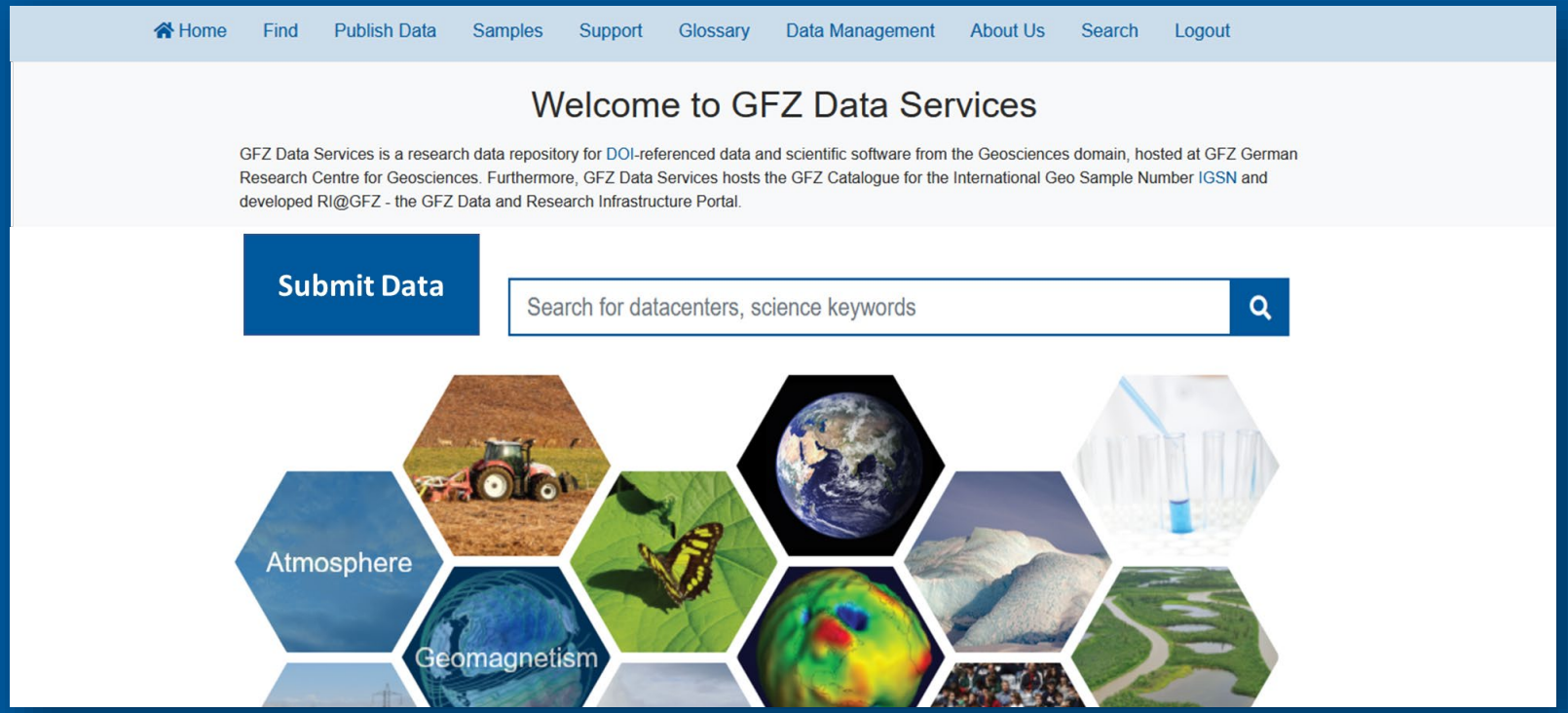
Introduction to data publications and IGSN



Kirsten Elger

FID GEO

E-PUBLISHING RESEARCH DATA DIGITISATION

Specialised Information Service for Geosciences

promoting Open Science in Geosciences

Electronic publishing of institutional literature not released in publishing houses as well as pre- and postprints of research articles.

Electronic publishing of curated research data via a domain repository for the Geosciences.

Digitisation on demand of literature and maps in the public domain, out of print, or on behalf of the publishing institutions or societies.

FID GEO @ GFZ

Melanie Lorenz

Kirsten Elger

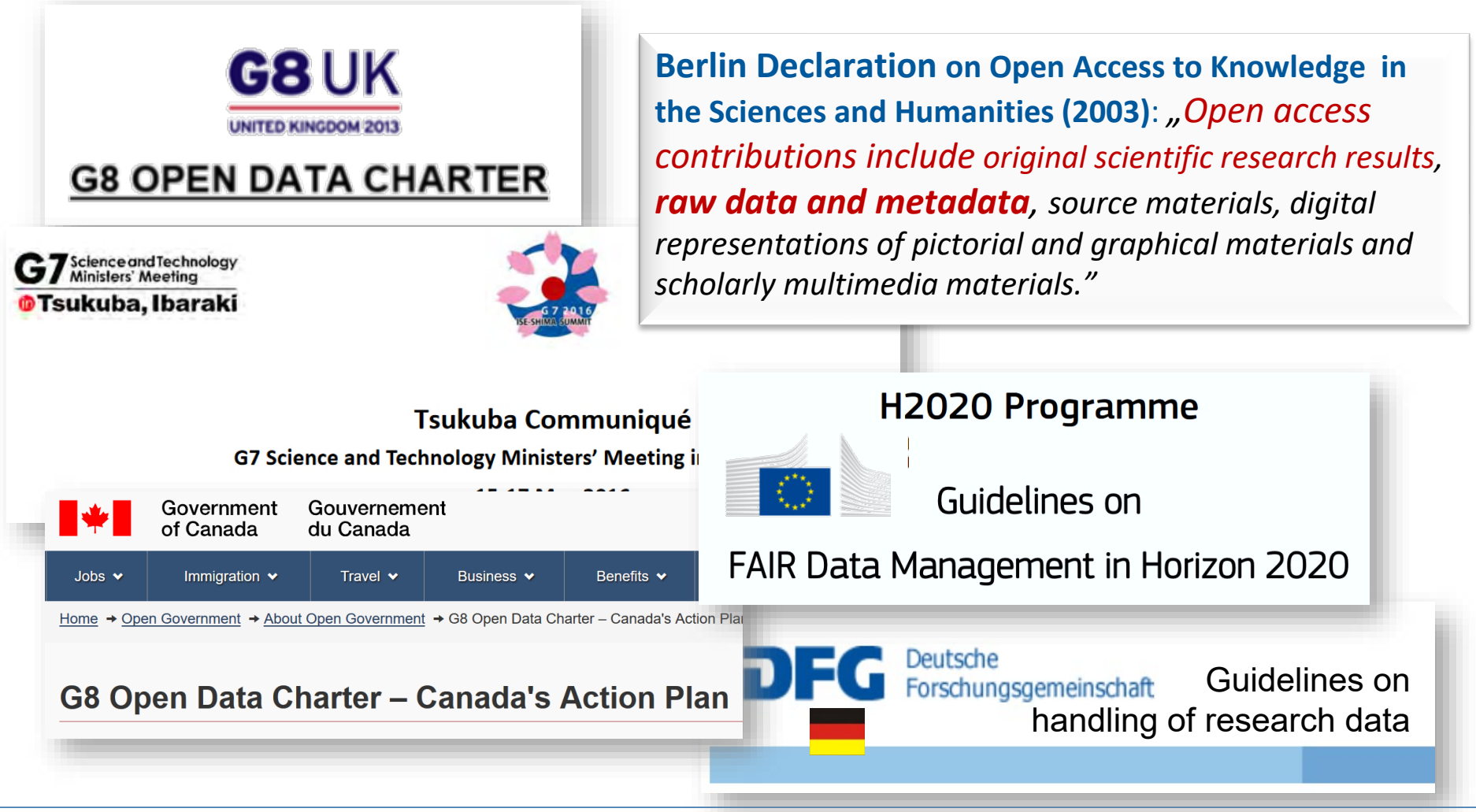
Marcel Meistring

SUB | NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

GFZ Helmholtz Centre Potsdam

Funded by: DFG

Open data – an international request



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

FAIR Principles

Findable

Accessible

Interoperable

Reusable

*Wilkinson et al., 2016 https://doi.org/10.1038/sdata.2016.18

Handling of Research Data

→ DFG promotes Open Access papers, open data, open source software

- **Research data includes** measurement data, laboratory values, audiovisual information, texts, survey or observation data, methodological test procedures and questionnaires. Compilations and simulations can likewise constitute a key outcome of academic research and are therefore also included under the term research data
- **DFG expects research projects to include a description of how research data is handled**. The description should be based on the checklist for handling research data. **The recommendation is that contact should be established as early as possible** during the project planning phase **with a research data centre or repository** where the research data can be deposited.
- **Making research data available**, developing methods and standards and building data infrastructures **are important contributions to the re-use of research findings as well as integrated part of the good research practice** and should be listed as part of a researcher's preliminary work or academic profile.

Publisher/Journal requirements for Open Science

SPRINGER NATURE

Reporting standards and availability of data, materials, code and protocols

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. A condition of publication in a Nature Portfolio journal is that **authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications**. Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any **restrictions must** also **be disclosed** in the submitted manuscript.

AGU ADVANCING EARTH AND SPACE SCIENCE

AGU requires that the underlying data needed to understand, evaluate, and build upon the reported research be available at the time of peer review and publication. Additionally, authors should make available software that has a significant impact on the research. This entails:

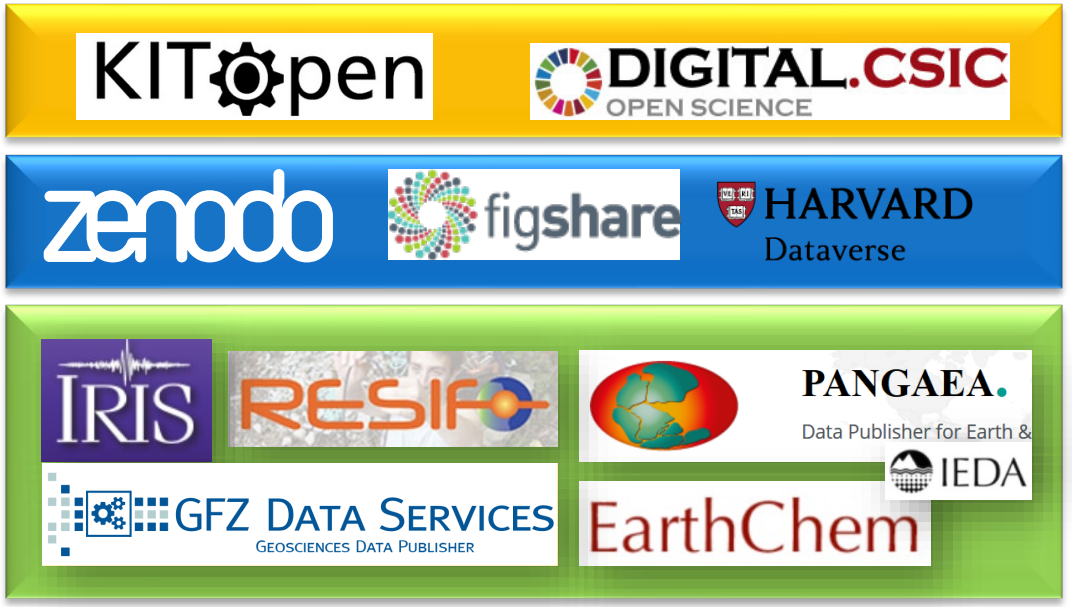
1. Depositing the data and software in a community accepted, trusted repository, as appropriate, and preferably with a DOI
2. Including an Availability Statement as a separate paragraph in the Open Research section explaining to the reader where and how to access the data and software
3. And including citation(s) to the deposited data and software, in the Reference Section.

Data Publications – best practice for FAIR sharing data

Publication of datasets as individual publications (with assigned persistent Identifier, e.g., DOI) **through research data repositories**

Research Data Repositories

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



"***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, https://doi.org/10.1029/2015EO022207

Data Publications – best practice for FAIR sharing data

Publication of datasets as individual publications (with assigned persistent Identifier, e.g., DOI) **through domain 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 → curation
- **Citable**: DOI-referenced datasets are citable just as journal articles (→ credit for researcher and institution)

How do I 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

What do I need for a data publication?

- Data
- Metadata

Contextual Metadata

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 from

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

highly variable between the disciplines but key information for data reuse

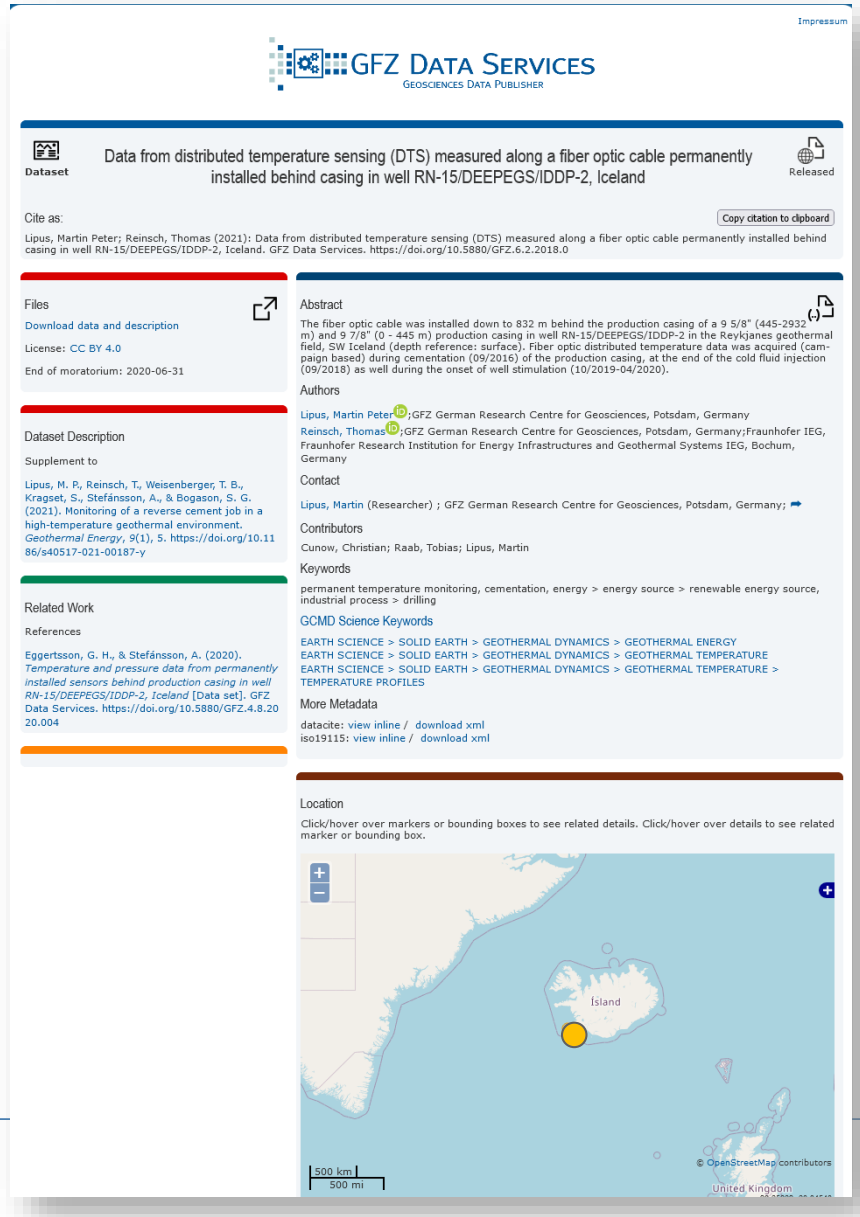
Metadata for data discovery: example DOI Landing Page

title citation

download data

key paper

related work



description/abstract

authors/ ORCID

keywords/control-led vocabularies

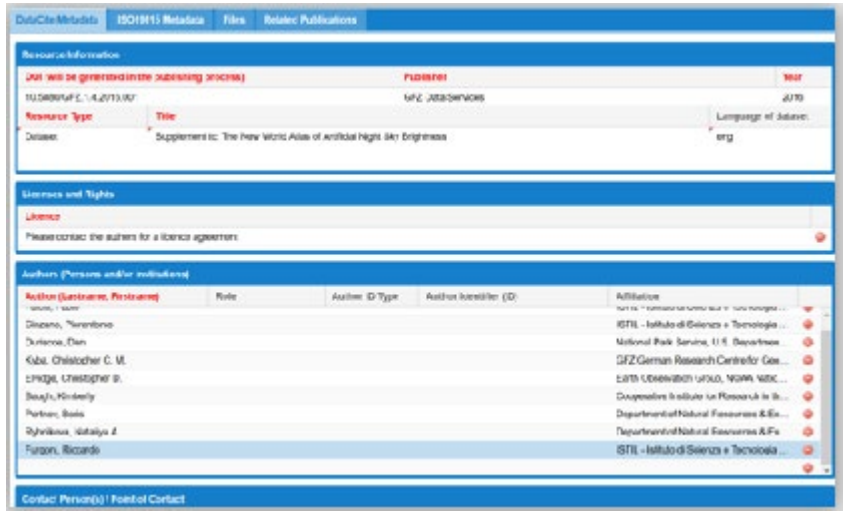
spatial coverage

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

Typical metdata standards for data discovery: **DataCite, ISO19115, Dublin Core**

Tools for data publications by GFZ Data Services

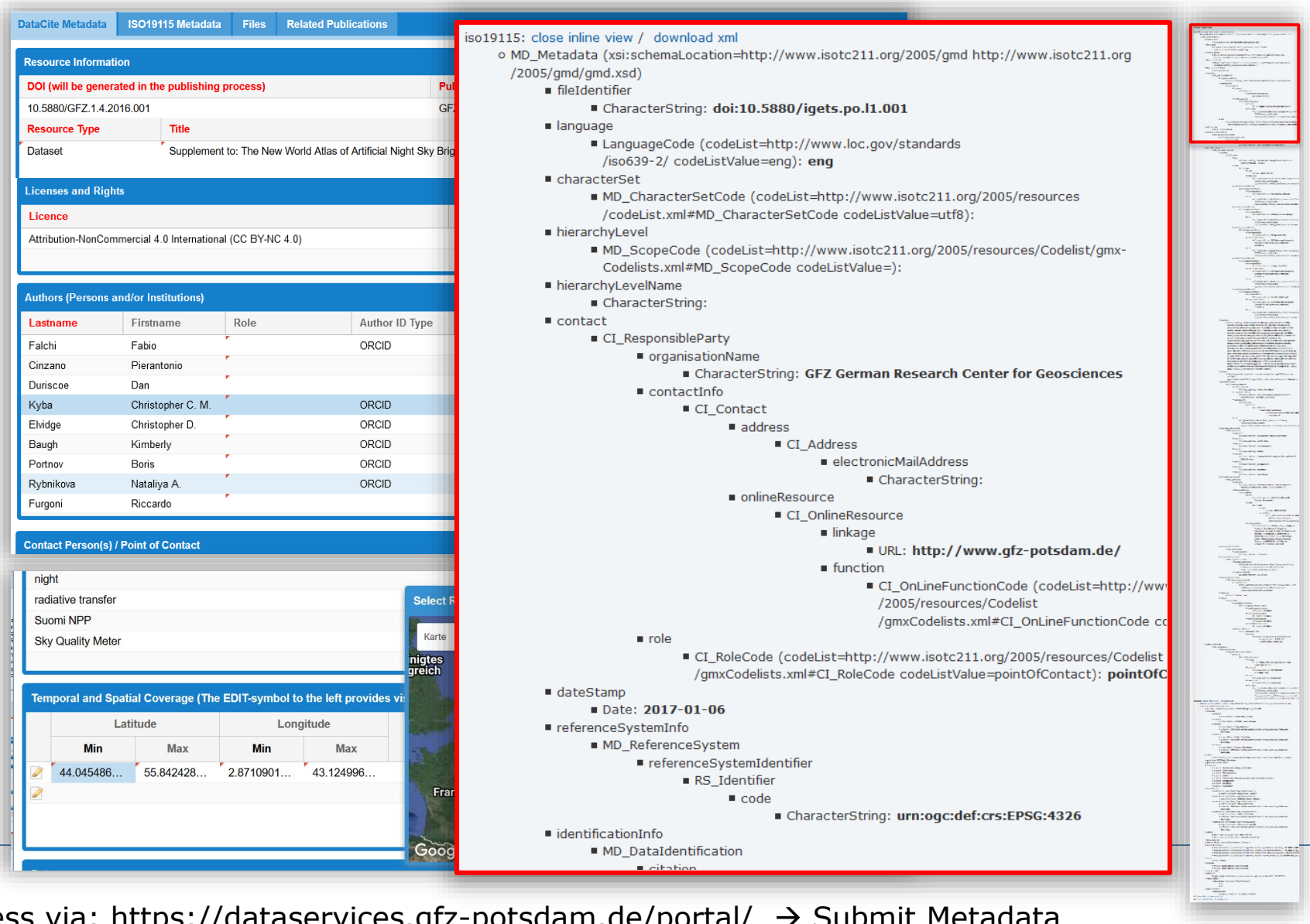
1. Discovery Metadata: via GFZ Metadata Editor



GFZ Metadata Editor (Java Script „translator“)

Input: provided by researchers

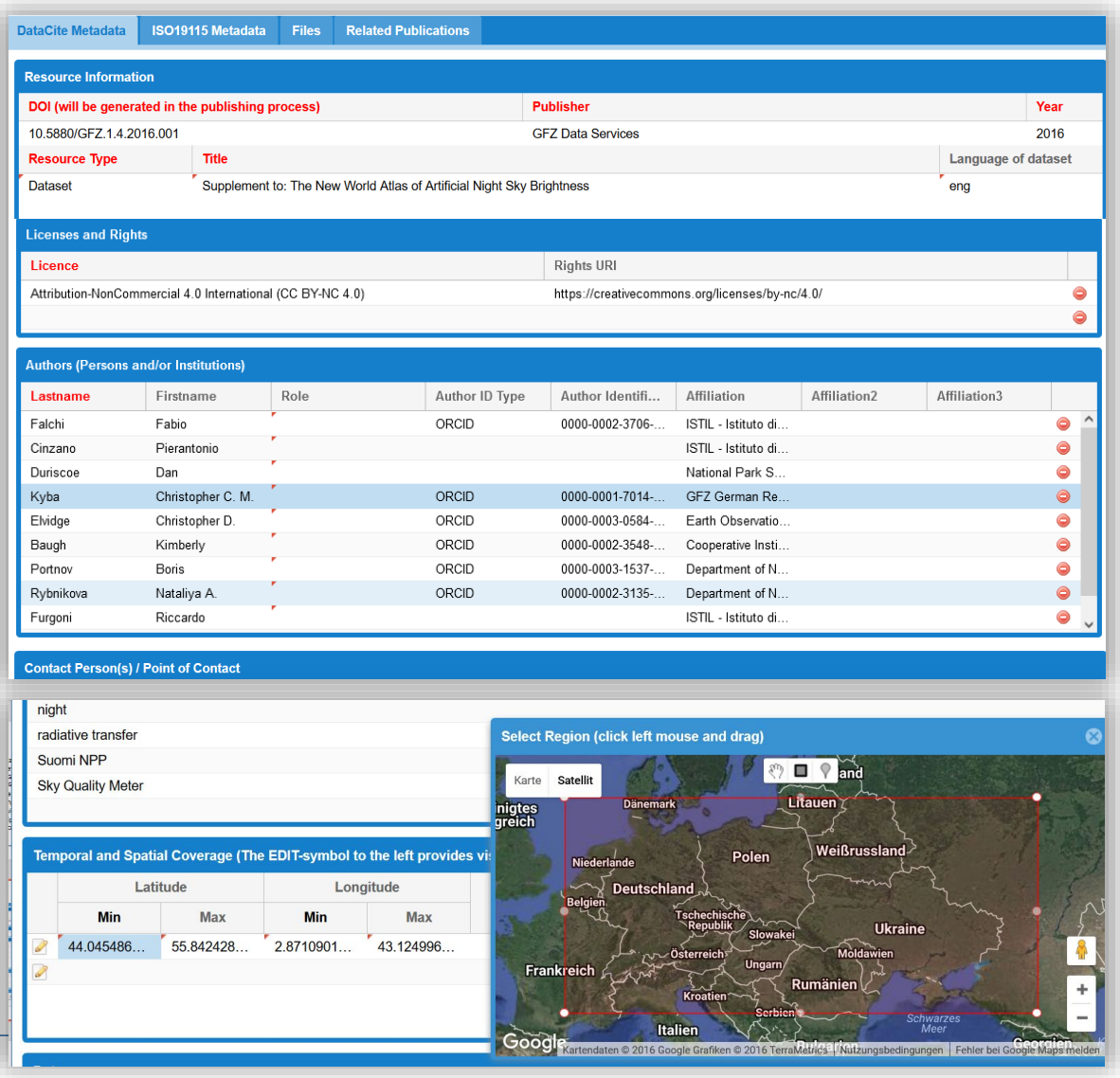
Output: XML (Extensible Markup Language): Metadata exchange format



Access via: https://dataservices.gfz-potsdam.de/portal/ → Submit Metadata

GFZ Metadata Editor (Java Script „translator“)

Input:
provided by researchers



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

DOI Landing Pages

Data Catalogue

Standardised API

„Special“ Features:

- Interactive map
- Searchable vocabulary lists

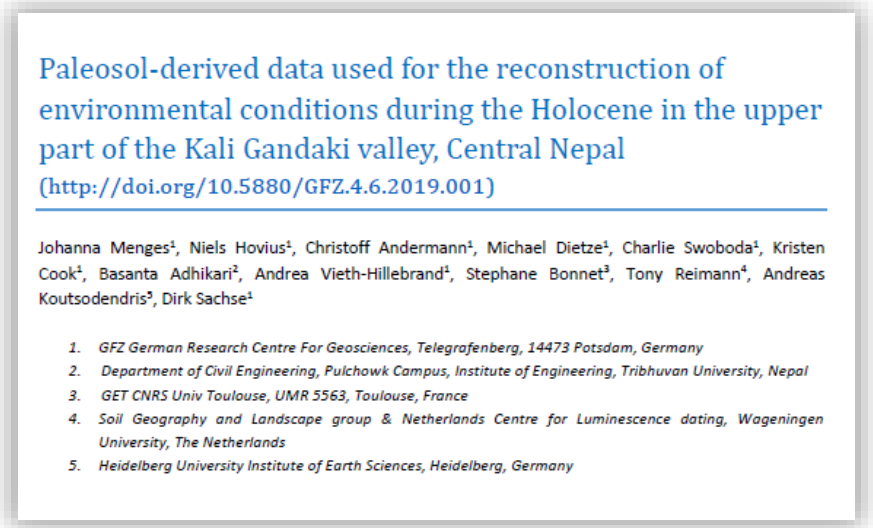
GFZ Helmholtz-Zentrum Potsdam

Access via: https://dataservices.gfz-potsdam.de/portal/ → Submit Metadata

Tools for data publications by GFZ Data Services

1. Discovery Metadata: via GFZ Metadata Editor
2. Contextual Metadata: via Data Description Templates

   (or data reports)



Paleosol-derived data used for the reconstruction of environmental conditions during the Holocene in the upper part of the Kali Gandaki valley, Central Nepal
(http://doi.org/10.5880/GFZ.4.6.2019.001)

Johanna Menges[1], Niels Hovius[1], Christoff Andermann[1], Michael Dietze[1], Charlie Swoboda[1], Kristen Cook[1], Basanta Adhikari[2], Andrea Vieth-Hillebrand[1], Stephane Bonnet[3], Tony Reimann[4], Andreas Koutsodendris[5], Dirk Sachse[1]

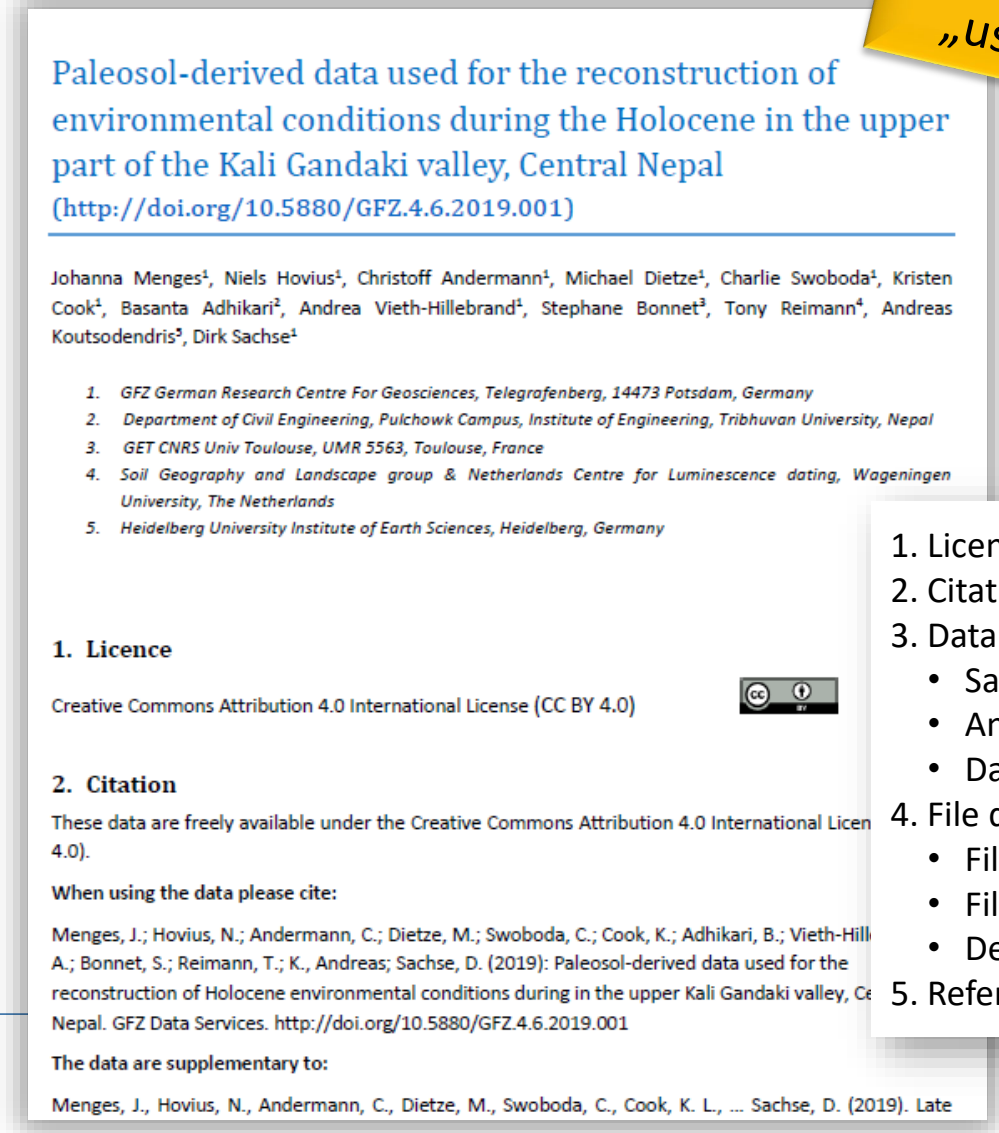
1. *GFZ German Research Centre For Geosciences, Telegrafenberg, 14473 Potsdam, Germany*
2. *Department of Civil Engineering, Pulchowk Campus, Institute of Engineering, Tribhuvan University, Nepal*
3. *GET CNRS Univ Toulouse, UMR 5563, Toulouse, France*
4. *Soil Geography and Landscape group & Netherlands Centre for Luminescence dating, Wageningen University, The Netherlands*
5. *Heidelberg University Institute of Earth Sciences, Heidelberg, Germany*

Data Description Templates

Available in „commented“ and „usable“ form

- 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

➢ Template via gfzpublic (https://gfzpublic.gfz-potsdam.de/pubman/item/item_5007103)



Paleosol-derived data used for the reconstruction of environmental conditions during the Holocene in the upper part of the Kali Gandaki valley, Central Nepal
(http://doi.org/10.5880/GFZ.4.6.2019.001)

Johanna Menges[1], Niels Hovius[1], Christoff Andermann[1], Michael Dietze[1], Charlie Swoboda[1], Kristen Cook[1], Basanta Adhikari[2], Andrea Vieth-Hillebrand[1], Stephane Bonnet[3], Tony Reimann[4], Andreas Koutsodendris[5], Dirk Sachse[1]

1. *GFZ German Research Centre For Geosciences, Telegrafenberg, 14473 Potsdam, Germany*
2. *Department of Civil Engineering, Pulchowk Campus, Institute of Engineering, Tribhuvan University, Nepal*
3. *GET CNRS Univ Toulouse, UMR 5563, Toulouse, France*
4. *Soil Geography and Landscape group & Netherlands Centre for Luminescence dating, Wageningen University, The Netherlands*
5. *Heidelberg University Institute of Earth Sciences, Heidelberg, Germany*

1. Licence

Creative Commons Attribution 4.0 International License (CC BY 4.0)

2. Citation

These data are freely available under the Creative Commons Attribution 4.0 International Licen 4.0).

When using the data please cite:

Menges, J.; Hovius, N.; Andermann, C.; Dietze, M.; Swoboda, C.; Cook, K.; Adhikari, B.; Vieth-Hill A.; Bonnet, S.; Reimann, T.; K., Andreas; Sachse, D. (2019): Paleosol-derived data used for the reconstruction of Holocene environmental conditions during in the upper Kali Gandaki valley, Ce Nepal. GFZ Data Services. http://doi.org/10.5880/GFZ.4.6.2019.001

The data are supplementary to:

Menges, J., Hovius, N., Andermann, C., Dietze, M., Swoboda, C., Cook, K. L., ... Sachse, D. (2019). Late

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

Tools for data publications by GFZ Data Services

1. **Discovery Metadata**: via GFZ Metadata Editor
2. **Contextual Metadata**: via Data Description Templates (or data reports)
3. **Data Discovery and access** via the Data Portal

https://dataservices.gfz-potsdam.de

GFZ Data Services

Profile

- Domain repository for the Geosciences since 2006
- DOIs for Data and software
- Data: real-time data streams, tables, maps, model data, …
- Online metadata editor
- Data description templates
- **Data curation by domain scientists**

https://dataservices.gfz-potsdam.de

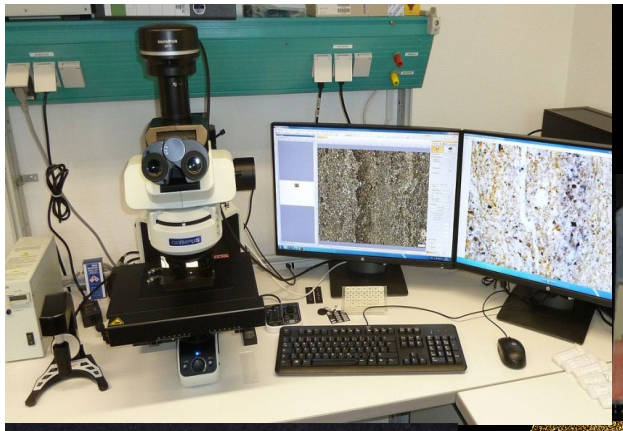
FAIR data

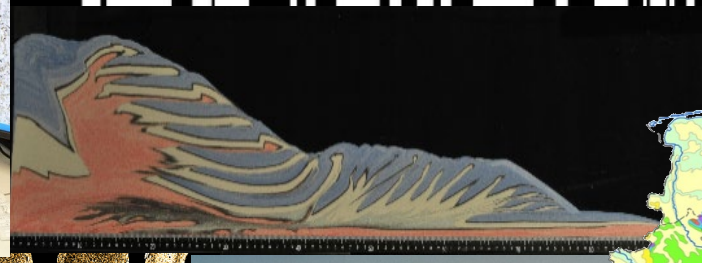
- International metadata standards (human & machine readable)
- Controlled vocabularies for „rich“ metadata
- PIDs ORCID Connecting Research and Researchers, IGSN, Crossref Funder Registry, doi, ROR in prep
- Open Licences for data and software
- OAI-PMH interface
- schema.org → Google Dataset Search

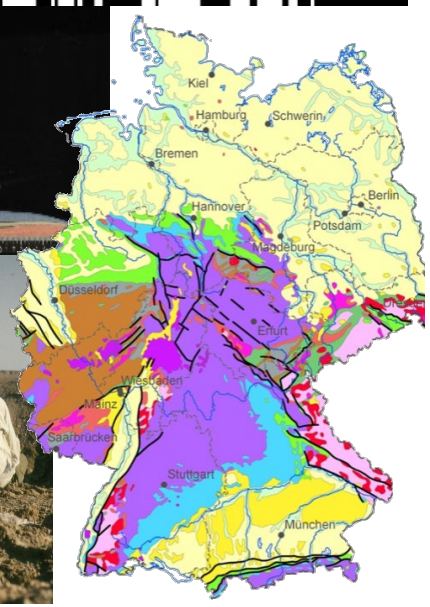
GFZ Data Services: Profile

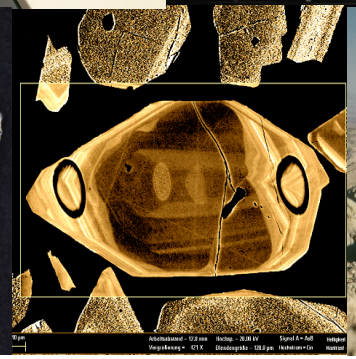
Focus:

1. **curation of long-tail data**

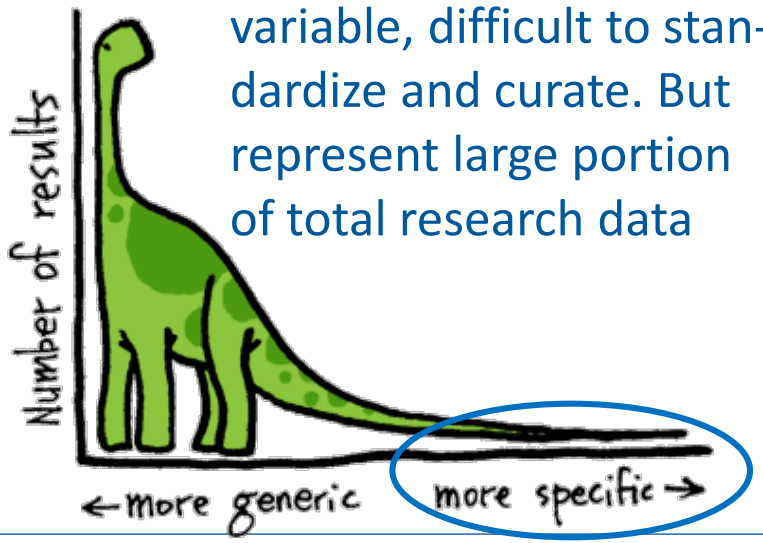








long-tail data: small in size, highly variable, difficult to standardize and curate. But represent large portion of total research data

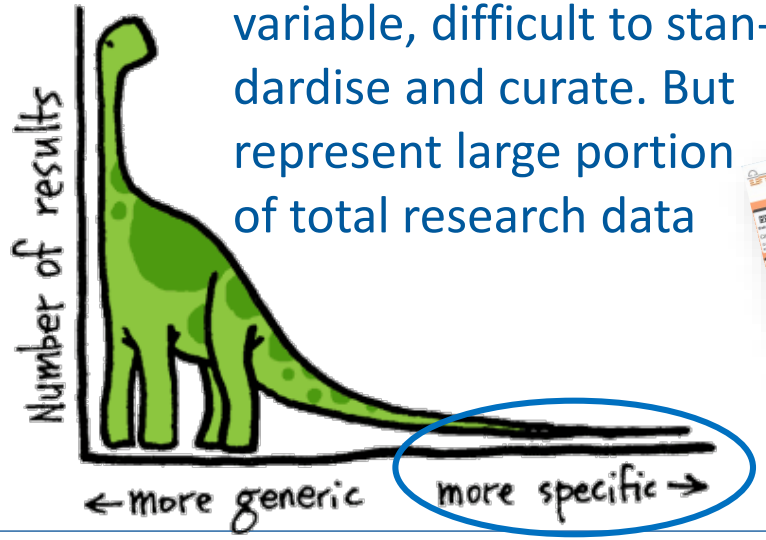


GFZ Data Services: Profile

Focus:

1. **curation of long-tail data**
2. **DOI minting services** for global monitoring networks/observatories in geodesy and geophysics and collaborative projects.

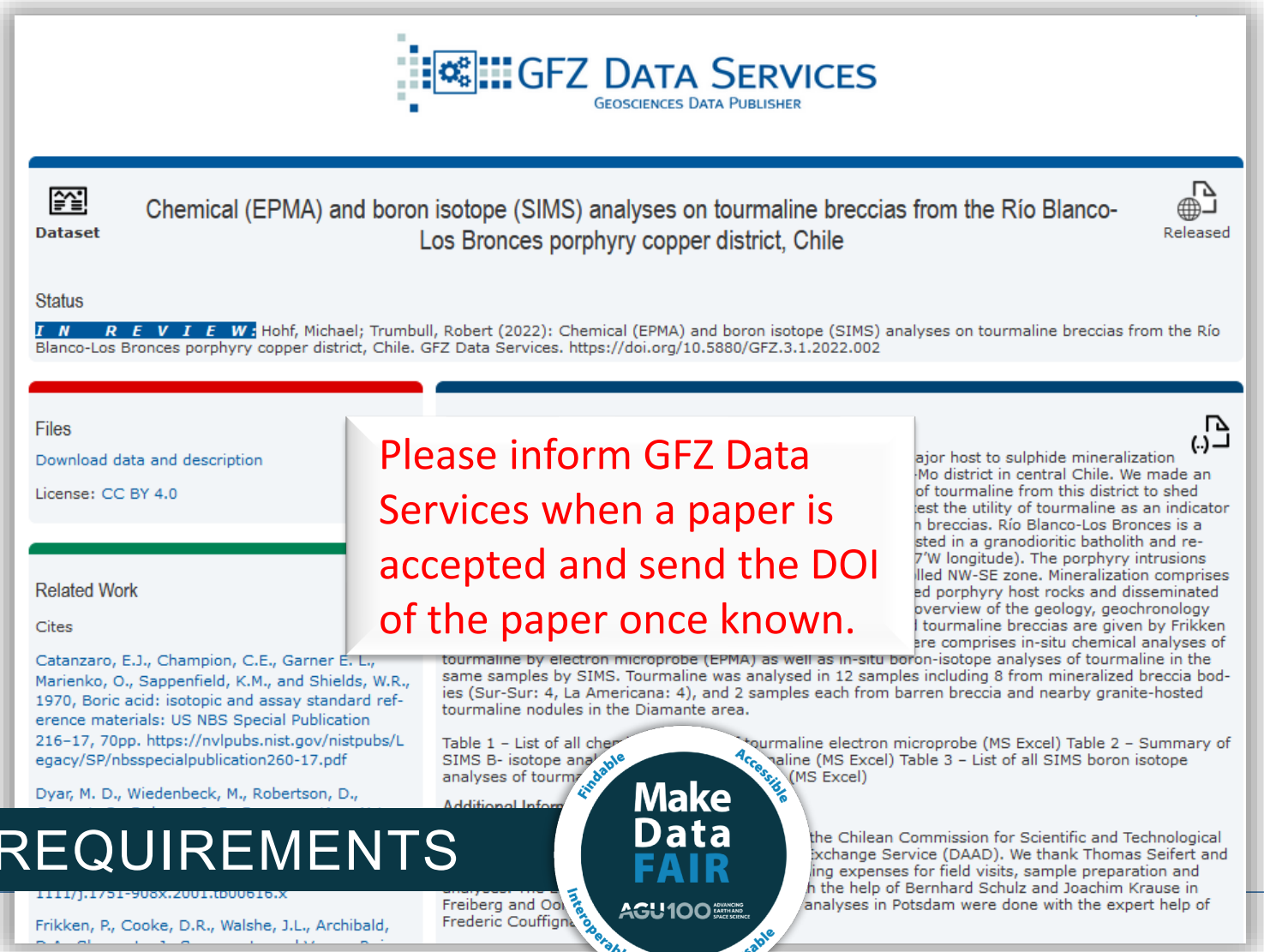
long-tail data: small in size, highly variable, difficult to standardise and curate. But represent large portion of total research data



Different layouts for DOI Landing Pages

Special Features: „Data in Review“ Links

- Link: https://dataservices.gfz-potsdam.de/panmetaworks/review/9c5de649b6b30c588f9fecad56a1c71dd56d1fb4f68ada89b9340002ff84abb7/
- Allows access to still unregistered data (for review purposes)
- DOI is reserved and citable
- Data can still be changed
- DOI registration when paper is accepted

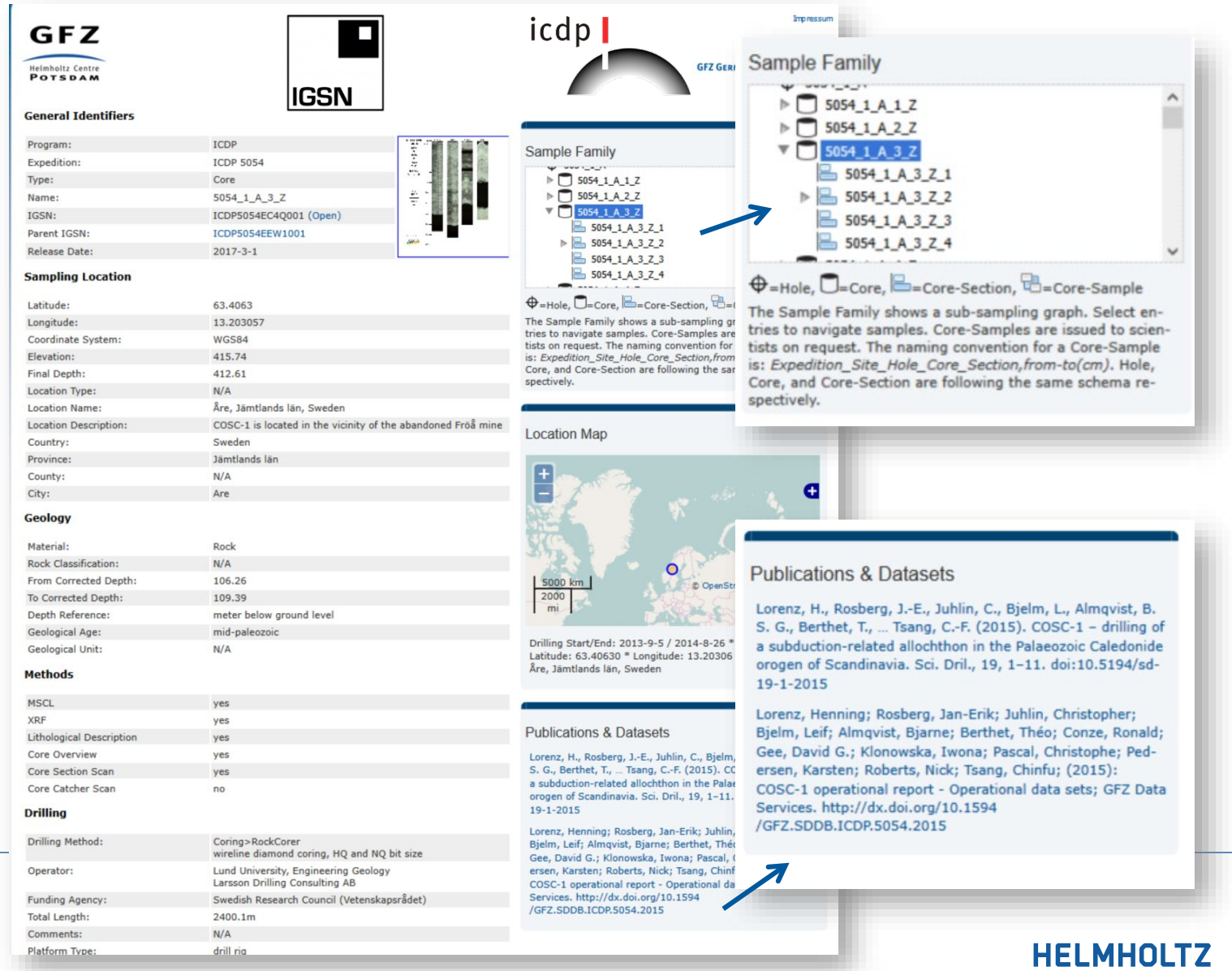


JOURNAL REQUIREMENTS

IGSN – International Generic Sample Number

- Globally unique identifier for physical samples and materials
- Closing the last gap for the full provenance of research results
- IGSN links to the online sample description (link, QR Code)

- For individual and hierarchical samples (e.g. drilling projects)
- IGSN are citable in papers and data publications



Persistent Identifier (PID) in data publications

for data, software, cross-references to related work

https://doi.org/10.5880/fidgeo.2021.049 (Data)

PID for physical samples, cross references to samples underlying measurements

https://igsn.org/GFFJH00AD (Rock sample)

uniquely identifying persons

https://orcid.org/0000-0002-1890-3940 (Max Wilke)

List of funders with DOIs

http://doi.org/10.13039/501100001659 (DFG)

New PID for Institutions

https://ror.org/03bnmw459 (Potsdam University)

→ PIDs ARE RESOLVABLE AND MACHINE-READABLE

PIDs and the provenance of research outcome

data

Table S3. Sierra Nevada analyses of plant samples

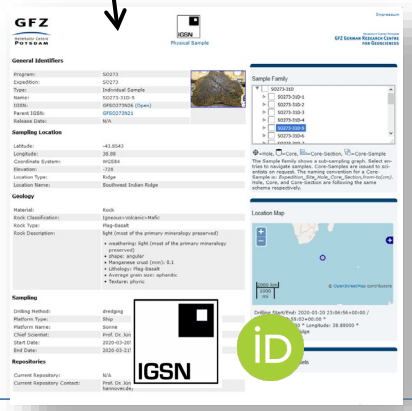
sample ID	IGSN	brief sample description	Element concentrations (µg/g) Al	Fe	Mn	Mg	Ca

Table SN1. Sierra Nevada analyses of soil, saprolite, rock, bedload sediment a

sample ID		IGSN	sample type	XRF lab	depth (cm)	Major element oxides (wt%) (SiO_2	TiO_2	Al_2O_3	Fe_2O_3
P301 regolith depth profile									
SN01	*	GFFB1002T	bulk soil	GFZ	7	36.2	0.55	11.4	3.88
SN02	*	GFFB1002U	bulk soil	GFZ	20	49.3	0.71	14.9	4.84
SN02c	*	GFFB1002U	exchangeable soil	-	20	n.a.	<lod	0.00	0.00
SN02r	*	GFFB1002U	residuum soil	-	20	n.a.	0.84	14.3	5.68
SN03	*	GFFB1002V	bulk soil	GFZ	30	57.5	0.88	17.6	6.44
SN04	*	GFFB1002R	bulk soil	GFZ	39	56.6	0.93	17.7	6.85
SN04e	*	GFFB1002R	exchangeable soil	-	39	n.a.	<lod	0.00	0.00
SN04r	*	GFFB1002R	residuum soil	-	39	n.a.	0.91	15.2	6.95

papers

sample



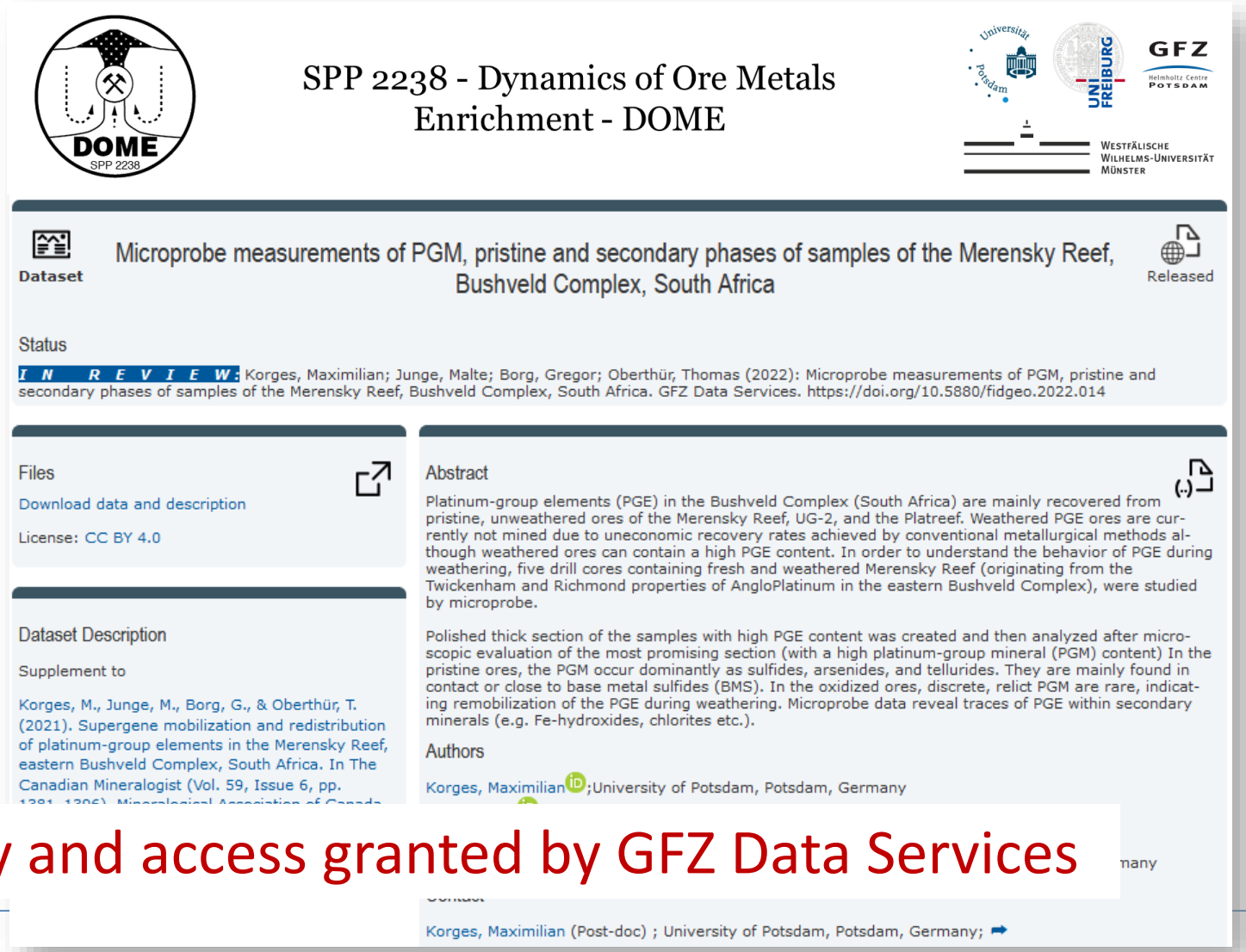
Sample description

Data publication

PIDs connect everything → FAIR

DOME and GFZ Data Services?

- DFG Requirement: Open Science
- Data publications timely correlated with Papers
- DOME Datacentre and Layout
- Examples for each data type (data description templates and meta-data element)
- Data tables aligned with intern. standards (OneGeochemistry, Georoc, EarthChem)



→ Long-term data availablity and access granted by GFZ Data Services