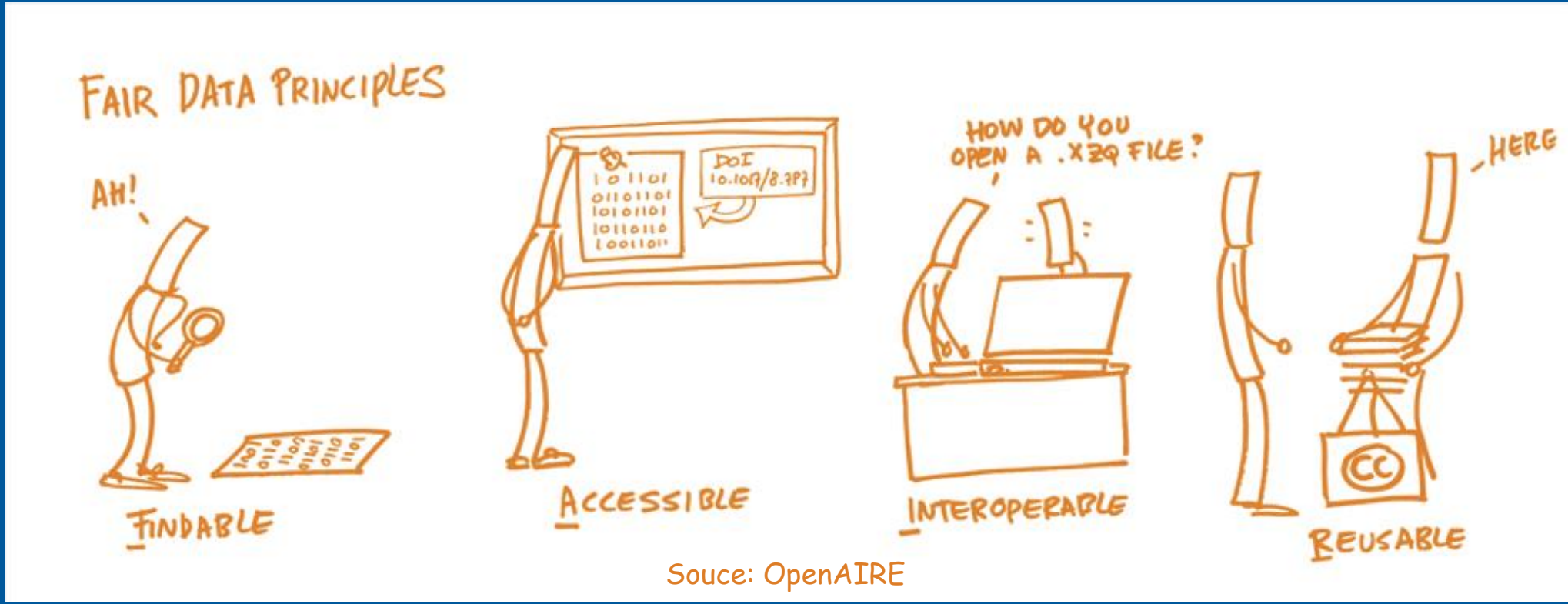
Die Rolle von Fachrepositorien und Datenjournals bei der Umsetzung der FAIR Prinzipien



Souce: OpenAIRE

Kirsten Elger, Boris Radosavljevic, Damian Ulbricht

Outline

- Introduction – open data and repositories
- COPDESS and Enabling FAIR Data
- Data Publications: technical background and best practices
- Data documentation: Practical tools and documents (by GFZ Data Services), reports, data journals
- Scholix: Incresasing the transparency of research products

Open data – an international request

→ following the **FAIR Principles** for Research Data Management

FAIR Principles

Findable

Accessible

Interoperable

Reusable

Data Publications – best practice for FAIR sharing data

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

- **Findable:** integration of standardised machine readable metadata in external data portals (e.g. DataCite, B2Find, Google Dataset Search)
- **Accessible:** via DOI, 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 researcher and institution)

Research Data Repositories

- permanent archives and access points to research data
- open access
- persistent identifier (ideally DOI)
- institutional, general, **domain**

"Domain repositories: These repositories **provide quality and standards** [for their domain]**, enriching and organizing data** from multiple sources **to facilitate new discoveries.** They are in many ways the **best stewards of the data** but are not currently well connected with most publishers, and many data are thus not finding their proper home."

Hanson et al.(2015) Eos, 96, doi:10.1029/2015EO022207

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

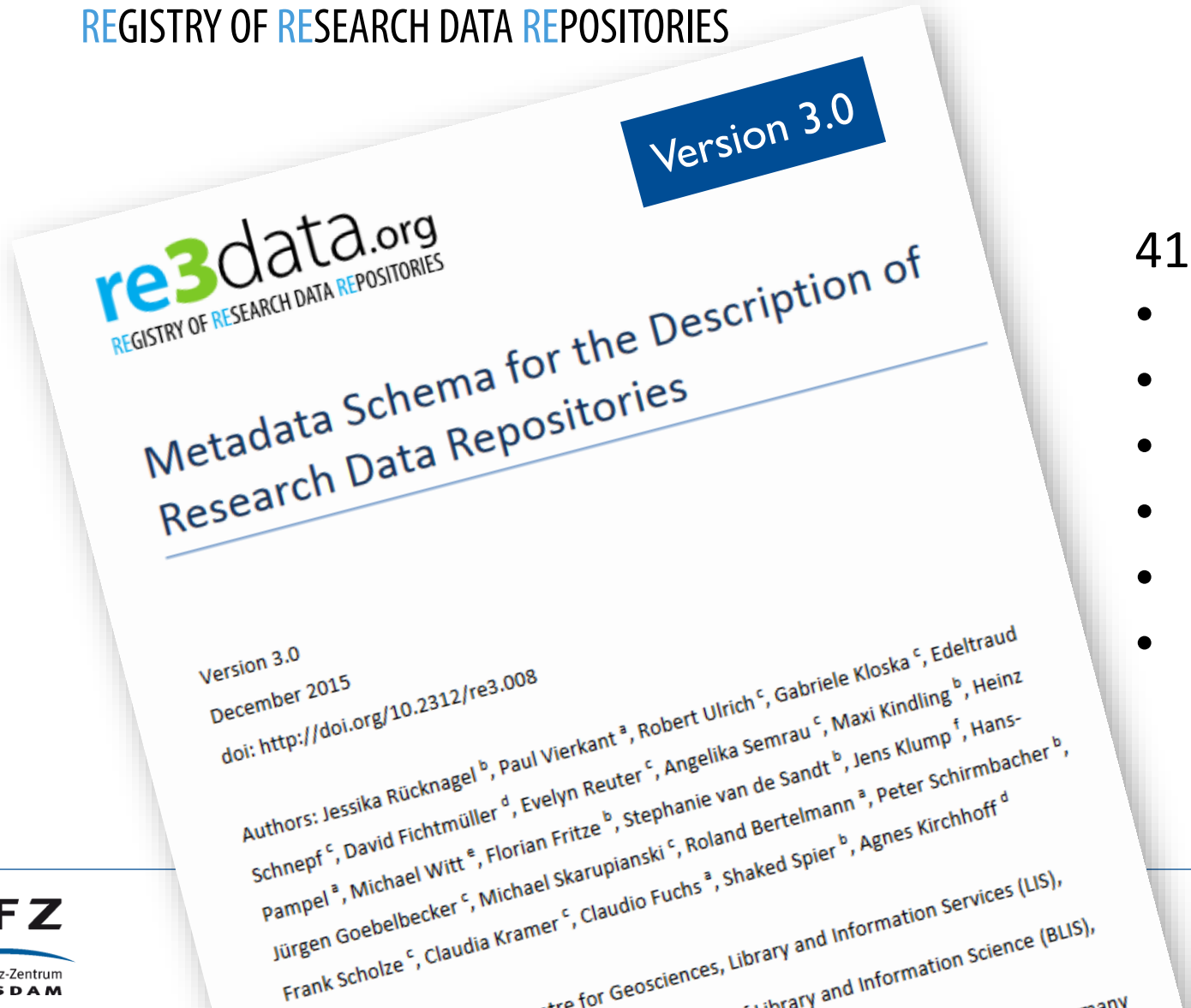
Mission

- **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

2466 registered repositories
(25 Feb 2020)

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Metadata Schema



41 Properties on

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

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

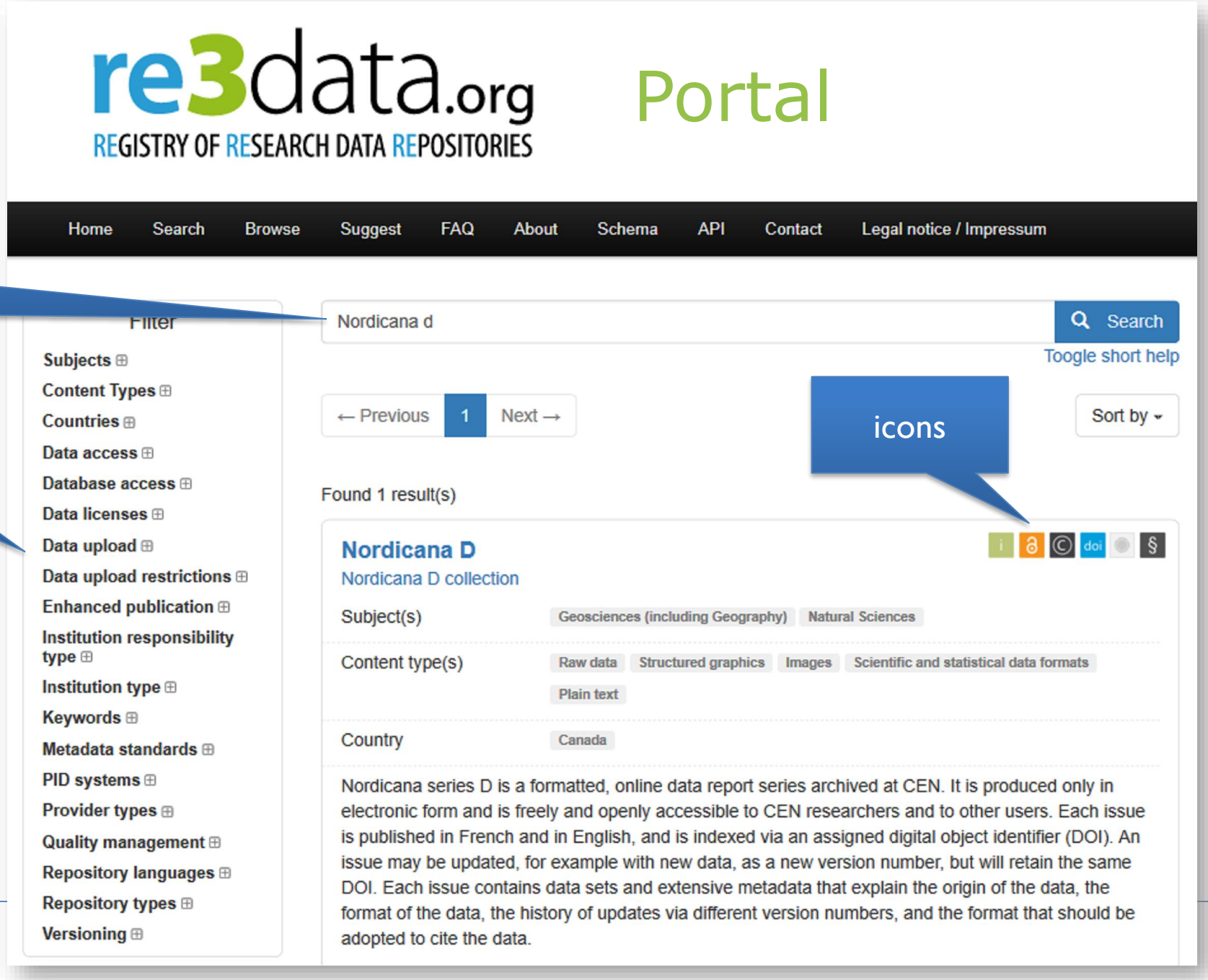
re3data.org Portal

REGISTRY OF RESEARCH DATA REPOSITORIES

simple search box

filters

icons



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Icons

facilitating the selection process of appropriate research data repositories

i — 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. — pi

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

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

News

New DFG Project:

re3data COREF

Community Driven Open Reference for Research Data Repositories

- 2020-2022
- KIT Karlsruhe Institute of Technology, Humboldt-Universität zu Berlin, GFZ Helmholtz Centre Potsdam, DataCite FIND, ACCESS, AND REUSE DATA
- To connect re3data as the reference for research data repositories with other services and infrastructures

Break – Questions?

Next: COPDESS and Enabling FAIR Data

Coalition for Publishing Data in the Earth and Space Sciences

COPDESS Statement of Commitment **2015**

Statement of Commitment from Earth and Space Science Publishers and Data Facilities


COPDESS @ AGU 2015

"Scholarly publication is a key high-value entry point in making data available, open, discoverable, and usable. Most publishers have statements related to the inclusion or release of data as part of publication, recognizing that inclusion of the full data enhances the value and is part of the integrity of the research. 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."

Coalition on Publishing Data in the Earth and Space Sciences

Data Publications are citable in research articles (COPDESS Statement of Commitment)

STATEMENT OF COMMITMENT

(January 2015)

- data should be stored in appropriate domain repositories.
- citations of data sets should be included within reference lists.
- include in research papers concise data availability statements.
- links to data sets in publications and corresponding links to journals in data facilities

http://www.copdess.org/statement-of-commitment/

After COPDESS: new Journal Policies

Data policy

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. Please find your appropriate data repository in the registry for research data repositories re3data.org. A data citation in a publication should resemble a bibliographic citation and be located in the publication's reference list. To foster the proper citation of data, Copernicus Publications requires all authors to provide a statement on the availability of underlying data as the last paragraph of each article (see section data availability). In addition, Copernicus Publications provides with Earth System Science Data (ESSD) a journal dedicated to the publication of data papers including peer review on data sets. Authors might consider submitting a data paper to ESSD in addition to their research paper in Copernicus Publications.

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Best practice following the Joint Declaration of Data Citation Principles initiated by FORCE 11:

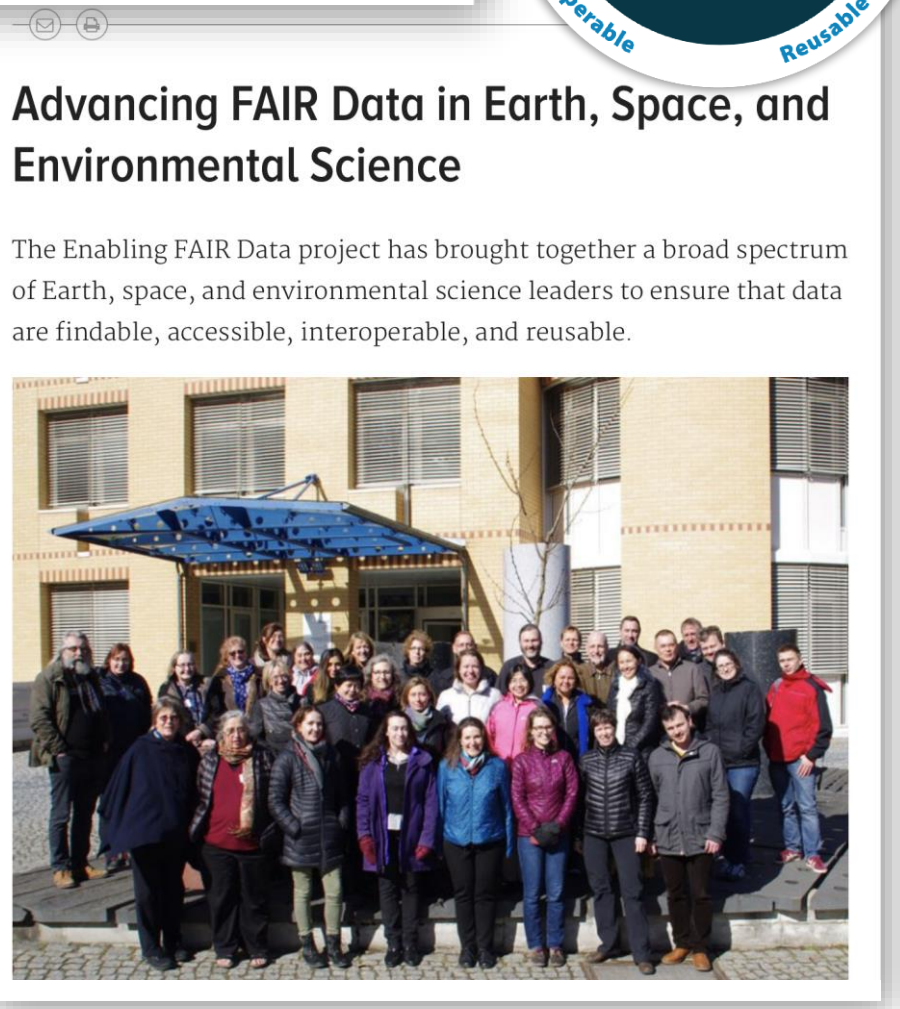
COPDESS

In addition to promoting these data citation principles, Copernicus Publications is a signatory of the Coalition on Publishing Data in the Earth and Space Sciences (COPDESS) commitment statement.

The Enabling FAIR Data project has brought together a broad spectrum of Earth, space, and environmental science leaders to ensure that data are findable, accessible, interoperable, and reusable.

From the Enabling FAIR Data Commitment Statement:

- Direct all core research outputs (data, software, samples and sample metadata) to trusted repositories.
  - Supplements will no longer be primary "archive" for data.
  - Data are cited via persistent identifier
- Adopt a shared set of author instructions (common set of expectations for authors in the ESES).
- Provide common expectations for publication peer review when evaluating science and determining if the data, metadata, and software are adequate.



http://www.copdess.org/enabling-fair-data-project/ and Stall, et al. (2018), Eos, 99, doi:10.1029/2018EO109301

How to cite a dataset?

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, Germany

Link to paper

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

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

1. Citation in the text

References

Heilbronner, R., Keulen, N., 2006. Grain size and grain shape analysis of fault rocks. Tectonophysics 427, 199–216.

Hubbert, M.K., 1951. Mechanical basis for certain familiar geologic structures. Geol. Soc. Am. Bull. 62, 1259–1273.

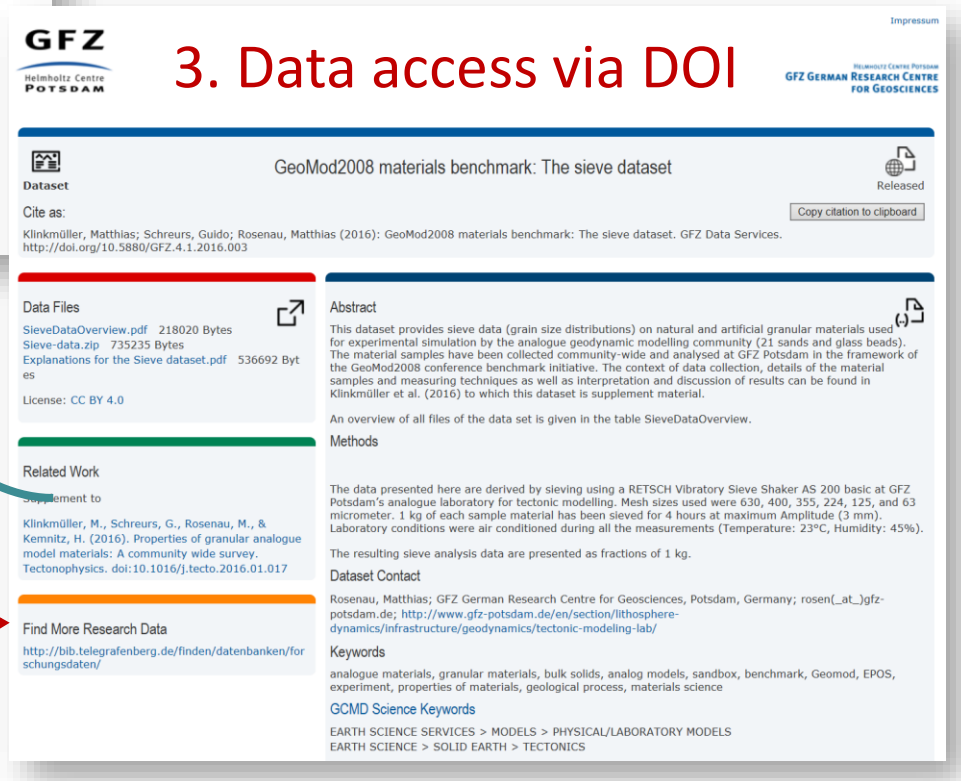
Klinkmüller, M., Schreurs, G., Rosenau, M., 2016a. GeoMod2008 materials benchmark: The ring shear test data set. GFZ Data Services. http://dx.doi.org/10.5880/GFZ.4.1.2016.002.

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.

2. Full reference with DOI in the References

3. Data access via DOI



GFZ Helmholtz Centre Potsdam — GFZ German Research Centre for Geosciences

Impressum

Dataset

GeoMod2008 materials benchmark: The sieve dataset

Released

Cite as:

Copy citation to clipboard

Klinkmüller, Matthias; Schreurs, Guido; Rosenau, Matthias (2016): GeoMod2008 materials benchmark: The sieve dataset. GFZ Data Services. http://doi.org/10.5880/GFZ.4.1.2016.003

Data Files

SieveDataOverview.pdf 218020 Bytes
Sieve-data.zip 735235 Bytes
Explanations for the Sieve dataset.pdf 536692 Bytes

License: CC BY 4.0

Related Work

Supplement to

Klinkmüller, M., Schreurs, G., Rosenau, M., & Kemnitz, H. (2016). Properties of granular analogue model materials: A community wide survey. Tectonophysics. doi:10.1016/j.tecto.2016.01.017

Find More Research Data

http://bib.telegrafenberg.de/finden/datenbanken/forschungsdaten/

Abstract

This dataset provides sieve data (grain size distributions) on natural and artificial granular materials used for experimental simulation by the analogue geodynamic modelling community (21 sands and glass beads). The material samples have been collected community-wide and analysed at GFZ Potsdam in the framework of the GeoMod2008 conference benchmark initiative. The context of data collection, details of the material samples and measuring techniques as well as interpretation and discussion of results can be found in Klinkmüller et al. (2016) to which this dataset is supplement material.

An overview of all files of the data set is given in the table SieveDataOverview.

Methods

The data presented here are derived by sieving using a RETSCH Vibratory Sieve Shaker AS 200 basic at GFZ Potsdam's analogue laboratory for tectonic modelling. Mesh sizes used were 630, 400, 355, 224, 125, and 63 micrometer. 1 kg of each sample material has been sieved for 4 hours at maximum Amplitude (3 mm). Laboratory conditions were air conditioned during all the measurements (Temperature: 23°C, Humidity: 45%).

The resulting sieve analysis data are presented as fractions of 1 kg.

Dataset Contact

Rosenau, Matthias; GFZ German Research Centre for Geosciences, Potsdam, Germany; rosen(_at_)gfz-potsdam.de; http://www.gfz-potsdam.de/en/section/lithosphere-dynamics/infrastructure/geodynamics/tectonic-modeling-lab/

Keywords

analogue materials, granular materials, bulk solids, analog models, sandbox, benchmark, Geomod, EPOS, experiment, properties of materials, geological process, materials science

GCMD Science Keywords

EARTH SCIENCE SERVICES > MODELS > PHYSICAL/LABORATORY MODELS
EARTH SCIENCE > SOLID EARTH > TECTONICS

Break – Questions?

Next: Data Publications: technical background and best practices

Data Publications - components

- Data
- Metadata

Types of Metadata

Descriptive metadata	For finding or understanding a resource
Administrative metadata - Technical metadata - Preservation metadata - Rights metadata	 - For decoding and rendering files - Long-term management of files - Intellectual property rights attached to content
Structural metadata	Relationships of parts of resources to one another
Markup languages	Integrates metadata and flags for other structural or semantic features within content

Descriptive Metadata: Contextual metadata
Metadata for data discovery

Source: NISO: Understanding Metadata – a Primer

Contextual Metadata

highly variable between the disciplines but key information for data reuse

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	m
11	DATE_START	Date	date of site start	date in UTC	dd-mmm-yyyy
12	DATE_END	Date	date of site end	date in UTC	dd-mmm-yyyy

README

Datei Bearbeiten Format Ansicht ?

AVERTISSEMENT / WARNING

Même si des efforts sont déployés pour

Although efforts are made to ensure th

AUTEUR(S) / AUTHOR(S)

Centre d'études nordiques

RÉSUMÉ / SUMMARY

Les données de ce numéro de Nordica

The datasets in this issue of Nordi

CITATION DES DONNÉES / DATA CITATI

CEN 2014. Données environnementale

CEN 2014. Environmental data fro

SITE(S)

Nom / Name

Ellesmere Parks Canada (ELLEPAR)

83.09396

DESCRIPTION

SITE: Ellesmere Parks Canada (ELLEPAR)
Profondeurs des / températures de sol (m) / Ground temperature depth range (m): 0.20

MESURE(S) / MEASUREMENT(S)

TYPE: Température moyenne du sol (Degré celsius (°C)) / Average ground temperature (Degree Celsius (°C))

Data Articles/ Reports

Metadata for data discovery: example DOI Landing Page

title
citation

download
data files

related
work

description/
abstract

Keywords

spatial
coverage

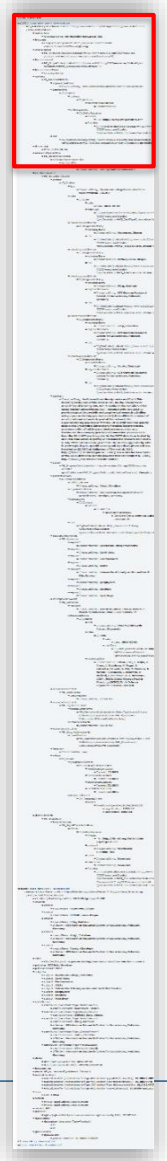
Essential for data discovery, DOI registration, etc: international standards across all disciplines

XML metadata following international standards (DataCite, ISO19115)

Machine-readable metadata standards

iso19115: close inline view / download xml

- MD_Metadata (xsi:schemaLocation=http://www.isotc211.org/2005/gmd http://www.isotc211.org/2005/gmd/gmd.xsd)
  - fileIdentifier
    - CharacterString: **doi:10.5880/igets.po.l1.001**
  - language
    - LanguageCode (codeList=http://www.loc.gov/standards/iso639-2/ codeListValue=eng): **eng**
  - characterSet
    - MD_CharacterSetCode (codeList=http://www.isotc211.org/2005/resources/codeList.xml#MD_CharacterSetCode codeListValue=utf8):
  - hierarchyLevel
    - MD_ScopeCode (codeList=http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml#MD_ScopeCode codeListValue=):
  - hierarchyLevelName
    - CharacterString:
  - contact
    - CI_ResponsibleParty
      - organisationName
        - CharacterString: **GFZ German Research Center for Geosciences**
      - contactInfo
        - CI_Contact
          - address
            - CI_Address
              - electronicMailAddress
                - CharacterString:
          - onlineResource
            - CI_OnlineResource
              - linkage
                - URL: **http://www.gfz-potsdam.de/**
              - function
                - CI_OnLineFunctionCode (codeList=http://ww /2005/resources/Codelist /gmxCodelists.xml#CI_OnLineFunctionCode co
      - role
        - CI_RoleCode (codeList=http://www.isotc211.org/2005/resources/Codelist /gmxCodelists.xml#CI_RoleCode codeListValue=pointOfContact): **pointOfC**
  - dateStamp
    - Date: **2017-01-06**
  - referenceSystemInfo
    - MD_ReferenceSystem
      - referenceSystemIdentifier
        - RS_Identifier
          - code
            - CharacterString: **urn:ogc:def:crs:EPSG:4326**
  - identificationInfo
    - MD_DataIdentification
      - citation



GFZ Data Services
Geosciences Data Publisher

- XML metadata in DataCite, ISO19115 standards
- Exchange via OAI-PMH interface

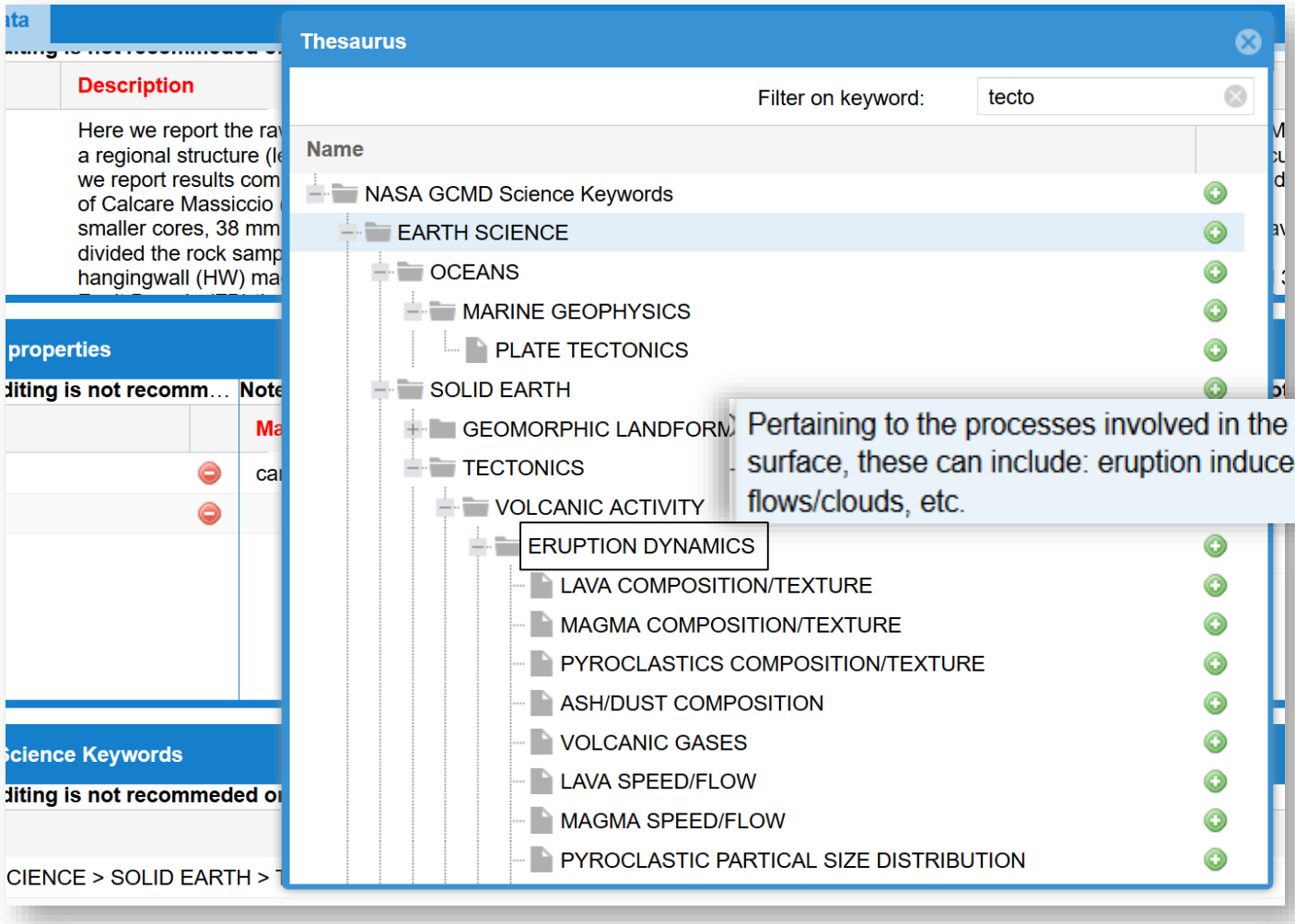
schema.org

- Machine-readable websites
- JSON LD
- Google Dataset Search

Controlled Vocabularies in the Metadata Editor

NASA Global Change Master Directory (GCMD) Keywords

Example: EARTH SCIENCE > SOLID EARTH > TECTONICS > VOLCANIC ACTIVITY > ERUPTION DYNAMICS



INSPIRE
Infrastructure for Spatial Information in Europe

INPRIRE ISO 19115 Keywords

RDF-Version

IUGS
CGI
Commission for the Management and Application of Geoscience Information
<GeoSciML/>

GeoSciML: Geoscience Vocabularies for Linked Data

Persistent Identifier in data publications

for data, software, cross-references to related work

List of funders with DOIs

uniquely identifying persons

PID for physical samples, cross references to samples underlying measurements

New PID for Institutions

and others...

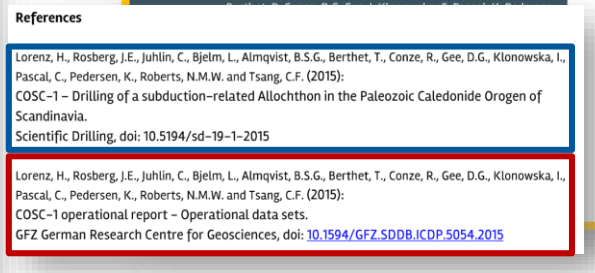
Linking papers, data, samples, …

DataCite related Identifier

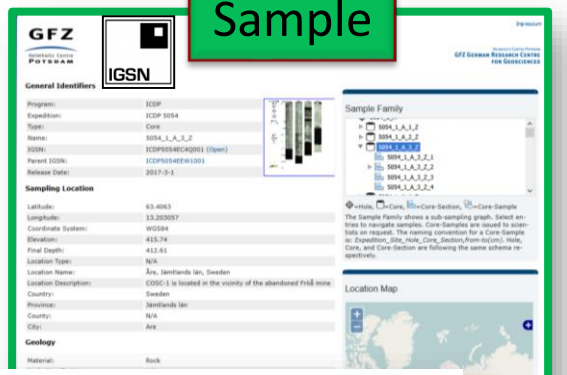
IsCitedBy	indicates that B… discovery).
Cites	indicates that A… discovery).
IsSupplementTo	
IsSupplement	
IsContinuedB	
Continues	
HasMetadata	
IsMetadataFo	
IsNewVersion	
IsPreviousVer	
IsPartOf	
HasPart	
IsReferencedI	
References	
IsDocumente	
Documents	
IsCompiledBy	
Compiles	
IsVariantForm	
IsOriginalFor	
IsIdenticalTo	

Data

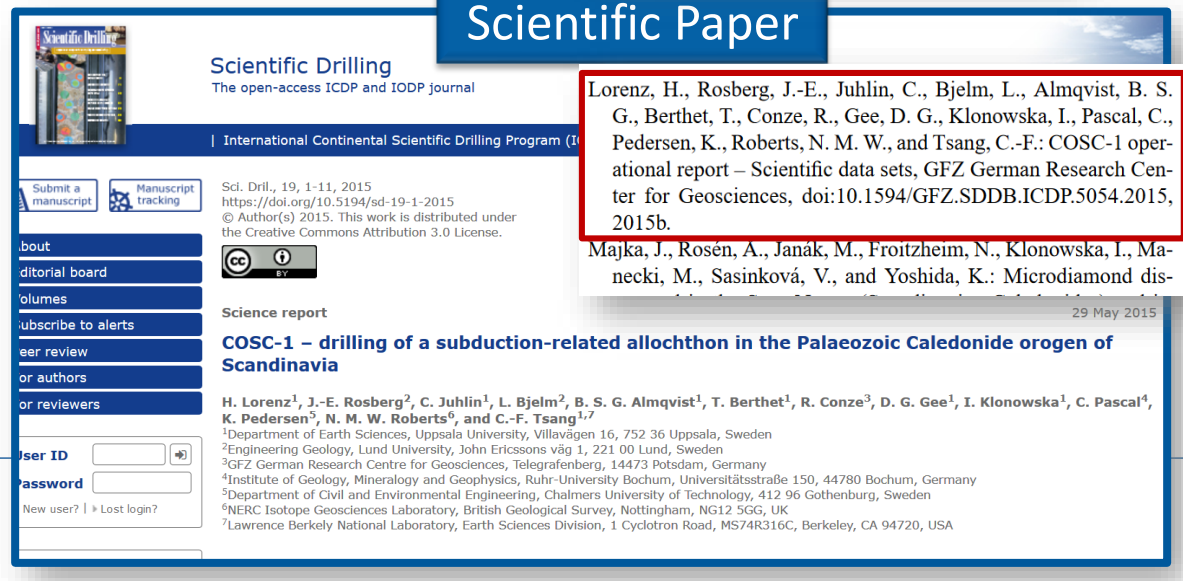
Data Report



Sample



Scientific Paper



Lorenz, H., Rosberg, J.-E., Juhlin, C., Bjelm, L., Almqvist, B. S. G., Berthet, T., Conze, R., Gee, D. G., Klonowska, I., Pascal, C., Pedersen, K., Roberts, N. M. W., and Tsang, C.-F.: COSC-1 operational report – Scientific data sets, GFZ German Research Center for Geosciences, doi:10.1594/GFZ.SDDB.ICDP.5054.2015, 2015b.

Majka, J., Rosén, A., Janák, M., Froitzheim, N., Klonowska, I., Manecki, M., Sasinková, V., and Yoshida, K.: Microdiamond dis-

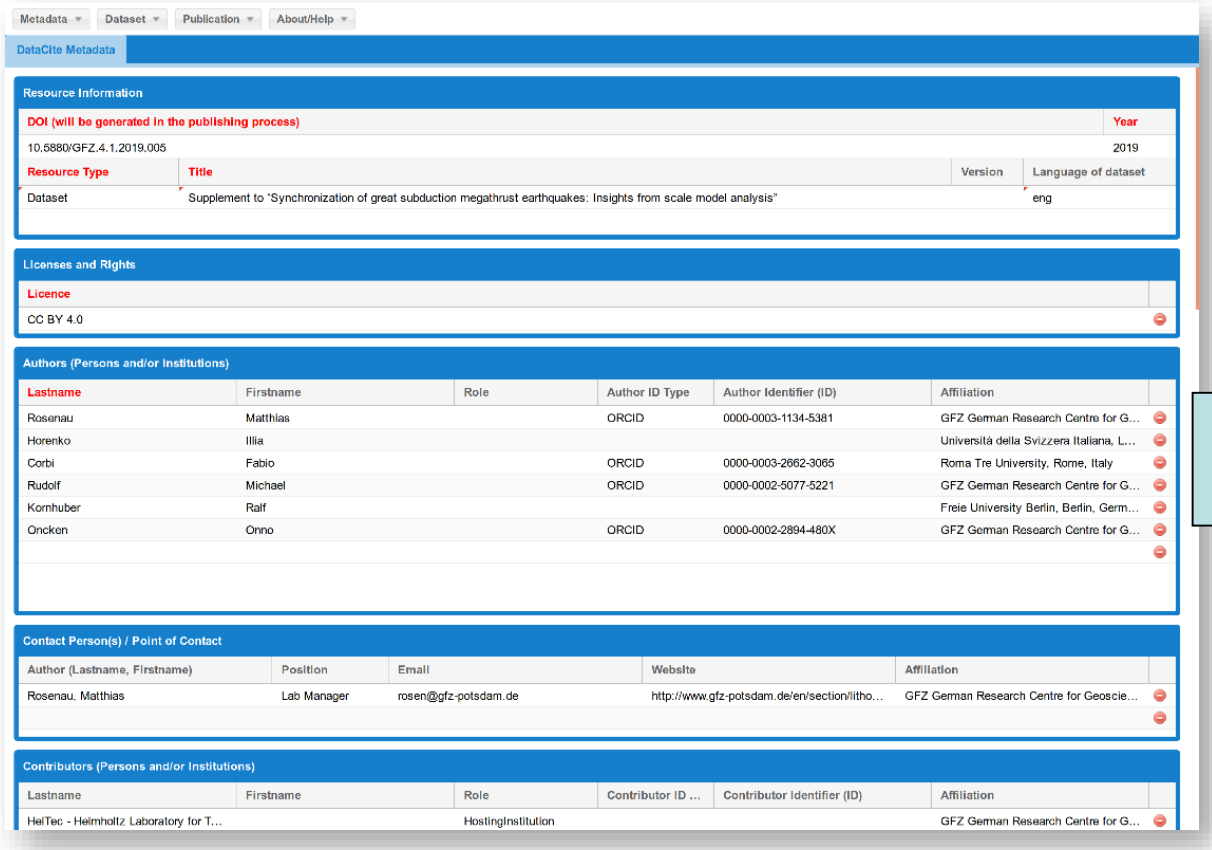
Break – Questions?

Next: Data documentation: Practical tools and guidelines, data reports, data journals

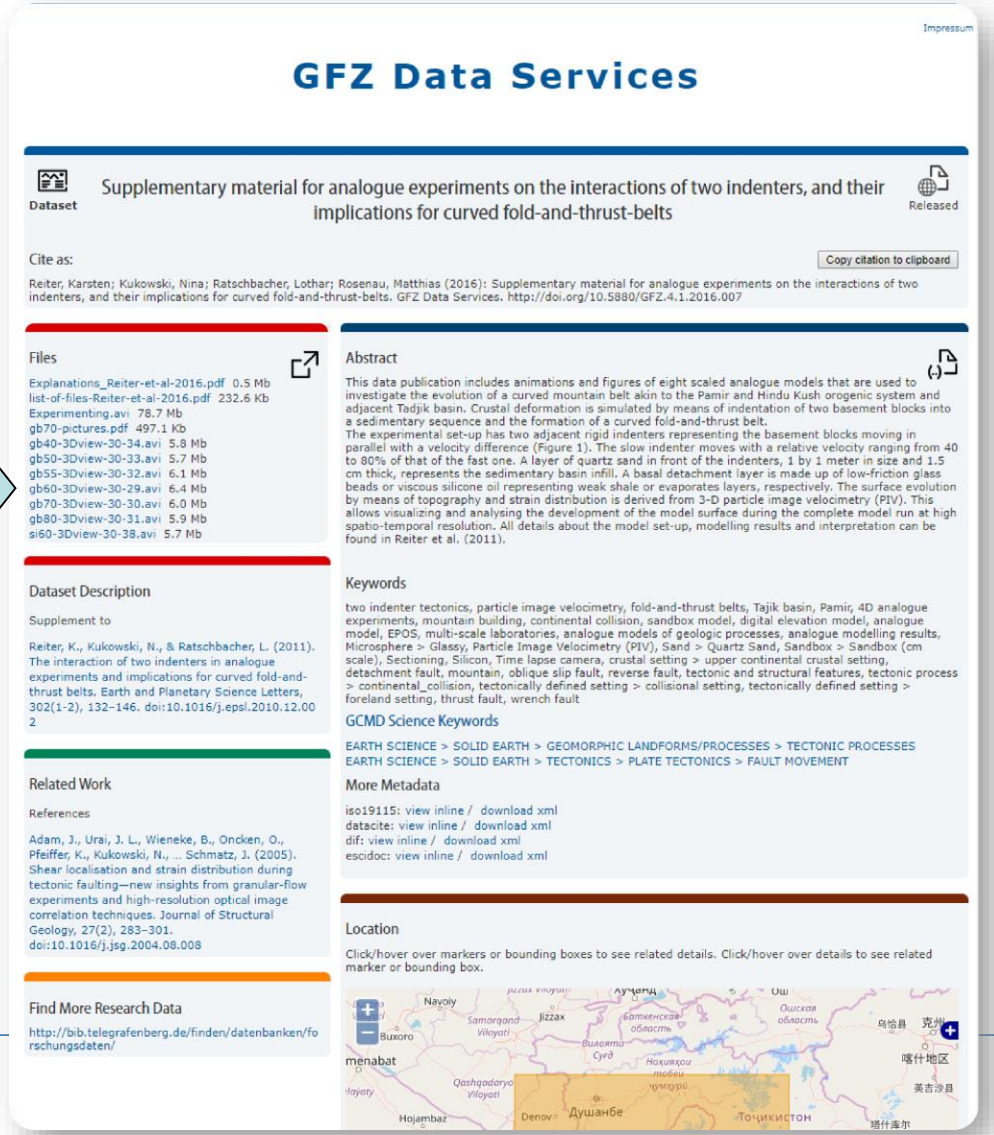
Outline

- Practical tools and documents (by GFZ Data Services)
  - GFZ Metadata Editor
  - Data description template
  - Machine-readable data tables

Standardised Metadata Workflow



XML

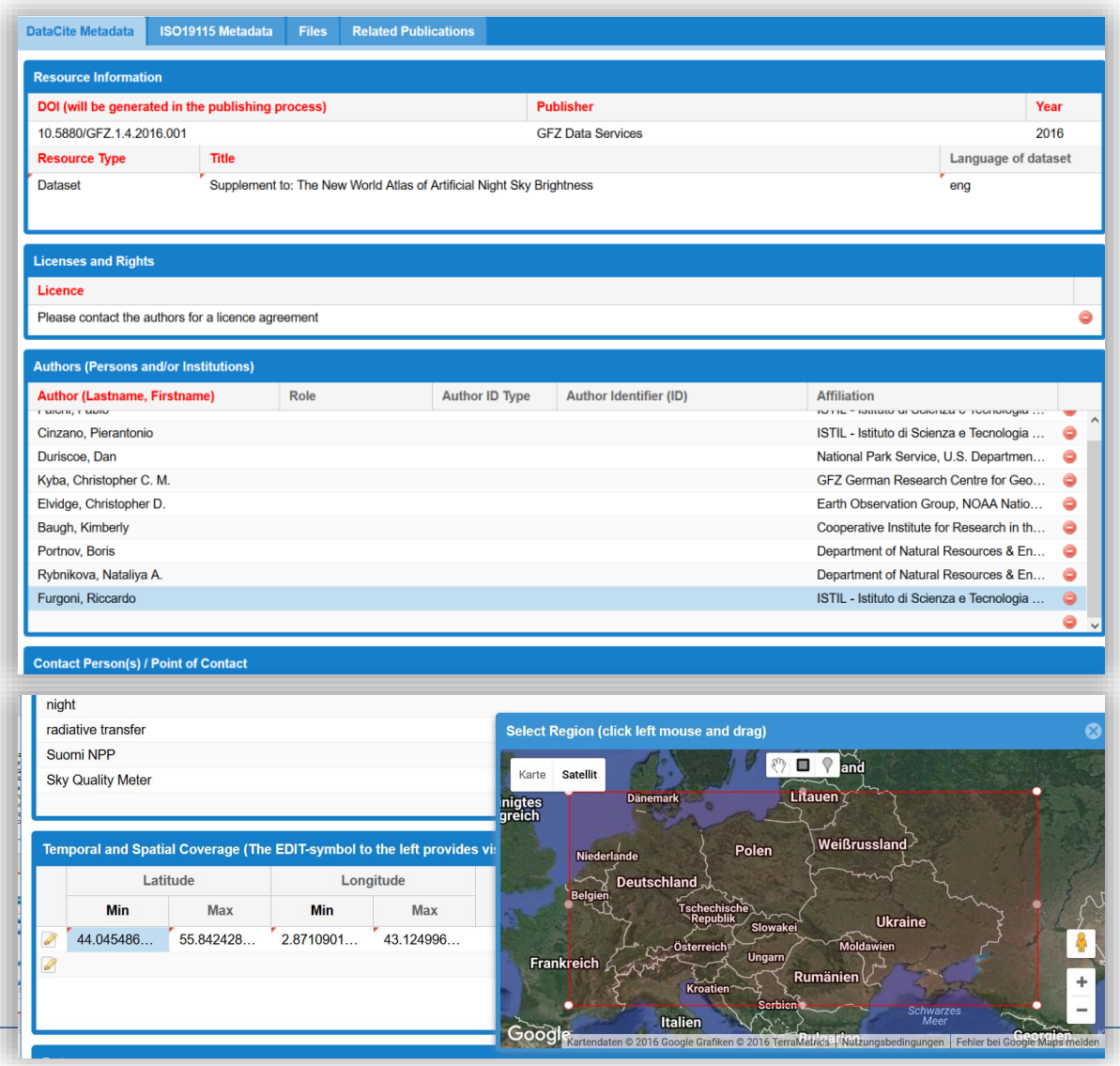


GFZ Metadata Editor (Java Script „translator“)

Input:
by scientists

„Special“ Features:

- Interactive map
- ORCID and Fundref
- Controlled vocabularies
- Multiple affiliations for authors



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

DataCite
FIND, ACCESS, AND REUSE DATA

GFZ Data Services Metadata Catalogue

EPOS, B2FIND. ENVRIplus, etc.

Access via: http://dataservices.gfz-potsdam.de/portal/about.html „Publishing step by step“

Description Template

- Many users are unaware of what a data publication represents and what to include in description
- Increase the quality of metadata
- Reduces curation workload
- Uniform format aids comprehension

Data Description Template

Descriptive Title of Dataset

Author-1[1], Author-2[2] …
- *Affiliation1, City, Country*
- *Affiliation2, City, Country*

•1. Licence

Creative Commons Attribution 4.0 International License (CC BY 4.0)

2. Citation

When using the data please cite:
Will be added by GFZ Data Services
The data are supplementary material to:
Please add citation and DOI of the key paper/s

- **3. Data Description**
  - **Sampling method**
  - **Analytical procedure**
  - **Data processing**
- **4. File description**
  - **File inventory**
  - **File naming convention**
  - **Description of data tables**
- **5. References**

Sampling method

How was sample obtained? Is the sample assigned with e.g. International Geo Sample Numbers?
How were the samples prepared for analysis?

Analytical procedure:

Instrument information, platform, calibration, standards used, number of aliquots, sample quantities,

Data processing

Describe how the analytical data was treated to obtain the dataset you would like to publish. What transformations, statistical methods did you apply?

3. Data Description

- Summary of the data description
- „Abstract“ on the DOI Landing Page

  → Internal part of metadata

  → transferred to other portals as machine-readable XML

  → essential for data discovery
- Understandable for the broader scientific community
- Scientific purpose for data collection (and project) and summary of technical data description

Example Description: Before

Stimulation data for each of the analyzed EGS projects.

The are provided in tabular form (CSV). The file names indate the project.

Definition of columns in the data tables (also in the header of the data):

- V = Cumulative injected volume (m³) - Ehyd = Applied hydraulic energy (J) - MaxM0 = Maximum observed seismic moment (Nm) - CumM0 = Cumulative seismic moment (Nm) - IE = Injection efficiency (-)

Example Description: After (= with the template)

The here provided data are part of a broader analysis of past and present stimulation projects, revealing that the temporal evolution and growth of maximum observed moment magnitudes may be linked directly to the injected fluid volume and hydraulic energy. Analyzed projects include the most prominent European Enhanced Geothermal System (EGS) projects in Basel, Switzerland (BAS) and Soultz-sous-Forêts (STZ), France. In Soultz, three different stimulations over the course of 10 years were performed in different wells and different depths. Therefore, we differentiate between the injections in 1993 (STZ93), 2000 (STZ00), and in 2003 (STZ03). We also included the deepest EGS Project to date (St1), located in Helsinki, Finland. Furthermore, we included the fluid-injection experiment from the German super deep scientific drilling hole (KTB), two Australian EGS projects, located at Paralana (Para) and the 2003 Cooper Basin (CBN) injection, as well as the EGS project near Pohang, South Korea. Finally, we also considered a single well injection period at the Berlín geothermal field (BGF), El Salvador, representing the only hydrothermal site considered here.

For each project the cumulative volume injected is provided along with the applied hydraulic energy, maximum observed seismic moment, cumulative seismic moment, and injection efficiency as tab separated ASCII files with the .csv extension. All stimulation files are combined into a single .zip archive. More details on processing steps and references herein can be found in the accompanying data description.

4. File inventory

Explanation of folder structure, file list and file contents included in data publication of mechanical data from rotary shear experiments on material derived from the Alpine Fault during the Deep Fault Drilling Project (phases 1A and 1B).

The zip-file contains folders for each individual experiment (33 in total), listed in **Table 1** in the Appendix of this document. In addition, Table1 is provided in the Data Files section of the DOI Landing Page (Table 1-Niemeijer-2017.pdf).

Each folder contains 5 different files and a subfolder (and each filename follows the same naming convention: the letter u, followed by a 3-digit number, indicating the experiment number):

- **datasheet.pdf**: Logsheet indicating the conditions of the material, sample material used and notes on the progress of the experiment, including times at which boundary conditions were changed.
- **u101AF_300**: Original data-file, a tab-separated text file with 12 columns. Note that the column headers and units in this file are NOT the correct headers, rearrangement and proper naming of the columns occurs within the XLOOK script.
- **u101AF_300l:** A "look" file, built from the original file using "asc2look" (see link below) of the data processing software "XLOOK", a program developed by Chris Marone which is available on github (https://github.com/PennStateRockandSedimentMechanics/xlook).

Example for a file inventory from Niemeijer et al. (http://doi.org/10.5880/icdp.5052.002)

4. File inventory/ list of files

It is part of the following data publication and should be cited as:
Rosenau, Matthias; Pohlenz, Andre; Kemnitz, Helga; Warsitzka, Michael (2018): Ring-shear test data of quartz sand G23 used for analogue experiments in the Helmholtz Laboratory for Tectonic Modelling (HelTec) at the GFZ German Research Centre for Geosciences in Potsdam. GFZ Data Services. http://doi.org/10.5880/GFZ.4.1.2019.004

ZIP folder	Folder size	File name	File format	Content
		2019-004_Rosenau-et-al_List_of_files	.pdf	List of files
		2019-004_Rosenau-et-al_Description-of-data	.pdf	Description of data and methods
Scripts	978 kb	RSTanalysis	.py	Python script for analysing and plotting friction and time series data (Mohr plot, histograms, shear curves)
		VSTanalysis	.py	Python script for analysing and plotting VST data
Data files	2932 kb	421-01_GFZ_quartzsandG23_vst	.pdf	Visualization of the VST data
		421-01_GFZ_quartzsandG23_vst	.txt	Table of the VST data: time, shear velocity, normal force, shear force
		426-01_GFZ_quartzsandG23_dynamic	.txt	Pairs of normal stress and corresponding shear strength for dynamic friction
		426-01_GFZ_quartzsandG23_hist	.pdf	Histograms of friction coefficients and cohesions
		426-01_GFZ_quartzsandG23_lineregr	.pdf	Mohr plot of friction data
		426-01_GFZ_quartzsandG23_peak	.txt	Pairs of normal stress and corresponding shear strength for peak friction
		426-01_GFZ_quartzsandG23_reactivation	.txt	Pairs of normal stress and corresponding shear strength for reactivation friction
		426-01_GFZ_quartzsandG23_ts	.pdf	Visualization of time series data (shear curves): Shear stress vs. shear displacement for 18 measurements
		426-01_GFZ_quartzsandG23_ts	.txt	Table of time series data for 18 measurements of shear stress (Pa, collumns 2-19) at given normal stresses (Pa, first cell in each column) vs. time (column 1)

Example for an associated „List of Files" (Rosenau et al, http://doi.org/10.5880/GFZ.4.1.2019.005)

Definition of data tables

Description of data tables

You should include a table explaining the column headers in case of tabular data:

Column header	unit	Description
SampleID		Sample Identifier
Lat	DD,dddd	Latitude in WGS84 in decimal degrees
Long	DD,dddd	Longitude in WGS84 in decimal degrees
MeasurementXY	xy	Value of XY in units xy

5. References

4. References

References for dataset and those cited above. If the data follows a standard format defined
community, please cite the standard. Does this data incorporate other authors' datasets? Ci
should contain a DOI (or URL for reports or websites). The DOI numbers shall be provided wi
https:// code facilitating to directly link to the cited paper (see Figure 3)

8. References

Corbi, F., F. Funiciello, S. Brizzi, S. Lallemand, and M. Rosenau (2017), Control of asperities size and spacing on seismic behavior of subduction megathrusts, Geophys. Res. Lett., 44, 8227–8235, https://doi.org/10.1002/2017GL074182

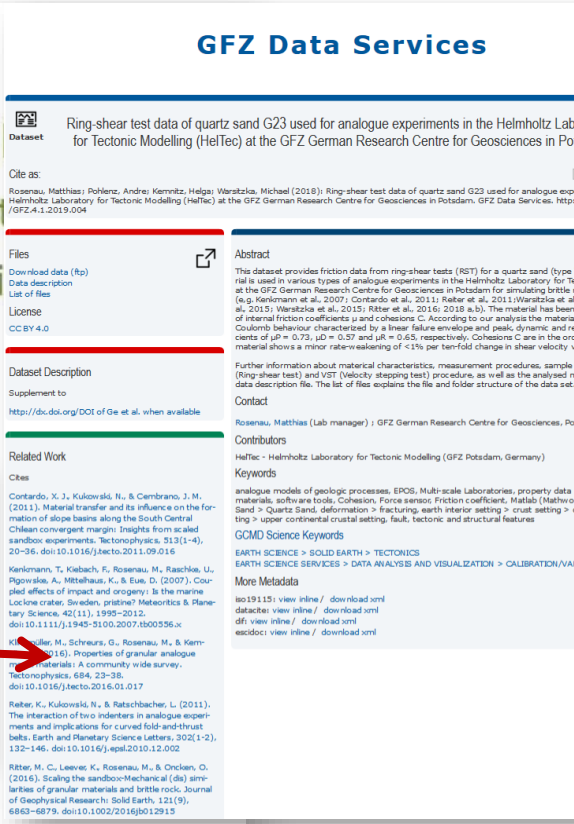


Figure 3: example of a reference with executable DOI link leading to the referred publication (example from Rosenau et al., http://doi.org/10.5880/GFZ.4.1.2019.005).

Machine Readable Tables

Table S1a Chemical composition of soil, saprolite and rock samples at CON and MIT (for colour coding see table caption)

sample ID	IGSN†	brief sample description	mean depth	XRF analyses: Raw data (major oxides)							
				SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MnO	MgO	CaO	Na_2O
			(m)	(wt%)	(wt%)	(wt%)	(wt%)	(wt%)	(wt%)	(wt%)	(wt%)
CON regolith - bedrock depth profile											
CON 14	GFDUH00LT	soil, Ah horizon	0.2	57	0.86	15	6.4	0.12	2.0	0.19	0.87
CON 13	GFDUH00LU	soil, Ah/Bw horizon	0.4	60	0.89	16	6.3	0.13	1.9	0.19	0.93
CON 12	GFDUH00LV	soil, Bw1 horizon	0.6	51	0.79	14	5.5	0.12	1.6	0.16	0.87
CON 11	GFDUH00LW	soil, Bw1 horizon	0.8	58	0.93	18	7.4	0.15	2.1	0.20	1.0
CON 10	GFDUH00LX	soil, Bw1 horizon	1.0	59	0.94	18	7.3	0.13	2.2	0.24	1.0
CON 9	GFDUH00LY	soil, Bw1 horizon	1.2	62	0.94	17	6.9	0.08	2.3	0.28	1.1

Machine readable data tables

Motivation:

- Easy access and data manipulation by machines (via scripts)
- Information can be easily read by a computer

- Especially for long data tables
- standardised tables
- Precondition for machine-learning, AI….

Things to keep in mind

- No combined cells
- No free lines or rows
- No metadata in the data table (definition of table heads in separate metadata spreadsheet)
- Separate data types in separate tables

Machine Readable Tables

These data are freely available under the Creative Commons Attribution 4.0 International Licence (CC BY 4.0)

when using the data please cite as:

Uhlig, D. and von Blanckenburg, F. (2019): Geochemical and isotope data on rock weathering, and nutrient balances during fast forest floor turnover in montane, temperate forest ecosystems. GFZ Data Services. http://doi.org/10.5880/GFZ.3.3.2019.004

The data are supplementary material to: Uhlig, D. and von Blanckenburg, F. (2019). How slow rock weathering balances nutrient loss during fast forest floor turnover in montane, temperate forest ecosystems. Frontiers in Earth Sciences. http://doi.org/10.3389/feart.2019.00159

sample ID	IGSN	brief sample description	mean depth (m)	SiO2 (wt%)	TiO2 (wt%)	Al2O3 (wt%)	Fe2O3 (wt%)	MnO (wt%)	MgO (wt%)	CaO (wt%)	Na2O (wt%)
CON 14	GFDUH00LT	soil, Ah horizon 0.2	57	0.86	15	6.4	0.12	2.0	0.19	0.87	
CON 13	GFDUH00LU	soil, Ah/Bw horizon	0.4	60	0.89	16	6.3	0.13	1.9	0.19	0.93
CON 12	GFDUH00LV	soil, Bw1 horizon 0.6	51	0.79	14	5.5	0.12	1.6	0.16	0.87	
CON 11	GFDUH00LW	soil, Bw1 horizon 0.8	58	0.93	18	7.4	0.15	2.1	0.20	1.0	

GFZ Data Reports/ Technical Reports

First Data Report published in 2011: persistently online accessible and citable with DOI

GFZ Data Reports/ Technical Reports:

- Flexible format for: “enhanced” data or software description, field guides
- standardised templates for each discipline/ project (ICDP, EnMAP)
- internal review by domain experts
- Project-specific design if required

Data Journals

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

Earth System Science Data

Copernicus.org
Meetings & Open Access Publications

- First data journal, launched in 2008
- „international, interdisciplinary journal for the publication of articles on original research data"
- **No interpretation of the data!**

- In 2020:
  - more than 500 peer-reviewed descriptions of easily- and freely-accessible data products from
  - more than 4000 data providers
  - archiving their products at more than 100 data centres
  - IF = 10.951

Break – Questions?

Next: Scholix: Incresasing the transparency of research products

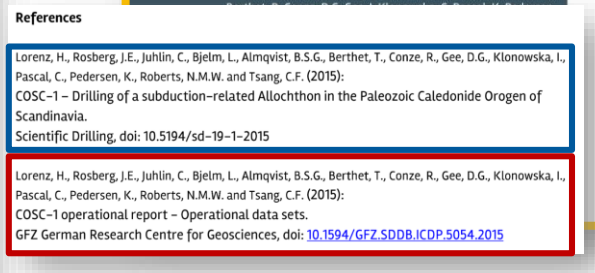
Linking papers, data, samples, …

DataCite related Identifier

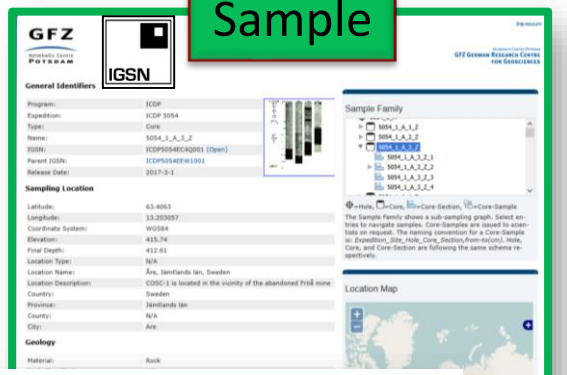
IsCitedBy	indicates that B… discovery).
Cites	indicates that A… discovery).
IsSupplementTo	
IsSupplement	
IsContinuedB	
Continues	
HasMetadata	
IsMetadataFo	
IsNewVersion	
IsPreviousVer	
IsPartOf	
HasPart	
IsReferencedI	
References	
IsDocumente	
Documents	
IsCompiledBy	
Compiles	
IsVariantForm	
IsOriginalFor	
IsIdenticalTo	

Data

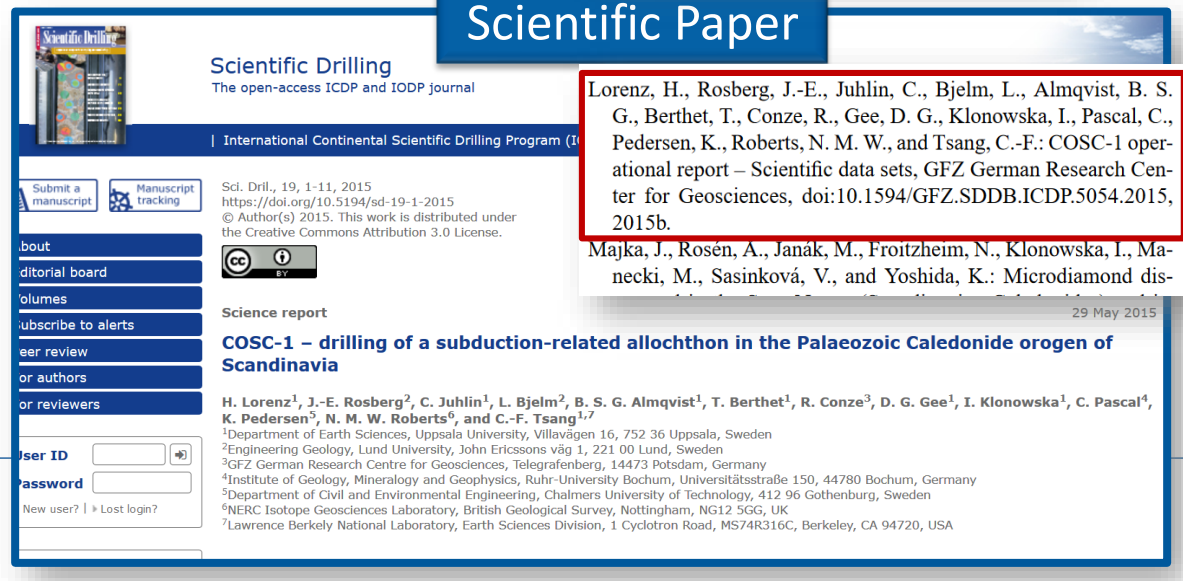
Data Report



Sample



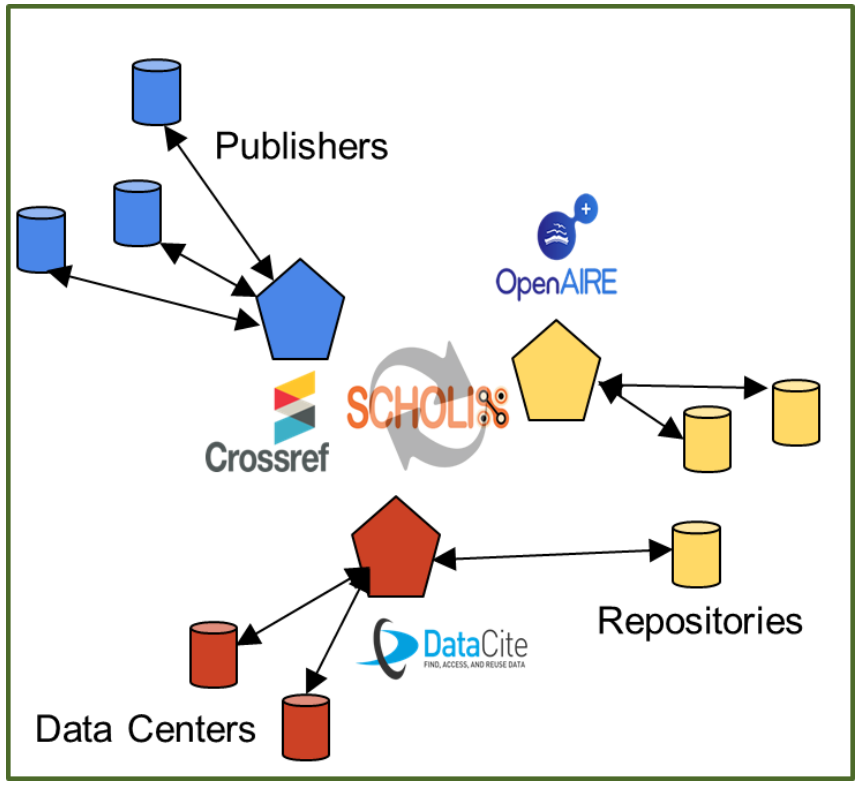
Scientific Paper



Lorenz, H., Rosberg, J.-E., Juhlin, C., Bjelm, L., Almqvist, B. S. G., Berthet, T., Conze, R., Gee, D. G., Klonowska, I., Pascal, C., Pedersen, K., Roberts, N. M. W., and Tsang, C.-F.: COSC-1 operational report – Scientific data sets, GFZ German Research Center for Geosciences, doi:10.1594/GFZ.SDDB.ICDP.5054.2015, 2015b.

Majka, J., Rosén, A., Janák, M., Froitzheim, N., Klonowska, I., Manecki, M., Sasinková, V., and Yoshida, K.: Microdiamond dis-

Scholix: linking data with papers



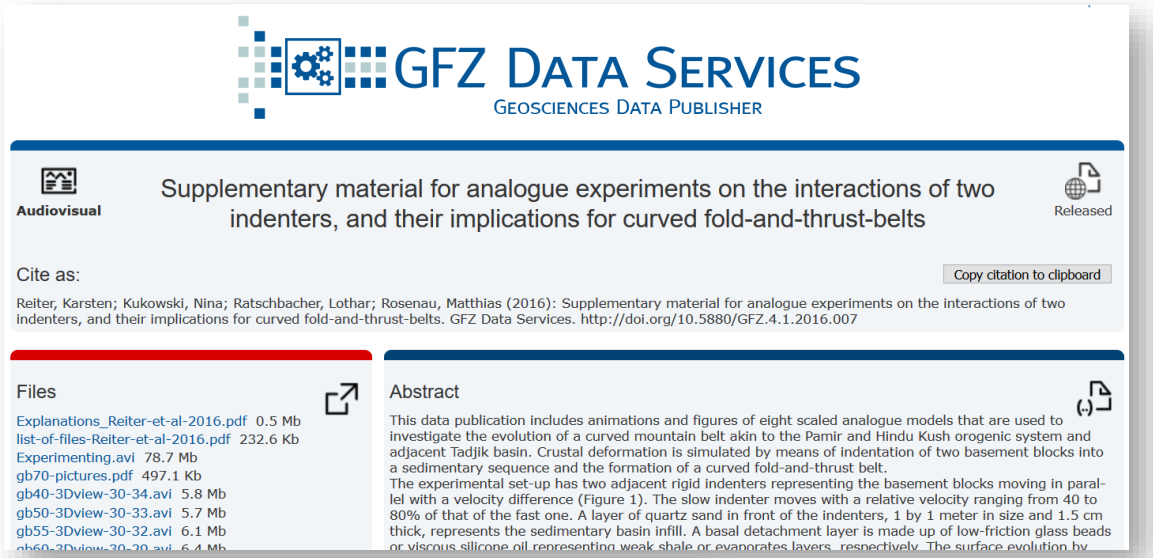
Credit: Wouter Haak, Elsevier

SCHOLIX

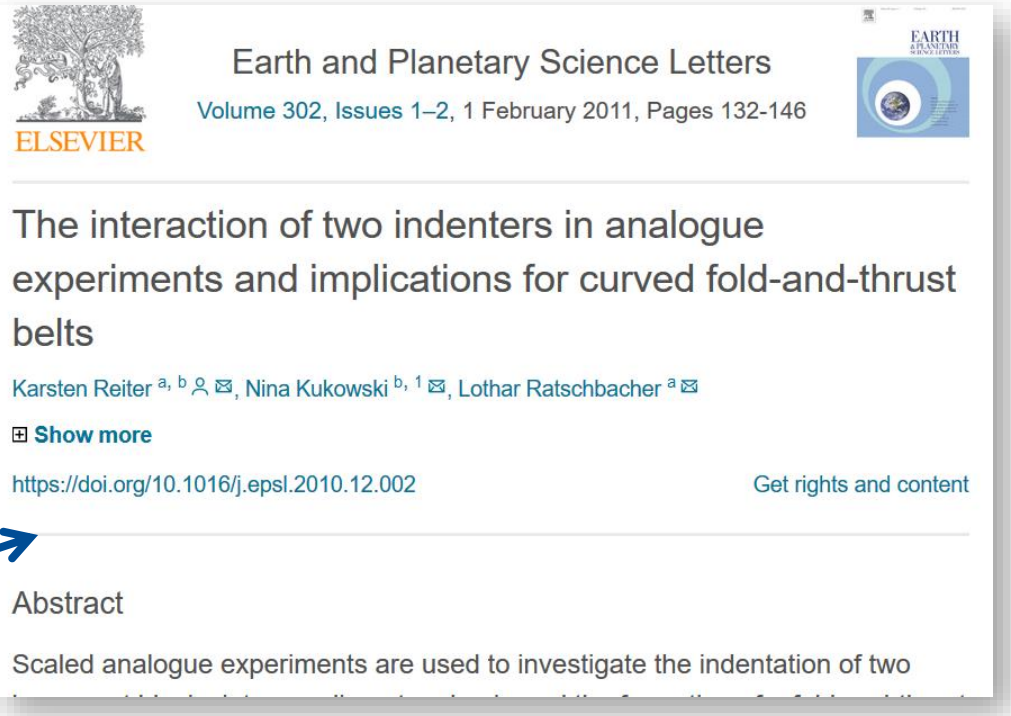
"interoperability framework for exchanging information about the links between scholarly literature and data"

Scholexplorer
The Data Literature Interlinking Service

The classical approach (before Scholix)

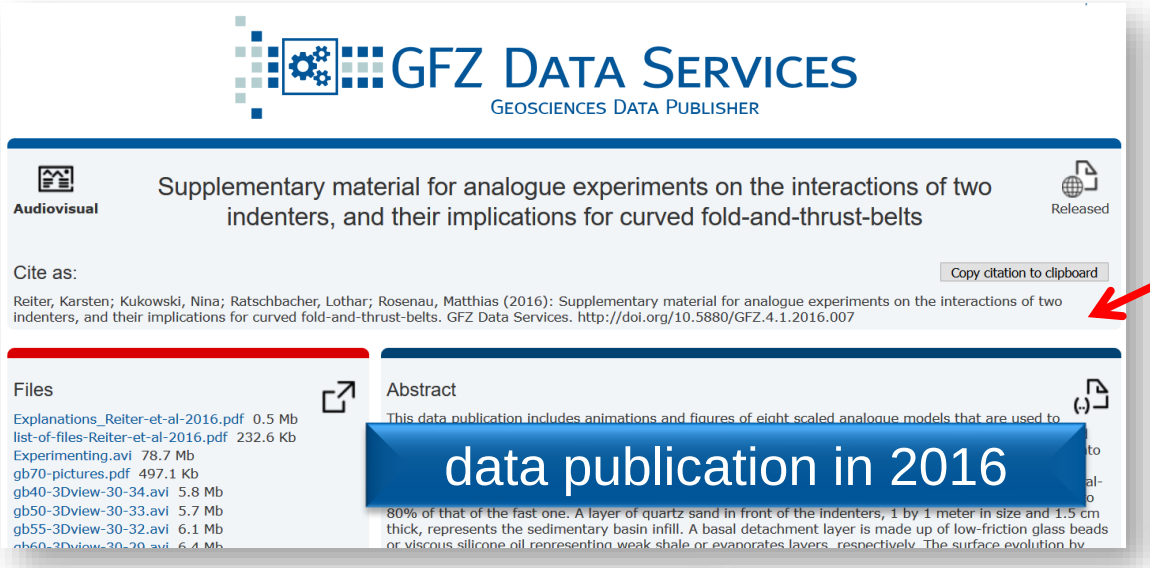


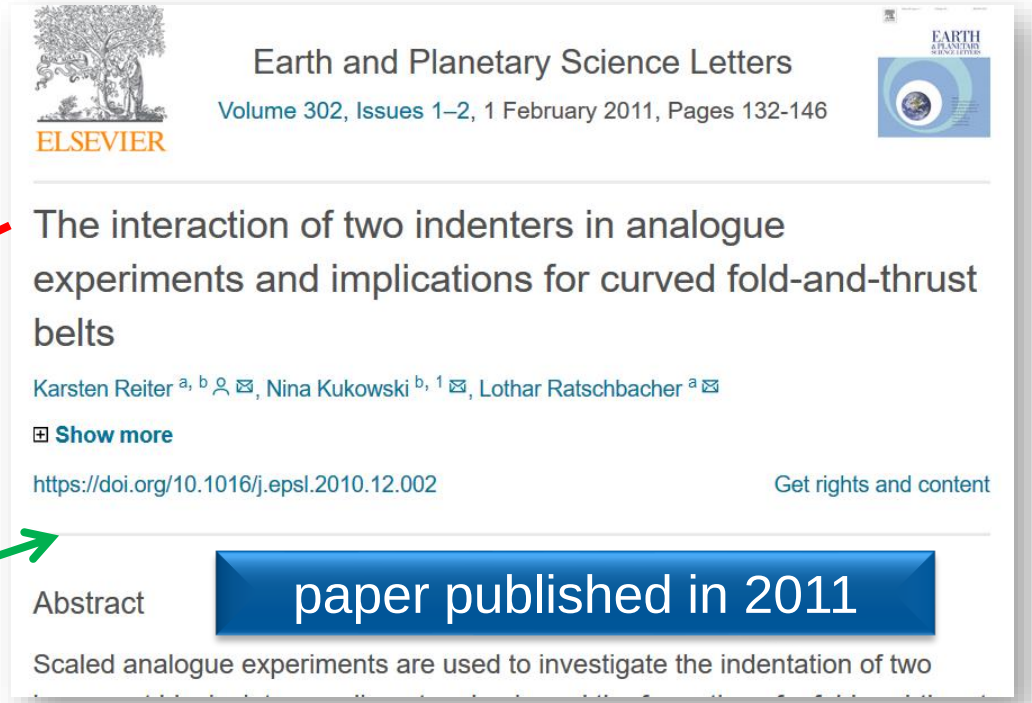
…supplement to an article from 2011



Data: http://doi.org/10.5880/GFZ.4.1.2016.007, Paper: https://doi.org/10.1016/j.epsl.2010.12.002

The classical approach (before Scholix)



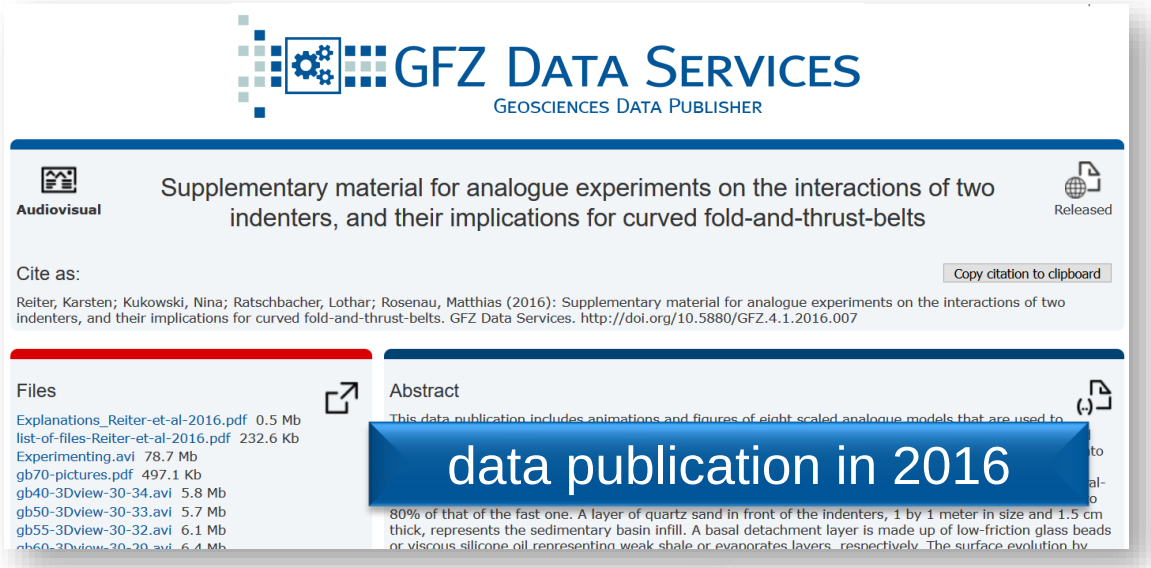


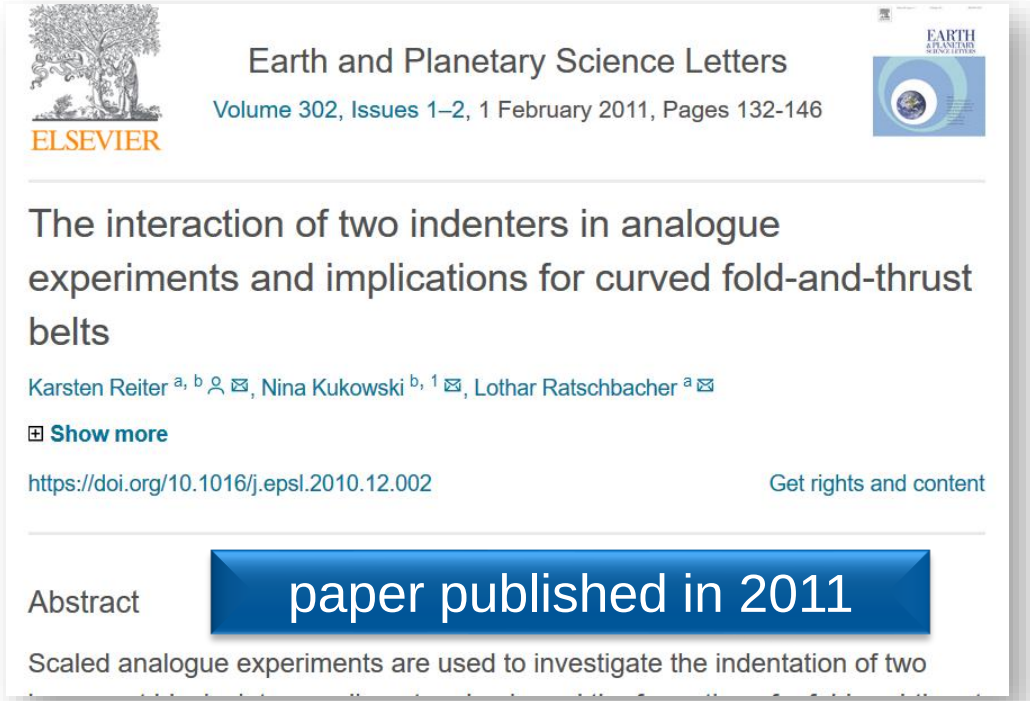
Link to paper in the metadata of the data publication

Is **not** possible to add the citation of the dataset to the paper 5 years after publishing

Data: http://doi.org/10.5880/GFZ.4.1.2016.007, Paper: https://doi.org/10.1016/j.epsl.2010.12.002

New possibilities for cross-linking data and papers





SCHOLIX

Link to data on the paper landing page

GFZ Helmholtz-Zentrum Potsdam

Data: http://doi.org/10.5880/GFZ.4.1.2016.007,
Paper: https://doi.org/10.1016/j.epsl.2010.12.002