
Telegrafenberg
14473 Potsdam

Malte Semmler
Inke Achterberg
Niedersächsische Staats- und
Universitätsbibliothek Göttingen
Platz der Göttinger Sieben
37073 Göttingen