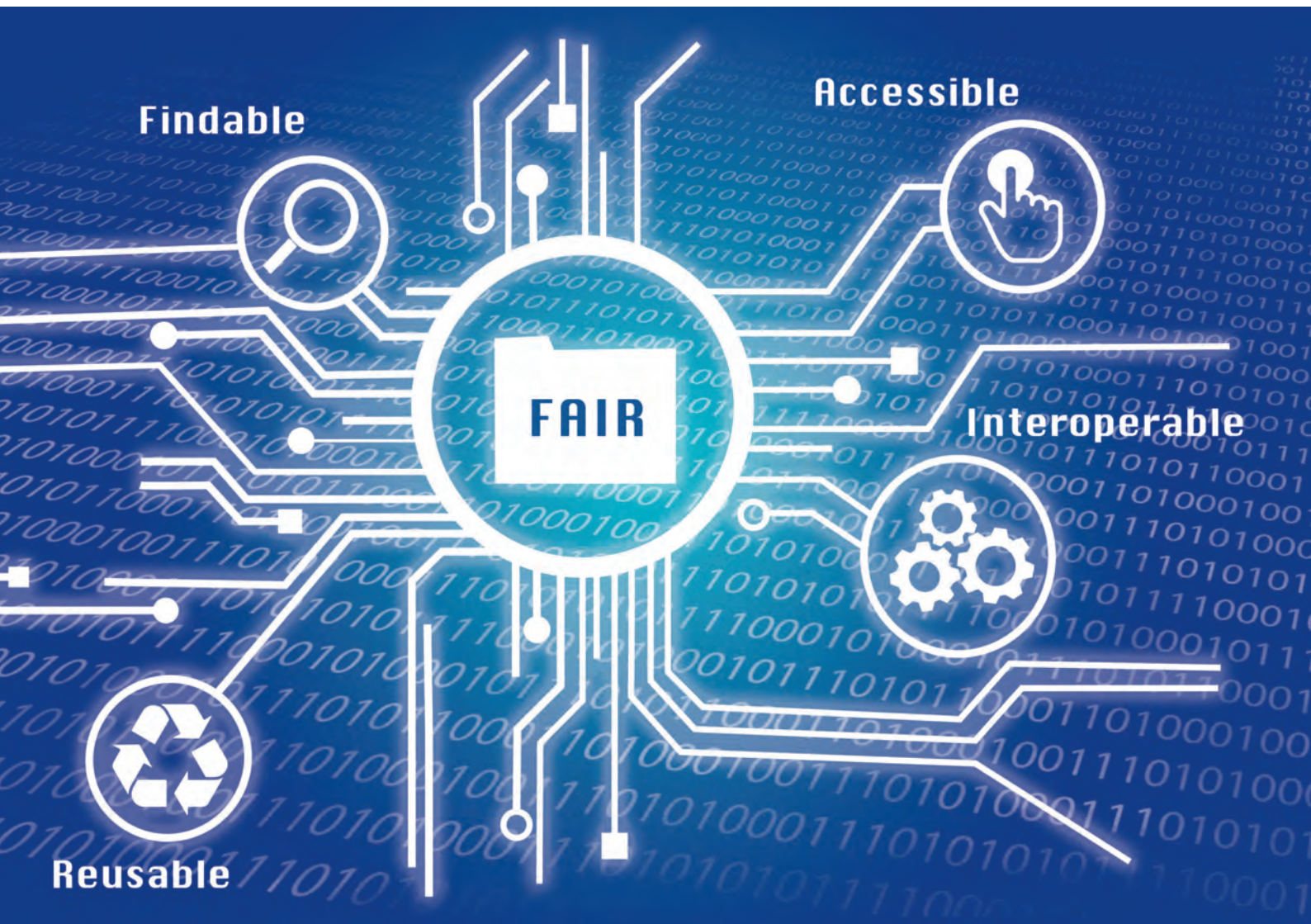


FAIR am GFZ

Kirsten Elger, Roland Bertelmann, Damian Ulbricht, Christian Haberland, Jörn Lauterjung, Boris Radosavljevic
Deutsches GeoForschungsZentrum GFZ, Potsdam

GFZ German Research Centre for Geosciences has a long tradition in research data management and practice in making data accessible to the scientific community. Since its publication in 2016, the „FAIR Guiding Principles for scientific data management and stewardship“ quickly evolved as a standard for research data management worldwide. GFZ is committed to these standards and has accordingly taken a series of actions which are described here. A key tool for FAIR data management at GFZ is the citable publication of research data with Digital Object Identifier (DOI). GFZ has a long tradition in research data management and practice in making data accessible to the scientific community. Since its publication in 2016, the „FAIR Guiding Principles for scientific data management and stewardship“ quickly evolved as a standard for research data management worldwide. GFZ is committed to these standards and has accordingly taken a series of actions which are described here. A key tool for FAIR data management at GFZ is the citable publication of research data, data collections and software with Digital Object Identifier (DOI) complementary to well-established disciplinary standards and data management.



FAIR-Prinzipien

Innerhalb weniger Jahre hat sich die griffige Bezeichnung FAIR Data international als Synonym für gutes Forschungsdatenmanagement durchgesetzt. Sowohl in der Diskussion der Praktiker beim Umgang mit Forschungsdaten, als auch im wissenschaftspolitischen Umfeld, wie etwa aktuell in den Positionsbestimmungen zur European Open Science Cloud (EOSC)¹, hat sich der Begriff FAIR Data etabliert. FAIR fügt sich zusammen aus

- Auffindbarkeit (**F**indable)
- Zugänglichkeit (**A**ccessible)
- Interoperabilität (**I**nteroperable)
- Wiederverwendbarkeit (**R**eusable)

und definiert so kompakt den angemessenen Umgang mit Forschungsdaten.

Die FAIR-Prinzipien² für das Forschungsdatenmanagement wurden erstmals 2016 über die Veröffentlichung „The FAIR Guiding Principles for scientific data management and stewardship“³ beschrieben. Zusammengefasst lässt sich sagen, dass die Prinzipien die Weitergabe und das Nachnutzen von Daten verbessern sollen, um sowohl reproduzierbare Forschungsergebnisse als auch eine Aufbereitung und ggfs. Neuinterpretation der Daten durch Menschen und Maschinen (Computer) zu ermöglichen. Die FAIR-Prinzipien können auf die Daten selbst und auf die sie beschreibenden Daten (ihre Metadaten) angewendet werden. Metadaten sind der Schlüssel für die Umsetzung der FAIR-Prinzipien. Sie dienen der Auffindbarkeit der Daten (oft über Datenportale) und sie beschreiben deren Zugänglichkeit, Interoperabilität und Nachnutzbarkeit. Nur wenn beispielsweise die vollständige Provenienz eines verfügbaren Datensatzes in den Metadaten beschrieben ist, kann später abgeschätzt werden, welche Datensätze sich für welche Weiterverarbeitungsansätze eignen.

FAIR Data ist auch ein zentrales Anliegen der Research Data Alliance (RDA). Die RDA hat den Anspruch, weltweit über Fachgrenzen hinweg nicht nur Wissen zum Forschungsdatenmanagement auszutauschen, sondern gemeinsame Lösungsansätze zum „barrierefreien“ Datenaustausch zu entwickeln und zu implementieren.

Die „European Commission Expert Group on FAIR data“ hat im Juni 2018 den EU Interim report „Turning FAIR data into reality“ veröffentlicht⁴. Der Report beschreibt das FAIR-Konzept und gibt konkrete Empfehlungen und Richtlinien, wie das Konzept umgesetzt und erreicht werden kann. Die FAIR-Prinzipien beziehen sich nicht nur auf Forschungsdaten allein, sondern beschreiben ein „Ökosystem“ aus Daten, Metadaten, wissenschaftlicher

Software, aber auch Workflows, Metriken und die Erfordernis dauerhafter Finanzierung von Infrastrukturen, die die Umsetzung der FAIR-Prinzipien erfüllen (z. B. Repositorien). Die eindeutige Identifizierung von Daten (DOI Digital Object Identifier, Handle, usw.), Personen (ORCID Open Researcher and Contributor ID) oder physischen Objekten, wie Proben (IGSN, International Geo Sample Number), aber auch Policies, Richtlinien, Standards und Datenrepositorien sind unverzichtbare Werkzeuge in einem solchen System.

Auszug aus „Turning FAIR data into reality“

Components of a FAIR data ecosystem (recommendation 4)

The realisation of FAIR data relies on, at minimum, the following essential components: policies, Data Management Plans (DMPs), identifiers, standards and repositories. There need to be registries cataloguing each component of the ecosystem and automated workflows between them.

- Registries need to be developed and implemented for all of the FAIR components and in such a way that they know of each other's existence and interact. Work should begin by enhancing existing registries for policies, standards and repositories to make these comprehensive, and initiate registries for DMPs and identifiers.

Stakeholders: Data services; Standard bodies; Global coordination fora.

- By default, the FAIR ecosystem as a whole and individual components should work for humans and for machines. Policies and DMPs should be machine-readable and actionable.

Stakeholders: Data services; Global coordination fora; Policymakers.

- The infrastructure components that are essential in specific contexts and fields, or for particular parts of research activities, should be clearly defined.

Stakeholders: Research communities; Data stewards; Global coordination fora.

- Testbeds need to be used to continually evaluate, evolve, and innovate the ecosystem.

Stakeholders: Data services; Data stewards.⁵



Kontakt: K. Elger
(kirsten.elger@gfz-potsdam.de)

Daten am GFZ: FAIR – avant la lettre

Das Deutsche GeoForschungsZentrum GFZ hat eine lange Tradition darin, Forschungsdaten öffentlich bereitzustellen. Einerseits ist die Forschung international vernetzt und Daten werden über schon etablierte externe Plattformen veröffentlicht und zugänglich gemacht. Andererseits betreibt das GFZ selbst im Rahmen der Modular Earth System Infrastructure (MESI)⁶ große Messnetze, mit denen Daten erarbeitet werden, die der wissenschaftlichen Gemeinschaft zur Verfügung stehen. Gerade hier spielt die Einbindung in internationale Strukturen eine große Rolle, um die Daten auffindbar und nach definierten Regeln in vereinbarten, fachspezifischen Formaten zugreifbar zu machen.

So werden die geomagnetischen Daten aus Observatorien des GFZ sowohl in dem vom British Geological Survey betriebenen World Data Centre for Geomagnetism archiviert, als auch über das internationale geomagnetische Konsortium INTERMAGNET bereitgestellt. Im Rahmen der International Association of Geodesy (IAG) ist das GFZ einer der führenden Partner bei vielen internationalen Services und betreibt mehrere Daten- und Analysezentren. GEOFON betreibt eines der größten Datenarchive für seismologische Daten

weltweit. All diese Daten werden nach internationalen disziplinären Standards erzeugt und vorgehalten, an deren (Weiter-)Entwicklung GFZ-Wissenschaftlerinnen und -Wissenschaftler häufig beteiligt waren und sind. Auch die Datenbank der World Stress Map⁷ zählt dazu, die als Weltzentrum globale Messungen zum aktuellen Spannungszustand der Erde sammelt und den Service anbietet, qualitätsgeprüfte Stresskarten von selbst gewählten Regionen zusammenzustellen.

Das GFZ nutzt seine international eingebundenen Forschungsinfrastrukturen auch, um den FAIRen Umgang mit Forschungsdaten über die Praxis am GFZ hinaus in Fachcommunities zu verankern. So hat beispielsweise GEOFON die persistente Identifikation von seismischen Netzwerken mit DOI (Digital Object Identifier) als Standard in die International Federation of Digital Seismic Networks (FDSN), den Dachverband der seismologischen Netzwerke, eingebracht und bereits 47⁸ permanente und temporäre seismologische Netzwerke persistent durch DOIs referenziert⁹.

Die Vielfalt und den aktuellen Stand der durch die Forschungsinfrastrukturen am GFZ bereitgestellten Forschungsdaten lässt sich am neuen Suchwerkzeug RI@GFZ¹⁰ ablesen. RI@GFZ wurde als

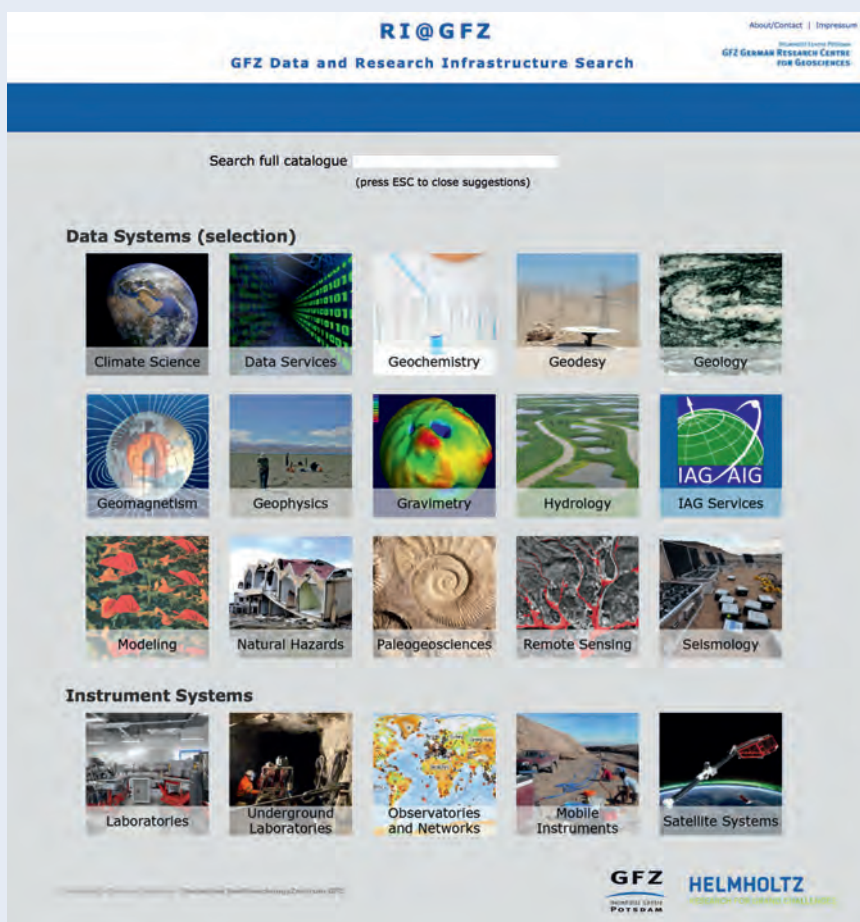


Abb. 1: RI@GFZ steht für Research Infrastructures am GFZ. Es ist ein einfaches, durchsuchbares Informationsportal zu Daten und Dateninfrastrukturen am GFZ, die öffentlich oder nach einer Registrierung zur Verfügung stehen. Das Portal ersetzt nicht die detaillierten Beschreibungen der Projekt-Webseiten, sondern bietet eine kurze inhaltliche Beschreibung und vor allem Links zu den Daten, Infrastrukturen, Projekten sowie Ansprechpersonen.

Fig. 1: RI@GFZ means Research Infrastructures at GFZ, a searchable information portal on GFZ data and data infrastructures which are publicly available with or without registration. The portal doesn't replace detailed descriptions on project webpages, but offers a short introduction and links to data, infrastructures, projects and contact persons.

einfacher Wegweiser zu den Forschungsdaten am GFZ entwickelt (Abb. 1) und bietet neben einer kurzen inhaltlichen Einführung zu den einzelnen Daten und Forschungsinfrastrukturen vor allem Links zu Daten, Websites sowie Ansprechpersonen.

Alle Daten, die in globale Netzwerke einfließen (vor allem aus Geophysik und Geodäsie) folgen internationalen fachspezifischen Standards, sowohl beim Format der Daten, als auch bei den beschreibenden Metadaten und sind somit innerhalb der Communities interoperabel und nachnutzbar. Für die interdisziplinäre Nutzung jedoch, insbesondere im Kontext der FAIR Prinzipien, sind maschinenlesbare Metadaten, vor allem für Lizenzen und Provenienz, aber auch die Vergabe von persistenten Identifiern (PID) für Daten, Personen etc. unverzichtbar. Hier bietet sich GFZ Data Services, das seit 2007 am GFZ betriebene Forschungsdatenrepositorium für geowissenschaftliche Daten und Software, als wichtiger Partner an. Forschungsdaten und Software werden hier mit DOIs, Autorinnen und Autoren mit ORCID¹¹ eindeutig identifiziert. Die assoziierten Metadaten und die Lizenzen folgen internationalen Standards und sind für Menschen und Maschinen lesbar.

Weitere Schritte zu FAIR

Das GFZ hat im März 2016 als erstes Helmholtz-Zentrum Grundsätze zum Umgang mit Forschungsdaten¹² verabschiedet und damit elementare Leitlinien gesetzt, an denen sich das Forschungsdatenmanagement am GFZ orientiert. In diesen Grundsätzen wird der explizite Wunsch geäußert, dass die Forschungsdaten, die am GFZ erzeugt und erhoben werden, so weit wie möglich öffentlich bereitgestellt werden – unter gleichzeitiger Würdigung der Leistung der Forschenden und nach angemessenen Embargozeiten. Auch hierbei ist die zitierbare Datenpublikation ein Schlüsselement. Die in den Richtlinien gesetzten Themen wissenschaftliche Anerkennung, Sicherung und Zugänglichkeit, Zitierbarkeit und Publikation, offene Lizenzen, Forschungsdateninfrastrukturen, Qualifizierung und Verantwortlichkeit sprechen bereits Umsetzungsstrategien an, die dem Geist der später veröffentlichten FAIR-Prinzipien entsprechen.

Die Grundsätze beinhalten auch die Unterstützung des GFZ für die „Joint Declaration of Data Citation Principles“¹³ von 2014 und für das „COPDESS Statement of Commitment from Earth and Space Science Publishers and Data Facilities“¹⁴. In der COPDESS-Erklärung von 2015 setzen Verlage, Datenanbieter und wissenschaftliche Gesellschaften gemeinsame Standards zur Publikation von Forschungsdaten im Kontext mit Artikeln in Zeitschriften. So wurde vereinbart, dass Forschungsdaten, die mit einer DOI veröffentlicht wurden, in den Literaturlisten wissenschaftlicher Artikel zitiert werden dürfen und dass Zeitschriften die Bereitstellung von Daten, die den Publikationen zu Grunde liegen, fordern können – eine inzwischen gängige Praxis in den Geowissenschaften. Auf dieser Basis wird aktuell unter Beteiligung des GFZ im Rahmen von der American Geophysical Union (AGU) geführten Initiative „Enabling FAIR Data in the Earth and Space Sciences“ an weiteren

Empfehlungen gearbeitet¹⁵. Hierbei spielen, wie bereits bei der COPDESS-Initiative, Datenpublikationen und ihre Verknüpfung mit wissenschaftlichen Veröffentlichungen eine wichtige Rolle auf dem Weg zu „fairen“ Daten, und es geht erneut um konkrete Verabredungen zwischen Verlagen und Repositorien. Auch auf der im Frühjahr 2018 vom GFZ als lokalem Organisator unterstützten internationalen Konferenz der Research Data Alliance¹⁶ standen die Rahmenbedingungen für die Zitierbarkeit von Forschungsdaten als elementarer Ansatz für „fares“ Datenmanagement im Mittelpunkt der Diskussion¹⁷.

Das GFZ war Projektpartner im DFG-Projekt STD-DOI (2003 bis 2008), das die Zitierbarkeit von Forschungsdaten mithilfe eines Digital Object Identifier (DOI) entwickelt und aus dem die Gründung von DataCite¹⁸, der heute größten Registrierungsagentur für Daten-DOIs hervorgegangen ist. Das auf dieser Erfahrung aufbauende fachspezifische Forschungsdaten-Repositorium für geowissenschaftliche Daten „GFZ Data Services“¹⁹ ist das zentrale Angebot für die Bereitstellung von Forschungsdaten nach FAIR-Prinzipien am GFZ (Abb. 2) und darüber hinaus. Über GFZ Data Services werden Daten, Datensammlungen, Datenprodukte und auch wissenschaftliche Software zitierbar veröffentlicht, ergänzt mit umfangreichen (standardisierten) Metadaten und bei Bedarf mit weitergehenden Beschreibungen (Data Reports, README).

Der Fokus von GFZ Data Services liegt einerseits auf dem sogenannten „long-tail“ der Forschungsdaten²⁰, den vielen kleinen und sehr unterschiedlichen Datensätzen, die von einzelnen Wissenschaftlerinnen und Wissenschaftlern oder kleineren Arbeitsgruppen erzeugt werden. Andererseits wurden in den letzten Jahren Publikationsservices für internationale Projekte und Dienste entwickelt (z. B. für die World Stress Map, die IAG Services IGETS²¹ und ICGEM²², das International Continental Scientific Drilling Programme ICDP²³, TERENO²⁴, EPOS²⁵ und andere). Das Angebot umfasst sowohl statische als auch dynamische Datentypen, mit dem gemeinsamen Nenner, dass die Forschungsdaten als Dateien angeboten werden. Da entsprechend keine direkte Suche in den Datensätzen möglich ist, wird besonderer Wert auf die beschreibenden Metadaten gelegt, um möglichst optimale und schnelle Suchergebnisse zu erreichen. Die Unterstützung internationaler Metadatenstandards und die Verschlagwortung von Datensätzen mit kontrollierten Vokabularien der Geowissenschaften (wie NASA GCMD Science Keywords, IUGS-CGI, INSPIRE) ermöglicht einerseits eine strukturierte Suche und erleichtert andererseits die Weitergabe von Metadaten an externe Portale, um die Auffindbarkeit der Daten zu erhöhen. Bei wissenschaftlicher Software wird stets auf die jeweilige Projektseite im Software-Repositorium (GitLab, GitHub) verlinkt, welches die Versionierung der Software managt.

Mehr als Daten ...

Neben der dauerhaften und eindeutigen Identifikation von Forschungsdaten wird es zunehmend relevant, weitere Objekte im Forschungsprozess und auch die agierenden Wissenschaftlerinnen und Wissenschaftler selbst eindeutig referenzieren zu

können. Dies erleichtert den Aufbau, die Verwaltung und die Zusammenführung von bestehenden Datenbanken.

Bei Akteuren im Bereich der Datenentstehung ist es beispielsweise vorteilhaft, nicht nur den Namen, sondern auch einen persistenten Identifier, wie die Open Researcher and Contributor-ID (ORCID) zu speichern – denn es gibt Namensgleichheiten, Namen werden abgekürzt, sie können sich ändern und bei Umlauten werden oft unterschiedliche Schreibweisen verwendet. An der Umsetzung einer breiten Benutzung von ORCID arbeitet das GFZ als Projektpartner im DFG-geförderten Projekt ORCID-DE aktiv mit. Das Projekt hat sich zum Ziel gesetzt, die Anwendung der ORCID des weltweiten Standards zur eindeutigen Referenzierung von Wissenschaftlerinnen und Wissenschaftlern in Deutschland zu fördern. Persistente Identifier für Proben lösen ein verwandtes Problem – sie machen die Zuordnung und Namen von Proben eindeutig. Für die Auswertung größerer Messkampagnen werden oft in verschiedenen Laboren Proben analysiert. Lange Zeit existierte kein eindeutiges Namensschema für Proben, mit der Folge, dass in Publikationen zwar Probenamen mit Messwerten genannt wurden – die Zusammenführung von Messergebnissen aber durch Namenswiederholungen sehr erschwert wurde. Um Proben in Publikationen besser verfolgen zu können und die gemessenen Daten

einzelnen Proben zuordnen zu können, wurde die International Geo Sample Number (IGSN) als persistenter Identifier entwickelt. Das GFZ gehört zu den Gründungsmitgliedern des IGSN e.V., dem Verein, der die Vergabe von IGSNs global koordiniert. Anwendungen am GFZ finden sich bei ICDP und den geochemisch arbeitenden Sektionen, für die bereits automatische Registrierungsservices implementiert wurden (Abb. 3). Die IGSN hilft dort bei der Katalogisierung und Präsentation von Probensammlungen und erlaubt es, Einzelproben in Veröffentlichungen zu referenzieren. Ende 2014 wurde, als wichtiger Schritt hin zur FAIR, die IGSN als Identifier in das Metadatenschema 4.0²⁶ von DataCite aufgenommen. Dies ermöglicht die direkte Verknüpfung von Proben mit publizierten Datensätzen und auch die Möglichkeit, IGSNs direkt in wissenschaftlichen Artikeln zu zitieren. Analog zu DOI-Links wird hier auch auf die online verfügbare Probenbeschreibung verlinkt.

Von GFZ-Beschäftigten in den Bereichen Geophysikalischer Gerätepool und GEOFON wird im Rahmen der RDA Working Group „Persistent Identification of Instruments“ aktiv an der Integration von PIDs für Geräte gearbeitet. So können die gerätespezifischen Informationen für Sensoren und Rekorder optimal und vor allem eindeutig mit den aufgezeichneten Daten verknüpft werden.

The screenshot shows the GFZ Data Services website. At the top left is the GFZ logo (Geophysikalisches Forschungszentrum Potsdam). The main heading is "GFZ Data Services". Below this is a search bar with the text "Search the Research Data Repository of GFZ Data Services below and read here how to publish data." To the left of the search bar are filters for "Search" (with a note "press ESC to close suggestions") and "Spatial Filter" (with a "Close Map" button). Below the filters is a "Datacenters" section listing various databases like CRC 1211 Database, ENMAP, FID GEO, etc. There is also a "Categories" section with checkboxes for "earth science" and "earth science services". On the right side, there is a world map showing search results. Below the map, it says "Found 6404 datasets." Two dataset entries are visible: "gms-index-mediator: a R-tree-based in-memory index for fast spatio-temporal queries for the GeoMultiSens platform" and "Data to Understanding soil and plant interaction by combining ground-based quantitative electromagnetic induction and airborne hyperspectral data". Each entry includes authors and an abstract.

Abb. 2: Der Katalog von GFZ Data Services erlaubt die Suche über die Karte, verfügt sowohl über eine facettierte als auch eine Freitextsuche. Die Katalogansicht zeigt die Titel, Autoren und die Beschreibung der Daten. Für mehr Informationen wird auf die DOI Landing Page weitergeleitet.

Fig. 2: The data catalogue of GFZ Data Services, a domain research data repository for the geosciences, allows searching via a map, faceted search and full text search. The entry shows title, authors and the data description. For more information it links to the DOI Landing Page of the respective dataset.

GFZ
Helmholtz Centre
POTSDAM

General Identifiers

Program:	ICDP
Expedition:	ICDP 5054
Type:	Core
Name:	5054_1_A_3_Z
IGSN:	ICDP5054EC4Q001 (Open)
Parent IGSN:	ICDP5054EEW1001
Release Date:	2017-3-1

Sampling Location

Latitude:	63.4063
Longitude:	13.203057
Coordinate System:	WGS84
Elevation:	415.74
Final Depth:	412.61
Location Type:	N/A
Location Name:	Åre, Jämtlands län, Sweden
Location Description:	COSC-1 is located in the vicinity of the abandoned Fröå mine
Country:	Sweden
Province:	Jämtlands län
County:	N/A
City:	Åre

Geology

Material:	Rock
Rock Classification:	N/A
From Corrected Depth:	106.26
To Corrected Depth:	109.39
Depth Reference:	meter below ground level
Geological Age:	mid-paleozoic
Geological Unit:	N/A

Methods

MSCL	yes
XRF	yes
Lithological Description	yes
Core Overview	yes
Core Section Scan	yes
Core Catcher Scan	no

Drilling

Drilling Method:	Coring>RockCorer wireline diamond coring, HQ and NQ bit size
Operator:	Lund University, Engineering Geology Larsson Drilling Consulting AB
Funding Agency:	Swedish Research Council (Vetenskapsrådet)
Total Length:	2400.1m
Comments:	N/A
Platform Type:	drill rig
Platform Name:	Atlas Copco CT20C
Platform Description:	slimhole wireline coring system
Chief Scientist:	Chris Juhlin
Start Date:	2013-9-5
End Date:	2014-8-26

Repositories

Current Repository:	Nationales Bohrkernlager BGR Berlin-Spandau
Current Repository Contact:	Jochen.Erbacher@bgr.de
Original Repository:	N/A
Original Repository Contact:	N/A

Sample Family

- 5054_1_A_1_Z
- 5054_1_A_2_Z
- 5054_1_A_3_Z
 - 5054_1_A_3_Z_1
 - 5054_1_A_3_Z_2
 - 5054_1_A_3_Z_3
 - 5054_1_A_3_Z_4

☐=Hole, ○=Core, ▭=Core-Section, ◻=Core-Sample

The Sample Family shows a sub-sampling graph. Select entries to navigate samples. Core-Samples are issued to scientists on request. The naming convention for a Core-Sample is: Expedition_Site_Hole_Core_Section_from-to(cm). Hole, Core, and Core-Section are following the same schema respectively.

Location Map

Drilling Start/End: 2013-9-5 / 2014-8-26 *
Latitude: 63.40630 ° Longitude: 13.20306 °
Åre, Jämtlands län, Sweden

Publications & Datasets

Lorenz, H., Rosberg, J.-E., Juhlin, C., Bjelm, L., Almqvist, B. S. G., Berthet, T., ... Tsang, C.-F. (2015). COSC-1 – drilling of a subduction-related allochthon in the Palaeozoic Caledonide orogen of Scandinavia. *Sci. Dril.*, 19, 1–11. doi:10.5194/sd-19-1-2015

Lorenz, Henning; Rosberg, Jan-Erik; Juhlin, Christopher; Bjelm, Leif; Almqvist, Bjorne; Berthet, Théo; Conze, Ronald; Gee, David G.; Klonowska, Iwona; Pascal, Christophe; Pedersen, Karsten; Roberts, Nick; Tsang, Chinfu; (2015): COSC-1 operational report - Operational data sets; GFZ Data Services. <http://dx.doi.org/10.1594/GFZ.SDD8.ICDP.5054.2015>

© Helmholtz-Zentrum Potsdam - Deutsches GeoForschungsZentrum GFZ

HELMHOLTZ
GEMEINSCHAFT

Abb. 3: Die Zielseite („IGSN Landing Page“) einer Probe von einer ICDP-Bohrung²⁷, für die eine International Geo Sample Number (IGSN) registriert wurde. Sie stellt eine Online-Probenbeschreibung dar, die durch verlinkte Text- und Datenpublikationen sowie den visualisierten „Stammbaum“ der Probe gekennzeichnet ist. Alle registrierten IGSNs sind auch in IGSN-Katalogen auffindbar. Diese existiert für jeden Registrierungsagenten; ein gemeinsamer internationaler Katalog befindet sich im Aufbau.

Fig. 3: Online description (IGSN Landing Page) of a sample from an ICDP-Drilling project that has been assigned with an International Geo Sample Number (IGSN). The sample description cross-links to text and data publications and offers a visualisation of the „family tree“ of a sample. All registered IGSNs are findable in IGSN catalogues that are managed by each IGSN agent. A global catalogue is in preparation.

„Faire“ Forschungsdaten erfordern zusätzlichen Aufwand im wissenschaftlichen Alltag. Über den Schwerpunkt der Aktivitäten auf den Bereich „Zugänglichkeit“ im Lebenszyklus von Forschungsdaten (vgl. Abb. 4) ergeben sich positive Anreize für die Wissenschaft (Zitierbarkeit, Sichtbarkeit der Forschungsdaten). Angebotene Hilfsmittel, wie Datenreports und Templates für Datenbeschreibungen, erleichtern diesen Aufwand und verbessern die Dokumentation der Daten.

Dadurch lässt sich die Diskussion zu Standards beim Umgang mit Forschungsdaten unter den Prinzipien von FAIR auf den gesamten Lebenszyklus ausdehnen. Metadaten, die nicht schon

beim Erstellen und Generieren von Forschungsdaten mitgedacht werden, stellen bei der Zugänglichmachung ein großes Hindernis dar und erzeugen zusätzlichen Aufwand.

Eine besondere Herausforderung liegt in der Rolle von wissenschaftlicher Software im Prozess der Erstellung und Bearbeitung von Forschungsdaten. Angebote wie eine „Software Carpentry“ helfen auch in diesem Feld am GFZ Standards besser zu etablieren.

Ausblick

Der GFZ-Bereich „Daten-, Informations- und IT-Dienste“ bündelt eine Reihe von Serviceangeboten zur Unterstützung von Wissenschaftlerinnen und Wissenschaftlern beim Arbeiten mit Forschungsdaten. Die Breite der Kompetenzen, die von den Einheiten „Bibliothek und Informationsdienste“, „IT-Dienste und IT-Betrieb“ und „eScience-Centre“ als Services angeboten werden, wird ergänzt und verknüpft mit dem Know-how in den wissenschaftlichen Arbeitsgruppen der Sektionen in den GFZ-Departments.

Diese Kompetenzen fließen in die Zusammenarbeit mit den anderen Helmholtz-Zentren des Forschungsbereichs Erde und Umwelt ein. Das GFZ verantwortet beispielsweise das gemeinsame Datenmanagement in MOSES²⁸ und ist an dem Projekt Digital Earth beteiligt. Ebenso ist die Erfahrung des GFZ in Projekten des Helmholtz-Inkubators „Information und Data Science“²⁹ gefragt. Auch in das europäische Projekt European Plate Observing System (EPOS) fließt aktuell die Expertise des GFZ aus der Arbeit an „fairen“ Daten ein.

Auf internationaler Ebene arbeiten GFZ-Forschende weiter in Verbänden, wie der International Association of Geodesy (IAG) und International Federation of Digital Seismograph Networks (FDSN), an der Gestaltung von Standards für „faire“ Daten mit. Herausforderungen sind dabei z. B. Vereinbarungen zur Bildung von referenzierbaren Objekten mit DOIs. Diese sollten vorhandene Arbeitsflüsse nicht verändern und gleichzeitig einen Mehrwert bieten, wie die Aufnahme in Referenzlisten von Artikeln.

Parallel dazu arbeitet das GFZ aktuell zusammen mit Verlagen und anderen Datenanbietern im Rahmen der Initiative der American Geophysical Union (AGU) „Enabling FAIR Data in the Earth and Space Sciences“ an Empfehlungen zur Verknüpfung von Daten und Artikeln mit.



Abb. 4: Der Data Life Cycle beschreibt den Lebenszyklus von Forschungsdaten. Nach Erhebung und Prozessierung folgt die Analyse der Daten. Spätestens mit der Veröffentlichung der Ergebnisse in wissenschaftlichen Artikeln müssen die Daten dauerhaft gespeichert (preserving data) und für andere zugänglich gemacht werden, um eine Nachnutzung zu ermöglichen. Besonders in den Bereichen der dauerhaften Archivierung und der Zugänglichkeit zu Forschungsdaten sind Datenpublikationen durch GFZ Data Services ein wichtiges Werkzeug, um die Daten im Sinne der FAIR-Prinzipien bereitzustellen.

Fig. 4: The data life cycle describes the stages involved in successful management and preservation of data for use and reuse. Data creation and processing is followed by data analysis and interpretation. To make data reusable it must be preserved permanently and made accessible at the latest when the results are published in a scientific article. GFZ Data Services are an important tool for FAIR research data especially with respect to permanent archiving and accessibility of research data.

Weiterführende Hinweise:

- ¹ <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>
- ² <https://www.force11.org/group/fairgroup/fairprinciples>
(dt. Übersetzung: <https://blogs.tib.eu/wp/tib/wp-content/uploads/sites/3/2017/09/Die-FAIR-Data-Prinzipien.pdf>)
- ³ Wilkinson et al. (2016), The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* 3, <http://doi.org/10.1038/sdata.2016.18>
- ⁴ Hodson, Jones et al. (2018), Turning FAIR data in to reality. Interim report of the European Commission Expert Group on FAIR data. <https://doi.org/10.5281/zenodo.1285272>
- ⁵ Hodson et al. (2018), S. 67
- ⁶ <https://www.gfz-potsdam.de/wissenschaftliche-infrastruktur/forschungsinfrastruktur/>
- ⁷ <http://www.world-stress-map.org/>
- ⁸ Stand: 14. Juli 2018
- ⁹ <http://dataservices.gfz-potsdam.de/portal/?fq=datacentre+facet:%22DOI+SEISNET%20-%20GEOFON%20Seismic%20Networks%22>
- ¹⁰ <http://dataservices.gfz-potsdam.de/mesi/>
- ¹¹ Open Researcher and Contributor-ID, <https://orcid.org/>
- ¹² http://media.gfz-potsdam.de/gfz/wv/doc/16/GFZ_Daten_Grundsaeetze+Erg_de.pdf
- ¹³ <https://www.force11.org/datacitationprinciples>
- ¹⁴ <http://www.copdess.org/statement-of-commitment/>
- ¹⁵ <http://www.copdess.org/home/enabling-fair-data-project/>
<https://www.gfz-potsdam.de/en/media-and-communication/news/all/article/make-data-fair-research-data-alliance-pre-meeting/>
- ¹⁶ <https://www.rd-alliance.org/plenaries/rda-eleventh-plenary-meeting-berlin-germany>
- ¹⁷ <https://www.gfz-potsdam.de/medien-kommunikation/meldungen/detailansicht/article/von-den-geowissenschaften-lernen-11-research-data-alliance-konferenz-zum-umgang-mit-forschungsdat/>
- ¹⁸ <http://www.datacite.org>
- ¹⁹ <http://dataservices.gfz-potsdam.de>
- ²⁰ http://www.forschungsdaten.org/index.php/Long_tail_of_research_data
- ²¹ <http://igets.gfz-potsdam.de>
- ²² <http://icgem.gfz-potsdam.de>
- ²³ <http://www.icdp-online.org>
- ²⁴ <http://www.tereno.net>
- ²⁵ EPOS European Plate Observing System, <https://www.epos-ip.org/>
- ²⁶ <https://schema.datacite.org/meta/kernel-4.0/>
- ²⁷ <http://igs.org/ICDP5054EHW1001>
- ²⁸ <http://moses.eskp.de/>
- ²⁹ https://www.helmholtz.de/forschung/information_data_science/