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

Kirsten Elger, Boris Radosavljevic, Damian Ulbricht

Source: OpenAIRE
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: Increasing 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.”

• 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)
Metadata Schema for the Description of Research Data Repositories

41 Properties on
- General information
- Responsibilities
- Policies
- Legal aspects
- Technical standards
- Quality standards

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Portal

Filter
Subjects
Content Types
Countries
Data access
Database access
Data licenses
Data upload
Data upload restrictions
Enhanced publication
Institution responsibility type
Institution type
Keywords
Metadata standards
PID systems
Provider types
Quality management
Repository languages
Repository types
Versioning

Nordicana D
Nordicana D collection

Subject(s)
Geosciences (including Geography)
Natural Sciences

Content type(s)
Raw data
Structured graphics
Images
Scientific and statistical data formats
Plain text

Country
Canada

Nordicana series D is a formatted, online data report series archived at CEN. It is produced only in electronic form and is freely and openly accessible to CEN researchers and to other users. Each issue is published in French and in English, and is indexed via an assigned digital object identifier (DOI). An issue may be updated, for example with new data, as a new version number, but will retain the same DOI. Each issue contains data sets and extensive metadata that explain the origin of the data, the format of the data, the history of updates via different version numbers, and the format that should be adopted to cite the data.
The research data repository provides additional information on its service.

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

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

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

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

Icons facilitating the selection process of appropriate research data repositories.
New **DFG** Project:
**re3data COREF**
Community Driven Open Reference for Research Data Repositories

- 2020-2022
- To connect re3data as the reference for research data repositories with other services and infrastructures
Break – Questions?

Next: COPDESS and Enabling FAIR Data
“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.

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.

Properties of granular analogue model materials: A community wide survey

M. Klinkmüller a, G. Schreurs a, M. Rosenau b, H. Kemnitz b

a Institute of Geological Sciences, University of Bern, Bürkistrasse 1 +3, CH-3012 Bern, Switzerland
b Helmholtz-Zentrum Potsdam, GFZ Deutsches Geoforschungszentrum, Telegrafenberg, D-14473 Potsdam, Germany

Presented 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).

How to cite a dataset?

1. Citation in the text
2. Full reference with DOI in the References
3. Data access via DOI

References


Break – Questions?

Next: Data Publications: technical background and best practices
Data Publications - components

- Data
- Metadata
### Types of Metadata

<table>
<thead>
<tr>
<th>Type of Metadata</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive metadata</td>
<td>For finding or understanding a resource</td>
</tr>
<tr>
<td>Administrative metadata</td>
<td>- For decoding and rendering files&lt;br&gt;- Long-term management of files&lt;br&gt;- Intellectual property rights attached to content</td>
</tr>
<tr>
<td>Structural metadata</td>
<td>Relationships of parts of resources to one another</td>
</tr>
<tr>
<td>Markup languages</td>
<td>Integrates metadata and flags for other structural or semantic features within content</td>
</tr>
</tbody>
</table>

**Descriptive Metadata:**
- Metadata for data discovery

**Contextual metadata**

Source: NISO: Understanding Metadata – a Primer
Definition of data labels

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Data Type</th>
<th>Description</th>
<th>Validation Test</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPEDITION</td>
<td>Numeric</td>
<td>expedition number</td>
<td>integer value</td>
<td></td>
</tr>
<tr>
<td>SITE</td>
<td>Integer</td>
<td>site number</td>
<td>integer value</td>
<td></td>
</tr>
<tr>
<td>NAME</td>
<td>Text</td>
<td>site name or locality</td>
<td>string of max. 40 characters</td>
<td></td>
</tr>
<tr>
<td>PLATFORM</td>
<td>Text</td>
<td>platform identifier</td>
<td>string of max. 40 characters</td>
<td></td>
</tr>
<tr>
<td>LATITUDE_XGE</td>
<td>Integer</td>
<td>degrees of site latitude (latitude of hole XGE)</td>
<td>integer value</td>
<td></td>
</tr>
<tr>
<td>LATITUDE_YGE</td>
<td>Integer</td>
<td>degrees of site longitude (longitude of hole XGE)</td>
<td>integer value</td>
<td></td>
</tr>
<tr>
<td>LATITUDE_XGR</td>
<td>Integer</td>
<td>degrees of site latitude (latitude of hole XGR)</td>
<td>integer value</td>
<td></td>
</tr>
<tr>
<td>LATITUDE_YGR</td>
<td>Integer</td>
<td>degrees of site longitude (longitude of hole XGR)</td>
<td>integer value</td>
<td></td>
</tr>
<tr>
<td>LONGITUDE_XGE</td>
<td>Double</td>
<td>decimal minutes of site longitude</td>
<td>integer between 0 and 180</td>
<td>deg.</td>
</tr>
<tr>
<td>LONGITUDE_YGE</td>
<td>Double</td>
<td>decimal minutes of site latitude</td>
<td>integer between 0 and 180</td>
<td>deg.</td>
</tr>
<tr>
<td>LONGITUDE_XGR</td>
<td>Double</td>
<td>decimal minutes of site longitude</td>
<td>integer between 0 and 180</td>
<td>deg.</td>
</tr>
<tr>
<td>LONGITUDE_YGR</td>
<td>Double</td>
<td>decimal minutes of site latitude</td>
<td>integer between 0 and 180</td>
<td>deg.</td>
</tr>
<tr>
<td>Date START</td>
<td>Date</td>
<td>date of site start</td>
<td>date in UTC</td>
<td></td>
</tr>
<tr>
<td>Date END</td>
<td>Date</td>
<td>date of site end</td>
<td>date in UTC</td>
<td></td>
</tr>
<tr>
<td>DESCRIPTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data Articles/Reports

Earth System Science Data

highly variable between the disciplines but key information for data reuse
Metadata for data discovery: example DOI Landing Page

**Title:**
COSC-1 operational report - Operational data sets

**Citation:**

**Description/Abstract:**

The Collision Sciences in Scandinavia (COSC-1) scientific drilling project focuses on the understanding of tectonic processes in a major mid-oceanic ridge in eastern Scandinavia and its consequences for modern analogues. The transport and development of volcanic-related hydrothermal systems occur in the area of the COSC-1 project. This project is located on the Mid-Atlantic Ridge (MAR) between the Iceland-Faroe Ridge and the North Atlantic basin. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothermal systems and their impact on the global climate. The scientific objectives of the COSC-1 project are to study the processes of the MAR and their impact on the global climate. The project aims to investigate the influence of volcanic-related hydrothermal systems on the global climate and the production of greenhouse gases. The COSC-1 project is a cooperative effort involving the participation of scientists from various disciplines. The project involves the study of volcanic-related hydrothal...
Machine-readable metadata standards

- XML metadata in DataCite, ISO19115 standards
- Exchange via OAI-PMH interface
- Machine-readable websites
- JSON LD
- Google Dataset Search
Controlled Vocabularies in the Metadata Editor

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

NASA Global Change Master Directory (GCMD) Keywords

INPRIRE ISO 19115 Keywords

RDF-Version

GeoSciML: Geoscience Vocabularies for Linked Data
Persistent Identifier in data publications

- **DOI**: for data, software, cross-references to related work
- **ORCID**: uniquely identifying persons
- **ROR**: New PID for Institutions
- **Crossref Funder Registry**: List of funders with DOIs
- **IGSN**: PID for physical samples, cross references to samples underlying measurements

and others...
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)

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-1¹, Author-2² ...
- 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 (m$^3$) - Ehyd = Applied hydraulic energy (J) - MaxM0 = Maximum observed seismic moment (Nm) - CumM0 = Cumulative seismic moment (Nm) - IE = Injection efficiency (−)
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

This data set is freely available under a Creative Commons Attribution 4.0 International (CC-BY 4.0) Licence. It is part of the following data publication and should be cited as:
Rosenau, Matthias; Pohlenz, Andre; Kemnitz, Helge; 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

<table>
<thead>
<tr>
<th>ZIP folder</th>
<th>Folder size</th>
<th>File name</th>
<th>File format</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2019-004_Rosenau-etal_Description-of-data</td>
<td>.pdf</td>
<td>Description of data and methods</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RSTanalysis</td>
<td>.py</td>
<td>Python script for analysing and plotting friction and time series data (Mohr plot, histograms, shear curves)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>VSTanalysis</td>
<td>.py</td>
<td>Python script for analysing and plotting VST data</td>
</tr>
<tr>
<td>Data files</td>
<td>2932 kb</td>
<td>421-01_GFZ_quartsandG23_vst</td>
<td>.pdf</td>
<td>Visualization of the VST data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>421-01_GFZ_quartsandG23_vst</td>
<td>.txt</td>
<td>Table of the VST data: time, shear velocity, normal force, shear force</td>
</tr>
<tr>
<td></td>
<td></td>
<td>426-01_GFZ_quartsandG23_dynamic</td>
<td>.txt</td>
<td>Pairs of normal stress and corresponding shear strength for dynamic friction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>426-01_GFZ_quartsandG23_hist</td>
<td>.pdf</td>
<td>Histograms of friction coefficients and cohesions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>426-01_GFZ_quartsandG23_lineragr</td>
<td>.pdf</td>
<td>Mohr plot of friction data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>426-01_GFZ_quartsandG23_peak</td>
<td>.txt</td>
<td>Pairs of normal stress and corresponding shear strength for peak friction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>426-01_GFZ_quartsandG23_reactivation</td>
<td>.txt</td>
<td>Pairs of normal stress and corresponding shear strength for reactivation friction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>426-01_GFZ_quartsandG23_ts</td>
<td>.pdf</td>
<td>Visualization of time series data (shear curves): Shear stress vs. shear displacement for 18 measurements</td>
</tr>
<tr>
<td></td>
<td></td>
<td>426-01_GFZ_quartsandG23_ts</td>
<td>.txt</td>
<td>Table of time series data for 18 measurements of shear stress (Pa, columns 2-19) at given normal stresses (Pa, first cell in each column) vs. time (column 1)</td>
</tr>
</tbody>
</table>
## Definition of data tables

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

<table>
<thead>
<tr>
<th>Column header</th>
<th>unit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SampleID</td>
<td></td>
<td>Sample Identifier</td>
</tr>
<tr>
<td>Lat</td>
<td>DD,dddd</td>
<td>Latitude in WGS84 in decimal degrees</td>
</tr>
<tr>
<td>Long</td>
<td>DD,dddd</td>
<td>Longitude in WGS84 in decimal degrees</td>
</tr>
<tr>
<td>MeasurementXY</td>
<td>xy</td>
<td>Value of XY in units xy</td>
</tr>
</tbody>
</table>
References for dataset and those cited above. If the data follows a standard format defined by the community, please cite the standard. Does this data incorporate other authors’ datasets? Citations should contain a DOI (or URL for reports or websites). The DOI numbers shall be provided within https:// code facilitating to directly link to the cited paper (see Figure 3).

8. References

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)

<table>
<thead>
<tr>
<th>sample ID</th>
<th>IGSN†</th>
<th>brief sample description</th>
<th>mean depth (m)</th>
<th>SiO₂ (wt%)</th>
<th>TiO₂ (wt%)</th>
<th>Al₂O₃ (wt%)</th>
<th>Fe₂O₃ (wt%)</th>
<th>MnO (wt%)</th>
<th>MgO (wt%)</th>
<th>CaO (wt%)</th>
<th>Na₂O (wt%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON 14</td>
<td>GFDUH00LT</td>
<td>soil, Ah horizon</td>
<td>0.2</td>
<td>57</td>
<td>0.86</td>
<td>15</td>
<td>6.4</td>
<td>0.12</td>
<td>2.0</td>
<td>0.19</td>
<td>0.87</td>
</tr>
<tr>
<td>CON 13</td>
<td>GFDUH00LU</td>
<td>soil, Ah/Bw horizon</td>
<td>0.4</td>
<td>60</td>
<td>0.89</td>
<td>16</td>
<td>6.3</td>
<td>0.13</td>
<td>1.9</td>
<td>0.19</td>
<td>0.93</td>
</tr>
<tr>
<td>CON 12</td>
<td>GFDUH00LV</td>
<td>soil, Bw1 horizon</td>
<td>0.6</td>
<td>51</td>
<td>0.79</td>
<td>14</td>
<td>5.5</td>
<td>0.12</td>
<td>1.6</td>
<td>0.16</td>
<td>0.87</td>
</tr>
<tr>
<td>CON 11</td>
<td>GFDUH00LW</td>
<td>soil, Bw1 horizon</td>
<td>0.8</td>
<td>58</td>
<td>0.93</td>
<td>18</td>
<td>7.4</td>
<td>0.15</td>
<td>2.1</td>
<td>0.20</td>
<td>1.0</td>
</tr>
<tr>
<td>CON 10</td>
<td>GFDUH00LX</td>
<td>soil, Bw1 horizon</td>
<td>1.0</td>
<td>59</td>
<td>0.94</td>
<td>18</td>
<td>7.3</td>
<td>0.13</td>
<td>2.2</td>
<td>0.24</td>
<td>1.0</td>
</tr>
<tr>
<td>CON 9</td>
<td>GFDUH00LY</td>
<td>soil, Bw1 horizon</td>
<td>1.2</td>
<td>62</td>
<td>0.94</td>
<td>17</td>
<td>6.9</td>
<td>0.08</td>
<td>2.3</td>
<td>0.28</td>
<td>1.1</td>
</tr>
</tbody>
</table>
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
# 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:


## Machine Readable Tables

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

- 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: Increasing the transparency of research products
Linking papers, data, samples, ...

DataCite related Identifier

Scientific Paper

Data Report

Sample

COSC-1 Drilling of a Subsection-Related Allochthon in the Palaeozoic Caledonide Orogen of Scandinavia


Majka, J., Rosen, A., Janák, M., Fritzsche, N., Klomowska, I., Mamecki, M., Sasinová, V., and Yoshida, K.: Microdiamond di...
Scholix: linking data with papers

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

Credit: Wouter Haak, Elsevier
The classical approach (before Scholix)

The classical approach (before Scholix)


Is not possible to add the citation of the dataset to the paper 5 years after publishing.
New possibilities for cross-linking data and papers

Data: http://doi.org/10.5880/GFZ.4.1.2016.007