Data archiving, sharing and publishing

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

GFZ Data Services

PANGAEA.
Data

zalf
Open Research Data

Kirsten Elger
GFZ Data Services,
GFZ German Research Centre for Geosciences, Potsdam

GFZ Helmholtz-Zentrum Potsdam
CC BY

kelger@gfz-potsdam.de

HELMHOLTZ GEMEINSCHAFT

Things to keep in mind when sharing data

A Painful (but True-to-life) Look at Data Availability and Reuse

https://scholarlykitchen.sspnet.org/2016/11/11/a-painful-but-true-to-life-look-at-data-availability-and-reuse/

Open Research Data – an international request

G8 UK
UNITED KINGDOM 2013

G8 OPEN DATA CHARTER

H2020 Programme

Guidelines on

FAIR Data Management in Horizon 2020

Government of Canada | Gouvernement du Canada

Jobs | Immigration | Travel | Business | Benefits | He

Home → Open Government → About Open Government → G8 Open Data Charter – Canada's Action Plan

G8 Open Data Charter – Canada's Action Plan

National Science Foundation
WHERE DISCOVERIES BEGIN

HOME | Research Areas | Funding | Awards

National Science Foundation (NSF)

About

Home

Open Data at NSF

G7 Science and Technology Ministers' Meeting in Tsukuba, Ibaraki

DFG Deutsche Forschungsgemeinschaft Handling of Research Data

Tsukuba Communiqué

G7 Science and Technology Ministers' Meeting in Tsukuba, Ibaraki

15-17 May 2016

……and many more

Why sharing data?

Sharing research data…

- encourages **scientific enquiry and** debate
- promotes innovation and **potential new data uses**
- leads to **new collaborations** between data users and data creators
- maximises **transparency** and accountability
- enables **scrutiny** of research findings
- encourages the improvement and **validation** of research methods
- **reduces the cost** of duplicating data collection
- increases the **impact** and **visibility** of research
- provides **credit** to the researcher as a research output in its own right
- provides great resources for education and training

(source: UK Data Archive, http://www.data-archive.ac.uk/create-manage/planning-for-sharing/why-share-data

OPEN ACCESS Freely available online

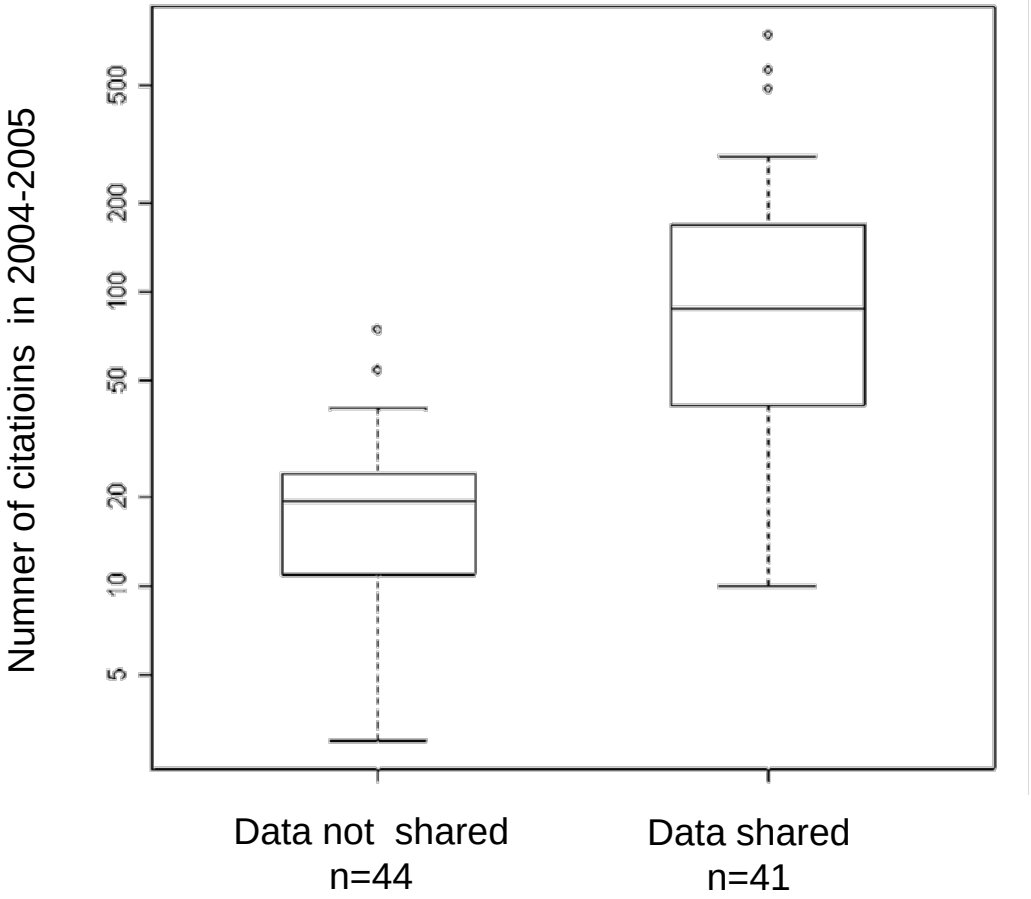
PLoS one

Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar*, Roger S. Day, Douglas B. Fridsma

Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America

"We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 48% of trials with publicly available microarray data received 85% of the aggregate citations. ***Publicly available data was*** *significantly (p = 0.006)* ***associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin*** *using linear regression."*



doi:10.1371/journal.pone.0000308

Best Practice: Data Publication

Publication of datasets as individual publications (with assigned persistent Identifier; DOI) **through data repositories**

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Data Repositories:

- permanent archives for research data
- Open Access
- disciplinary, institutional, general
- persistent identifier (ideally DOI)
- re3data.org helps to find repositories

Best Practice: Data Publication

Publication of datasets as individual publications (with assigned persistent Identifier; DOI) **through data repositories**

- **Findable:** integration of standardised metadata in external data portals (e.g. DataCite, EUDAT)
- **Accessible:** persistent data storage and access guaranteed by the publisher (= data repository)
- **Documented:** with metadata for discovery and reuse
- **Citable**: DOI-referenced datasets are citable just as journal articles (→ credit for the researcher)

Coalition on Publishing Data in the Earth and Space Sciences

COPDESS

GOAL

OPEN DATA in the EARTH and SPACE SCIENCES

SITUATION TODAY

1. Scholarly publication is a key high value entry point in making data available, open, discoverable, and usable.
2. Unfortunately, the vast majority of data submitted along with publications are in formats and forms of storage that makes discovery and reuse difficult or impossible.

STATEMENT OF COMMITMENT

42 SIGNATURES (October 2016)

COPDESS

www.copdess.org/statementofcommitment

Coalition on Publishing Data in the Earth and Space Sciences

COPDESS

GOAL

OPEN DATA in the EARTH and SPACE SCIENCES

SITUATION TODAY

1. Scholarly publication is a key high value entry point in making data available, open, discoverable, and usable.
2. Unfortunately, the vast majority of data submitted along with publications are in formats and forms of storage that makes discovery and reuse difficult or impossible.

STATEMENT OF COMMITMENT

- To promote **metadata information and domain standards**, [...], to help **simplify and standardize deposition and reuse.**
- To promote referencing of data sets using the **Joint Declaration of Data Citation Principles,** in which **citations of data sets should be included within reference lists.**
- To include in research papers concise statements indicating where data reside and **clarifying availability.**
- To promote and **implement links to data sets in publications** and corresponding links to journals in data facilities via persistent identifiers.

(January 2015)

New Journal Policies 2016

SPRINGER NATURE

RESEARCH DATA POLICY:

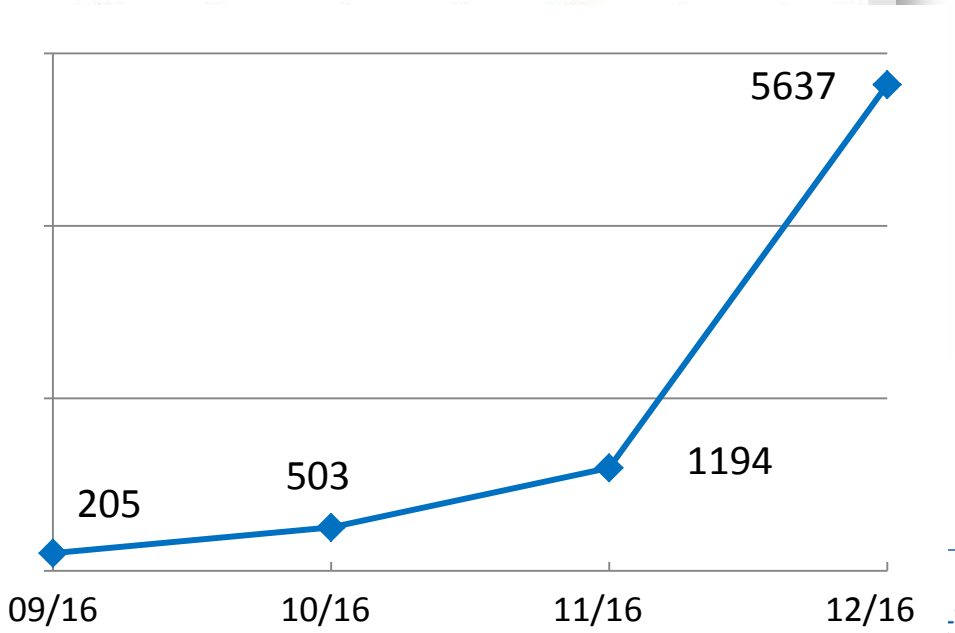
"The journal encourages authors, where possible and applicable, to deposit data that support the findings of their research in a public repository […] Datasets that are assigned digital object identifiers (DOIs) by a data repository may be cited in the reference list."

Copernicus Publications recommends depositing data that correspond to journal articles in reliable (public) data repositories, assigning digital object identifiers, and properly citing data sets as individual contributions.

Tracking Data Publications

DataCite Statistics

Datacentre	DOI Registrations				Metadata			
	Total	This Year	Last 30 Days	Last 7 Days	Searchable	Hidden	Missing	Ratio
TIB.ADLNET - Romania ADL Association	1 193	0	0	0	998	195	0	100%
TIB.AIP - Leibniz-Institut für Astrophysik Potsdam (AIP)	85 288	0	0	0	85 288	0	0	100%
TIB.AMA - AMA Service GmbH	1 893	0	0	0	1 889	4	0	100%
TIB.ASTONE - Institut für Wirtschaftsinformatik, Westfälische Wilhelms-Universität Münster	213	0	0	0	213	0	0	100%
TIB.AWI - Alfred-Wegener-Institut								
TIB.BAFG - Bundesanstalt für Gewässerkunde								
TIB.BAUA-DOI - Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)								
TIB.BAW - Bundesanstalt für Wasserbau								
TIB.BEILST - Beilstein-Institut zur Förderung der Chemischen Wissenschaften								
TIB.BIKF - Biodiversity and Climate Research Centre								
TIB.CASIMIR - Stiftung Schloss Friedenstein, Gotha								
TIB.CTT - Cellular Therapy and Transplantation								
TIB.DAGST - Dagstuhl								
TIB.DESY - DESY - Deutsches Elektronen-Synchrotron								



DOI hits for GFZ Datasets of the World Stress Map

What do I need for a data publication/ What is important when I want to share my data?

1. Data
2. Metadata

Metadata and Metadata

1. Structural metadata (disciplinary data description)

Header of sensor data

```
ENVI
description = {
  Create New File Result [Tue Jan 05 17:34:54 2016]}
samples = 70
lines   = 393
bands   = 1
header offset = 0
file type = ENVI Classification
data type = 1
interleave = bsq
sensor type = Unknown
classes = 62
class lookup = {
   0,   0,   0,  20,  75,   0,  40, 105,  10, 113, 160,
 188, 185, 115, 220, 185, 255, 205,  25, 255, 165,  20,
 225,  25,   0, 225,  25,   0, 225,  25,   0, 250, 150,
 250, 150,   0, 250, 150,   0, 230, 120,   0, 230, 120,
  25,  85, 245,  40, 145, 255,  40, 145, 255,  40, 145, 2
  40, 145, 255,  40, 145, 255, 130, 210, 255, 130, 210, 2
 250, 160, 185, 255, 195, 195, 255, 195, 195, 145,  25,
 218, 112, 214, 218, 112, 214, 218, 112, 214,  30,  30, 18
 255, 210,   0, 160, 215,  50, 160, 215,  50, 160, 215,  5
 176, 131, 255, 100,  40, 180, 255, 255,  75, 255, 255,  7
 225, 205, 170, 225, 205, 170, 147, 115,  48, 147, 115,  48
 147, 115,  48,  80,   0, 115,  80,   0, 115, 180, 180, 186
 180, 180, 180,  60,  60,  60}
class names = {
 Not_Classified, calcite_abundant, calcite, calcite.7+muscov
 calcite.8+montmorillonite_Ca.2, calcite.8+montmorillonite_N
 carbonate_Fe_bearing, dolomite, dolomite.5+montmorillonite_
 dolomite.25+calcite.25+mont_Na.5, epidote, chlorite_lowFe,
 chlorite+muscovite, muscovite_lowAl, muscovite_medAl, muscov
 muscovite_Fe-rich, illite, illite_gds4, kaolinite_wxl, kaolinite_pxl,
```

ReadMe

ds_000011593_mn_ReadMe - Editor

Datei Bearbeiten Format Ansicht ?

AVERTISSEMENT / WARNING

Même si des efforts sont déployés pour assurer la qualité des données et des métadonnées, nous ne pouvons garan

Although efforts are made to ensure the quality of the data and metadata, we cannot guarantee that they contain

AUTEUR(S) / AUTHOR(S)

Centre d'études nordiques

RÉSUMÉ / SUMMARY

Les données de ce numéro de Nordica

The datasets in this issue of Nord

CITATION DES DONNÉES / DATA CITAT

CEN 2014. Données environnementa

CEN 2014. Environmental data fr

SITE(S)

Nom / Name

Ellesmere Parks Canada (ELL

DESCRIPTION

SITE: Ellesmere Parks Ca

Profondeurs des / tempéra

MESURE(S) / MEASUREMENT(

TYPE: Température moye

d'Ellesmere au Nunavut, Canada. On

ere Island, Nunav

Definition of data labels

Metadata of the Data Tables

Sites

	Column Name	Data Type	Description	Validation Text	Unit
1	EXPEDITION	Numeric	expedition number	integer value	#
2	SITE	Numeric	site number	integer value	#
3	NAME	Text	site name or locality	text string of max. 40 characters	#
4	PLATFORM	Text	platform identifier, C=Chikyu, J=Joides, M=Mission Specific, R=Drill Rig	text string of max. 1 character	#
5	LATITUDE_DEG	Integer	decimal degrees of site latitude (latitude of hole 'A')	integer value between 0 and 90	deg.
6	LATITUDE_MIN	Double	decimal minutes of site latitude (latitude of hole 'A')	real value	min.
7	LATITUDE_DIR	Text	direction latitude	text string of max. 1 character	#
8	LONGITUDE_DEG	Integer	decimal degrees of site longitude (longitude of hole 'A')	integer value between 0 and 180	deg.
9	LONGITUDE_MIN	Double	decimal minutes of site longitude	real value	min.
10	LONGITUDE_DIR	Text	direction site longitude	text string of max. 1 character	#
11	DATE_START	Date	date of site start	date in UTC	m
12	DATE_END	Date	date of site end	date in UTC	dd-mmm-yyyy
					dd-mmm-yyyy

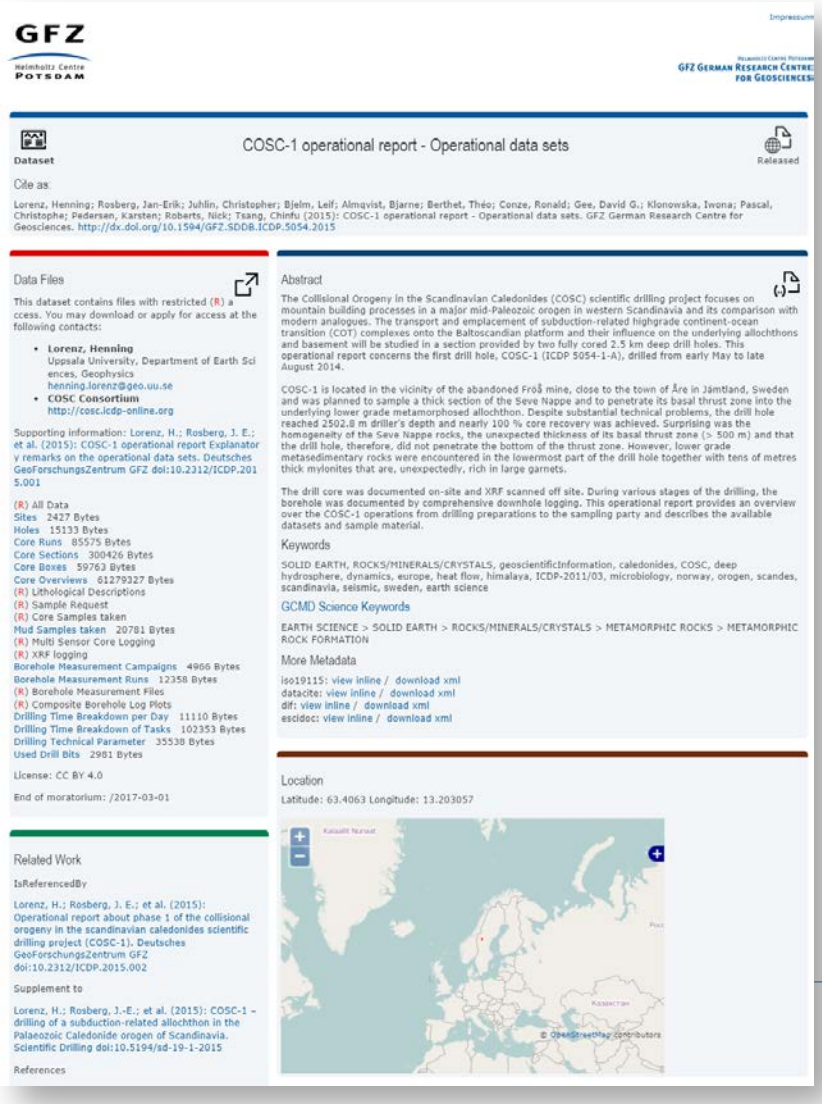
Metadata and Metadata

2. Metadata for data discovery: example DOI landing page

title
citation

download
data files

related
work



description/
abstract

Keywords

spatial
coverage

Who did **what**, **when** and **where**?

Metadata and Metadata

2. Metadata for data discovery: machine-readable form

title
citation

download
data files

related
work

Standardised metadata:
machine to machine communication

XML
(Extensible Markup Language)
Metadata exchange format

What do I need for a data publication?

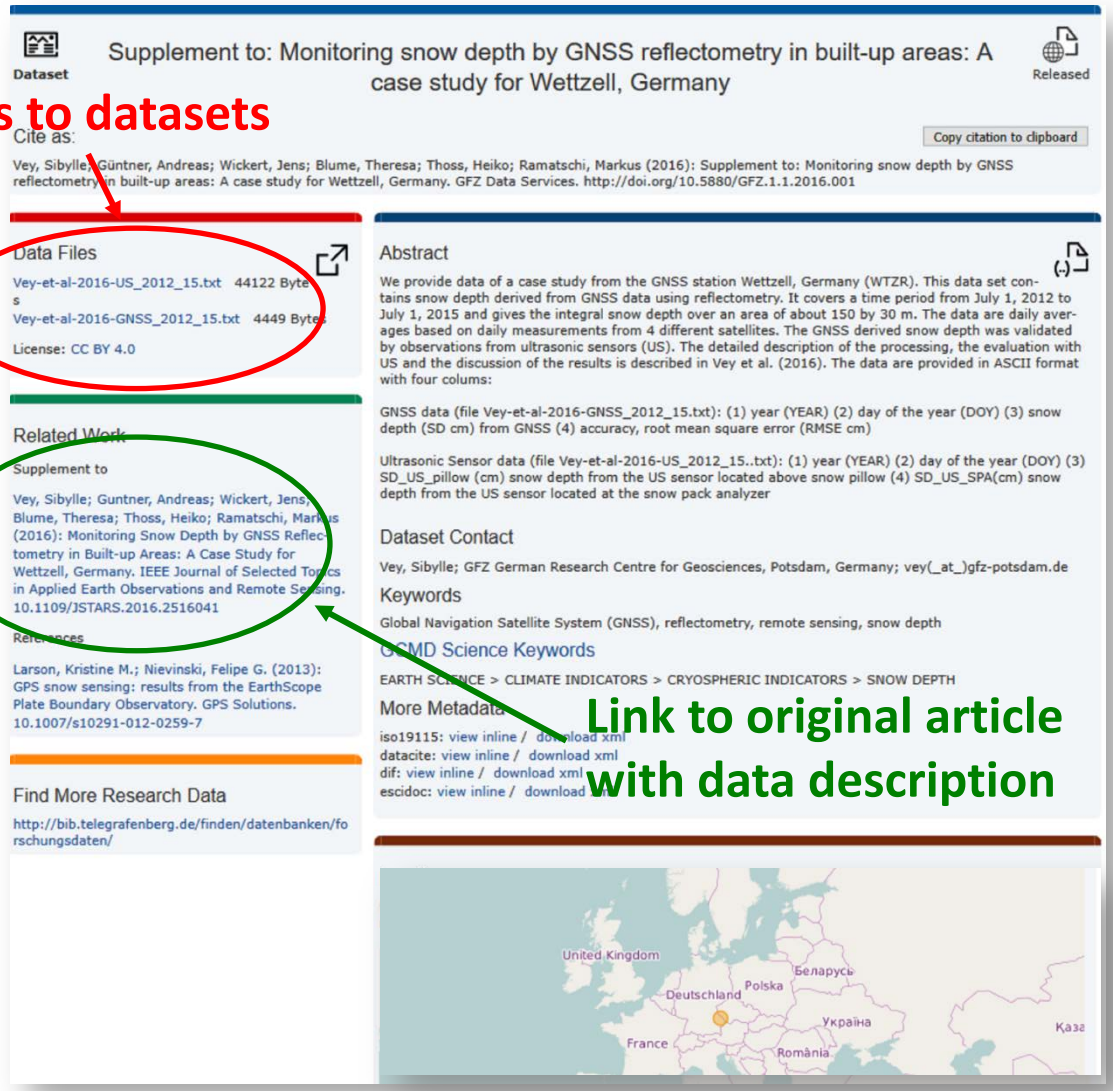
1. Research data
2. Structural/ contextual metadata for data documentation and re-use
3. Metadata for data discovery (standardised, readable for for humans and for machines)

→ **Digital object identifier (DOI)**

Formats for data publication (and their description)

1. Data publication as „supplementary material“ to journal articles (data description in the article, additional README or explanatory file with the dataset if required)
2. Data publication together with an article in a Data Journal
3. Standalone data publication with Data Report or “README”

Exampe 1: Data Supplements



We recommend…

- to publish **data supplements in open access data repositories**
- synchronous to the publication of the scientific article with **cross-references between the article and the dataset**

Example 2: Data Journals

Peer-reviewed articles with the description of datasets, data collections, data infrastructures, etc.

Example 3: GFZ Data Reports

2011: first Data Report published as a new series of the traditional Scientific Technical Report series of GFZ (persistently online accessible and citable with DOI)

GFZ Data Reports:

- Flexible format – “enhanced data description“
- standardised templates for each discipline
- internal review by domain experts
- Project-specific design if required

Citing a dataset

“A data citation in a publication should resemble a bibliographic citation and be located in the publication's reference list.” (COPERNICUS Data Policy)

Contents lists available at ScienceDirect

Tectonophysics

journal homepage: www.elsevier.com/locate/tecto

Properties of granular analogue model materials: A community wide survey

M. Klinkmüller [a], G. Schreurs [a,1], M. Rosenau [b], H. Kemnitz [b]

[a] Institute of Geological Sciences, University of Bern, Baltzerstrasse 1 +3, CH-3012 Bern, Switzerland
[b] Helmholtz-Zentrum Potsdam, GFZ Deutsches GeoForschungsZentrum, Telegrafenberg, D-14473 Potsdam, German

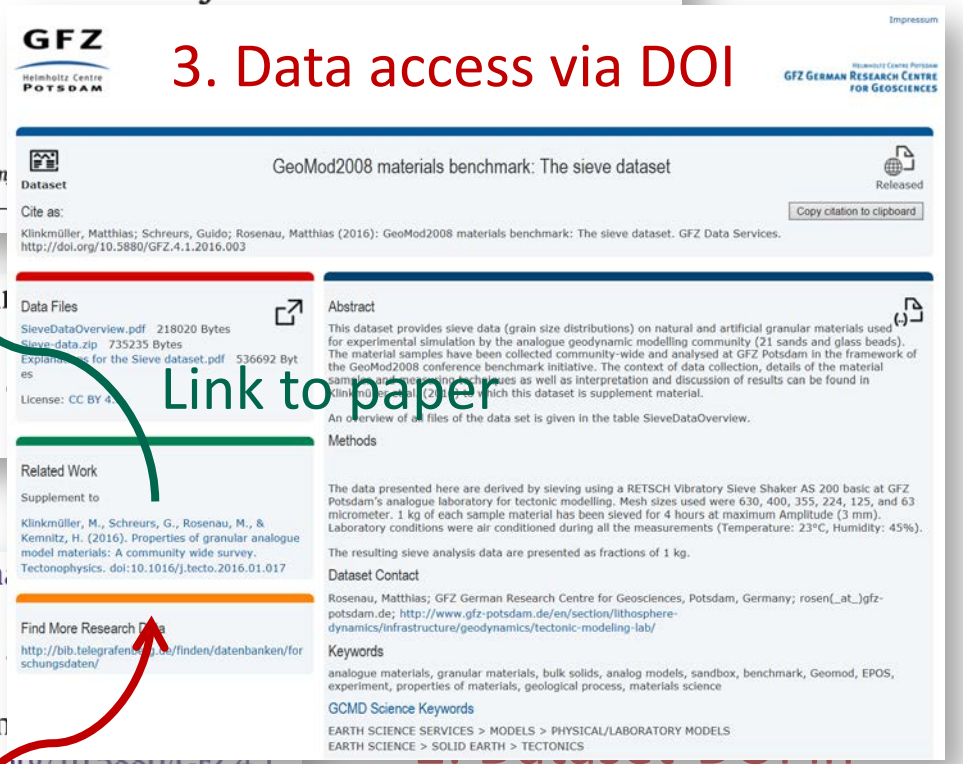
sented as grain size distribution curves, in which particle grai
plotted against cumulative weight percentage (Fig. 2).

The original sieve data have been published open access
available in Klinkmüller et al. (2016b).

References

Heilbronner, R., Keulen, N., 2006. Grain size and grain shape an
Tectonophysics 427, 199–216.
Hubbert, M.K., 1951. Mechanical basis for certain familiar geologic
Am. Bull. 62, 1259–1273.
Klinkmüller, M., Schreurs, G., Rosenau, M., 2016a. GeoMod2008 m
The ring shear test data set. GFZ Data Services. http://dx.doi.org/10.5880/GFZ.4.1.
2016.002.
Klinkmüller, M., Schreurs, G., Rosenau, M., 2016b. GeoMod2008 materials benchmark:
The sieve data set. GFZ Data Services. http://dx.doi.org/10.5880/GFZ.4.1.2016.003.
Klinkmüller, M., Kemnitz, H., Schreurs, G., Rosenau, M., 2016c. GeoMod2008 materials
benchmark: The SEM image data set. GFZ Data Services. http://dx.doi.org/10.5880/
GFZ.4.1.2016.004.

3. Data access via DOI



Link to paper

the References

GFZ Helmholtz Centre Potsdam

GFZ Data Services

About/Contact | Impressum

GFZ German Research Centre for Geosciences

Search the Research Data Repository of GFZ Data Services below and read here how to publish data.

Search
(press ESC to close suggestions)

elge

- Andersen, Per Helge (1) - author
- Elgered, Gunnar (1) - author
- Elger, Kirsten (10) - contributor

46.07323062

Current Selection (Link)

geo:[46.07323062540835,0.87890625 TO 58...

Datacenters

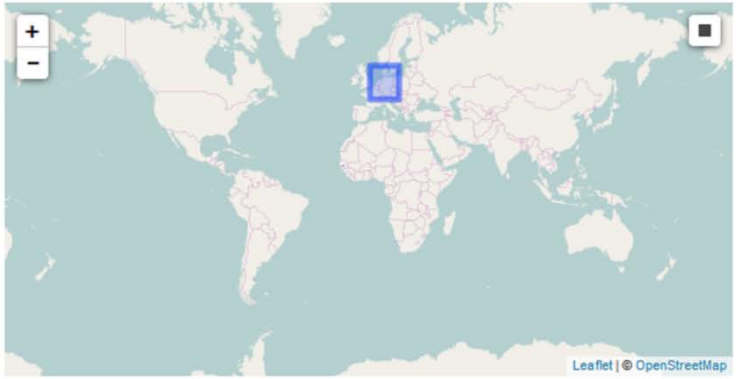
- EnMAP
- GEOFON Seismic Networks
- GFZ German Research Centre for Geosciences
- GIPP Geophysical Instrument Pool Potsdam
- ICGEM International Centre for Global Earth M...
- IGETS International Geodynamics and Earth Ti...
- PIK Potsdam Institute for Climate Impact Rese...
- SDDB Scientific Drilling Database
- SFB806 and CRC806-Database
- TERENO
- WDS World Stress Map

Top Categories

- agriculture
- atmosphere
- biosphere
- climate indicators
- data analysis and visualization
- data management/data handling
- human dimensions
- land surface
- models
- paleoclimate
- solid earth
- spectral/engineering
- terrestrial hydrosphere

Top Sub-Categories

- agricultural plant science
- atmospheric phenomena
- atmospheric pressure
- atmospheric radiation
- atmospheric temperature
- atmospheric water vapor
- atmospheric winds
- atmospheric/ocean indicators
- data search and retrieval
- deep drilling
- erosion/sedimentation
- geochemistry



Found 275 datasets.

National Network of Seismic Stations of Slovakia

Authors: ESI SAS (Earth Science Institute of the Slovak Academy of Sciences)
Abstract: The Earth Sciences Institute of the Slovak Academy of Sciences (ESI SAS) operates the National Network of Seismic Stations (NNSS) and analyzes instrumental and macroseismic data from earthquakes. The main purpose of the instrumental network is the continuous monitoring of the seismicity of Slovakia more

Superconducting Gravimeter Data from Potsdam - Level 1

Authors: Neumeyer, Jürgen; Dittfeld, Hans-Jürgen; Pflug, Hartmut et al.
Abstract: An International Geodynamics and Earth Tide Service (IGETS) was established in 2015 by the International Association of Geodesy IAG. IGETS continues the activities of the Global Geodynamics Project (GGP) between 1997 and 2015 to provide support to geodetic and geophysical research activities using more

Stress Map of the Mediterranean and Central Europe 2016

Authors: Heidbach, Oliver; Custodio, Susana; Kingdon, Andrew et al.
Abstract: The Stress Map of the Mediterranean and Central Europe 2016 displays 5011 A-C quality stress data records of the upper 40 km of the Earth's crust from the WSM database release 2016 (Heidbach et al, 2016, http://doi.org/10.5880/WSM.2016.001). Focal mechanism solutions determined as being potentially more

Metadata Catalogue

- spatial search via map
- filter + facetted search
- basic information (title, authors, abstract
- link to the DOI landing page

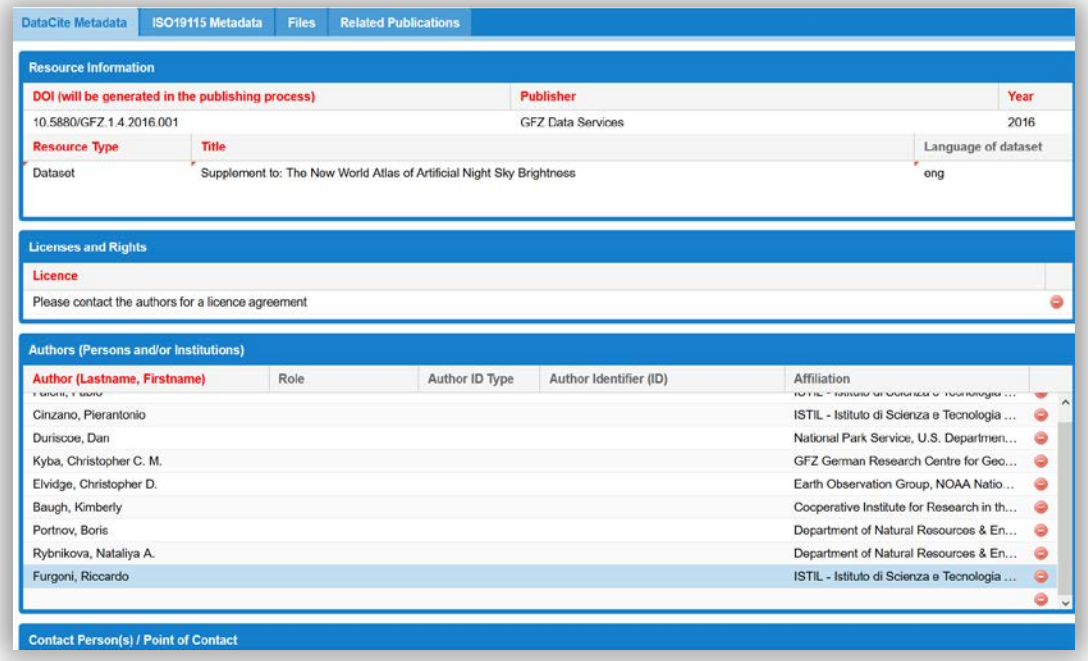
http://dataservices.gfz-potsdam.de

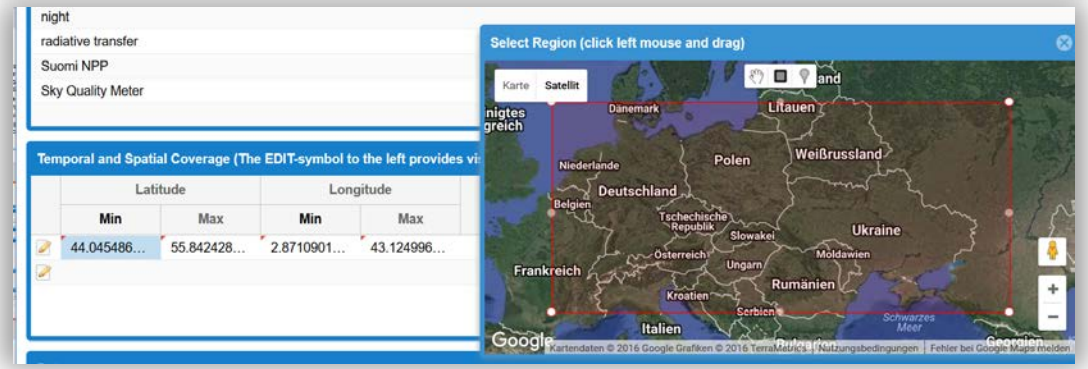
Project-specific DOI Landing Pages/ Datacentres

Datacenters

- EnMAP
- GEOFON Seismic Events
- GEOFON Seismic Networks
- GFZ German Research Centre for Geosciences
- GIPP Geophysical Instrument Pool Potsdam
- ICGEM International Centre for Global Earth M...
- IGETS International Geodynamics and Earth Ti...
- ISDC Information System and Data Center
- PIK Potsdam Institute for Climate Impact Rese...
- SDDB Scientific Drilling Database
- SFB806 and CRC806-Database
- TERENO
- TR32DB Transregio 32 Database
- WDS World Stress Map

XML Metadata Editor





Standardised XML files (Datacite, ISO 19115, NASA GCMD DIF, Dublin Core)

DataCite
FIND, ACCESS, AND REUSE DATA

GFZ Data Services Metadata Catalogue

How do I find an appropriate data repository?

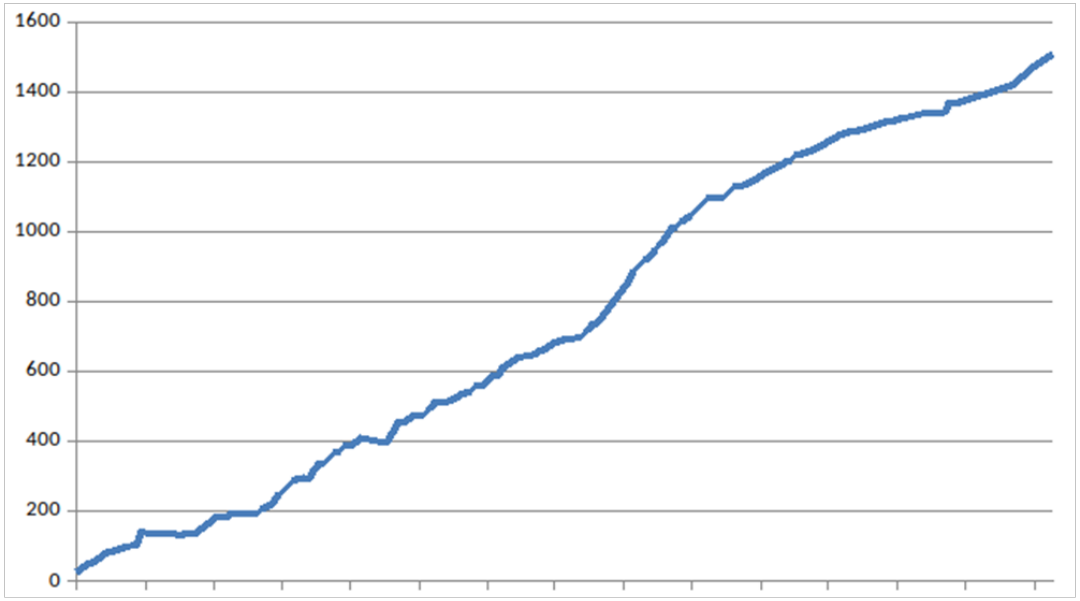
re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

- **global registry** of research data repositories
- covers **all academic disciplines**
- presents repositories and portals for the **permanent storage** and **access** of research data sets to researchers, funding bodies, publishers and scholarly institutions.
- promotes a culture of **sharing, increased access and better visibility** of research data

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

1775 repositories indexed (06.12.2016)



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

re3data.org - metadata schema

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Schema for the Description of Research Data Repositories

Version 2.2
December 2014
doi: http://doi.org/10.2312/re3.006

Authors: Paul Vierkant [b], Shaked Spier [b], Jessika Rücknagel , [b] Heinz Pampel [a], Florian Fritze [b], Jens Gundlach [c], David Fichtmüller [d], Maxi Kindling [b], Agnes Kirchhoff [d], Hans-Jürgen Goebelbecker [c], Jens Klump [f], Gabriele Kloska [c], Evelyn Reuter [c], Angelika Semrau [c], Edeltraud Schnepf [c], Michael Skarupianski [c], Roland Bertelmann [a], Peter Schirmbacher [b], Frank Scholze [c], Claudia Kramer [c], Michael Witt [e], Claudio Fuchs [a], Robert Ulrich [c]

[a] GFZ German Research Centre for Geosciences, Library and Information Services (LIS), Germany
[b] Humboldt-Universität zu Berlin, Berlin School of Library and Information Science (BLIS), Germany
[c] Karlsruhe Institute of Technology (KIT), KIT Library, Germany
[d] Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Germany
[e] Purdue University Libraries, West Lafayette, USA
[f] Commonwealth Scientific and Industrial Research Organisation (CSIRO), Mineral Resources Flagship, Kensington WA, Australia

Contact
info@re3data.org
http://www.re3data.org

Except where otherwise noted, this work is licensed under http://creativecommons.org/licenses/by/3.0/

Version 3.0

41 Properties on

- General information
- Responsibilities
- Policies
- Legal aspects
- Technical standards
- Quality standards

GFZ Helmholtz Centre Potsdam

Vierkant, P., et al. (2015). Schema for the Description of Research Data Repositories. Version 3.0. http://doi.org/10.2312/re3.008

HELMHOLTZ ASSOCIATION

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

Icons – facilitating the selection process of appropriate research data repositories

The research data repository provides additional information on its service.

The research data repository is either certified or supports a repository standard.

The research data repository provides a policy.

The research data repository provides open/restricted/closed access to its data.

The research data repository uses a persistent identifier system to make its provided data persistent, unique and citable.

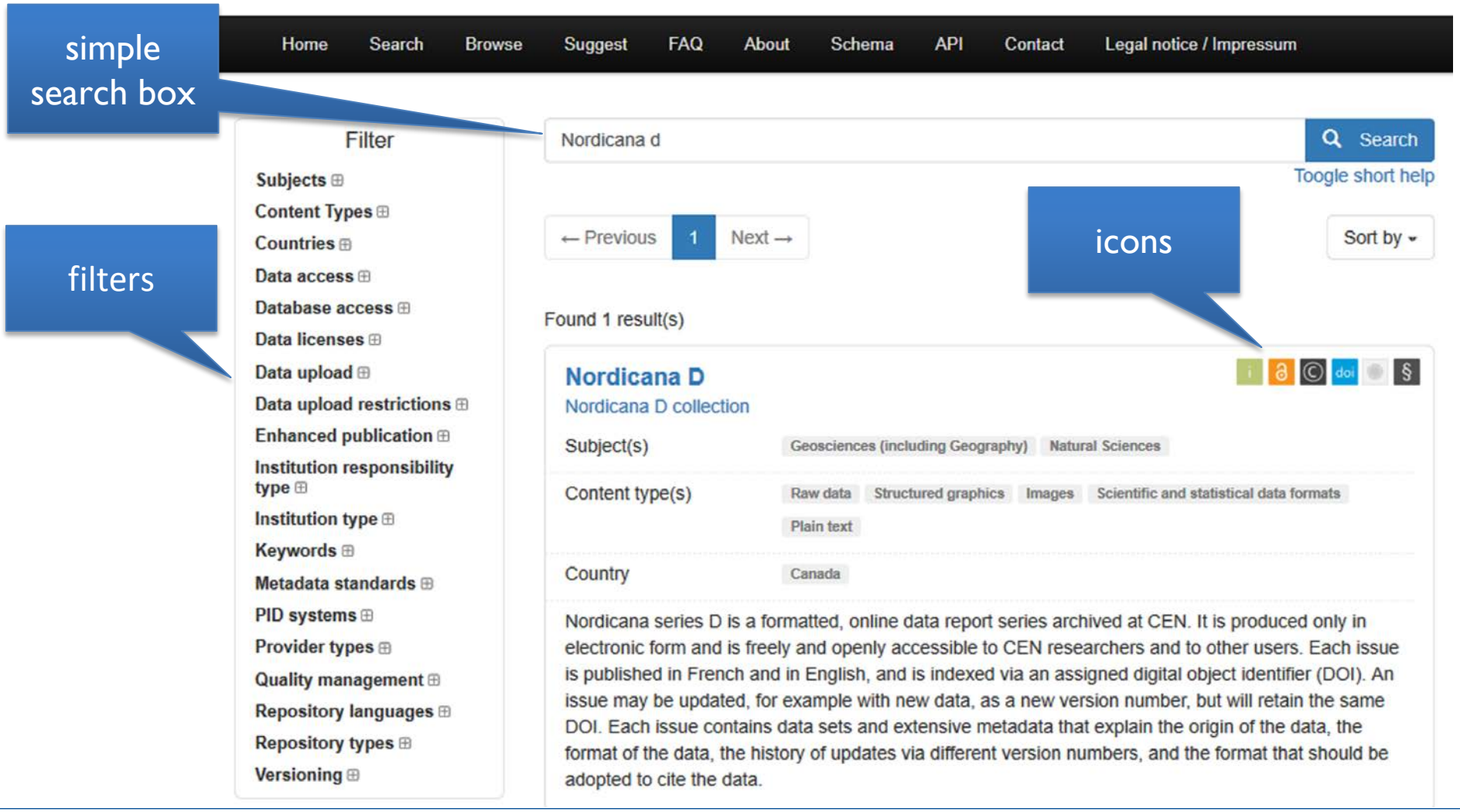
The terms of use and licenses of the data are provided by the research data repository.

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

overview

simple search box

filters



Questions?

Comments?

Thank you for you attention!