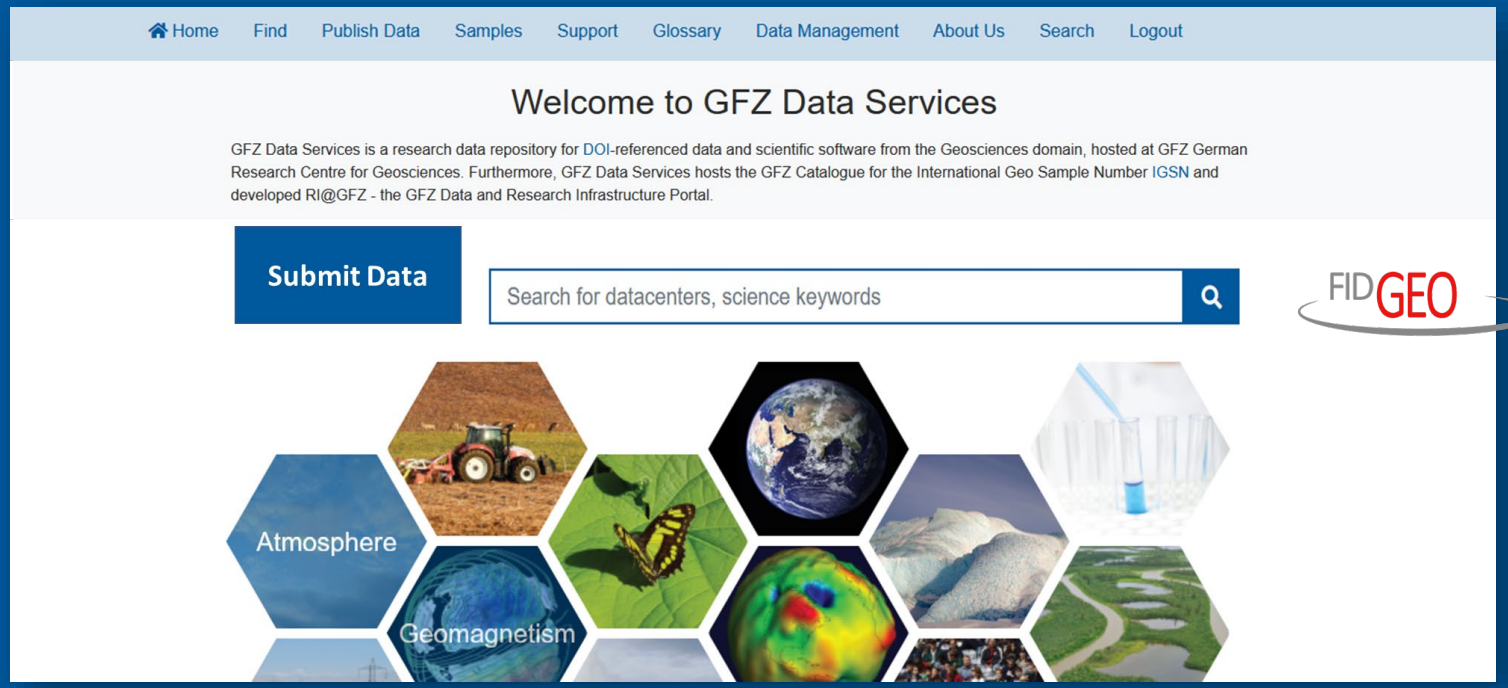
GFZ Data Services – A domain repository for geosciences data



Florian Ott, Kirsten Elger, Simone Frenzel
GFZ German Research Centre for Geosciences, Potsdam, Germany

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)

GFZ Data Services: Research Data Repository

Profile

- Domain repository for the Geosciences since 2006
- DOIs for Data and software
- Data: real-time data streams, tables, maps, model data, …
- **Data curation by domain scientists**
- Partner of: FIDGEO

https://dataservices.gfz-potsdam.de

FAIR data

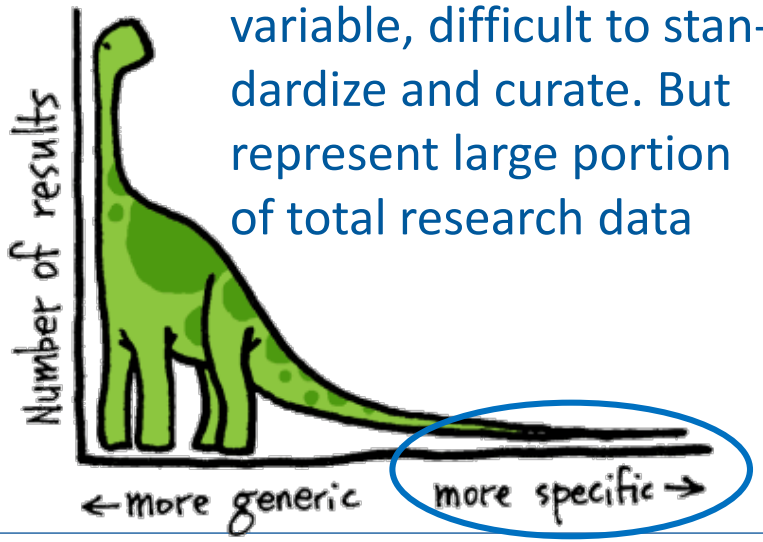
- International metadata standards (human & machine readable)
- Controlled vocabularies for „rich“ metadata
- PIDs ORCID Connecting Research and Researchers, IGSN, Crossref Funder Registry, doi, ROR in prep
- Open Licences for data and software
- OAI-PMH interface
- schema.org → Google Dataset Search

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

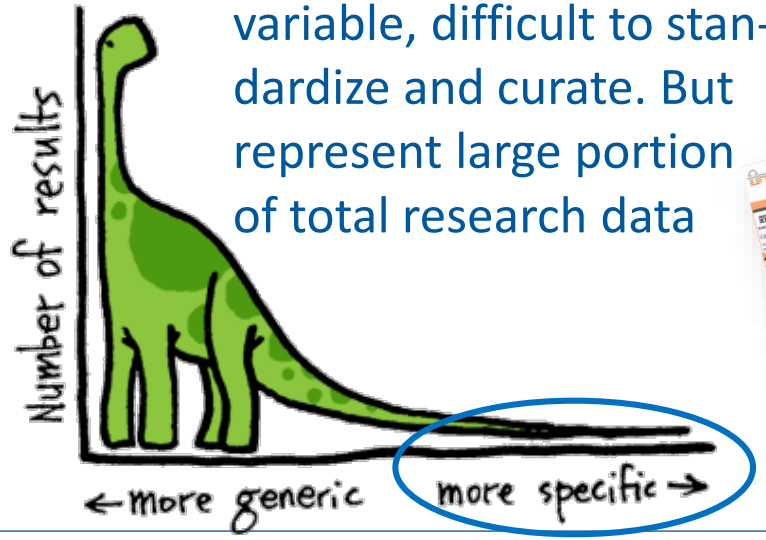


GFZ Data Services: Profile

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 standardize and curate. But represent large portion of total research data



Different layouts for DOI Landing Pages

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

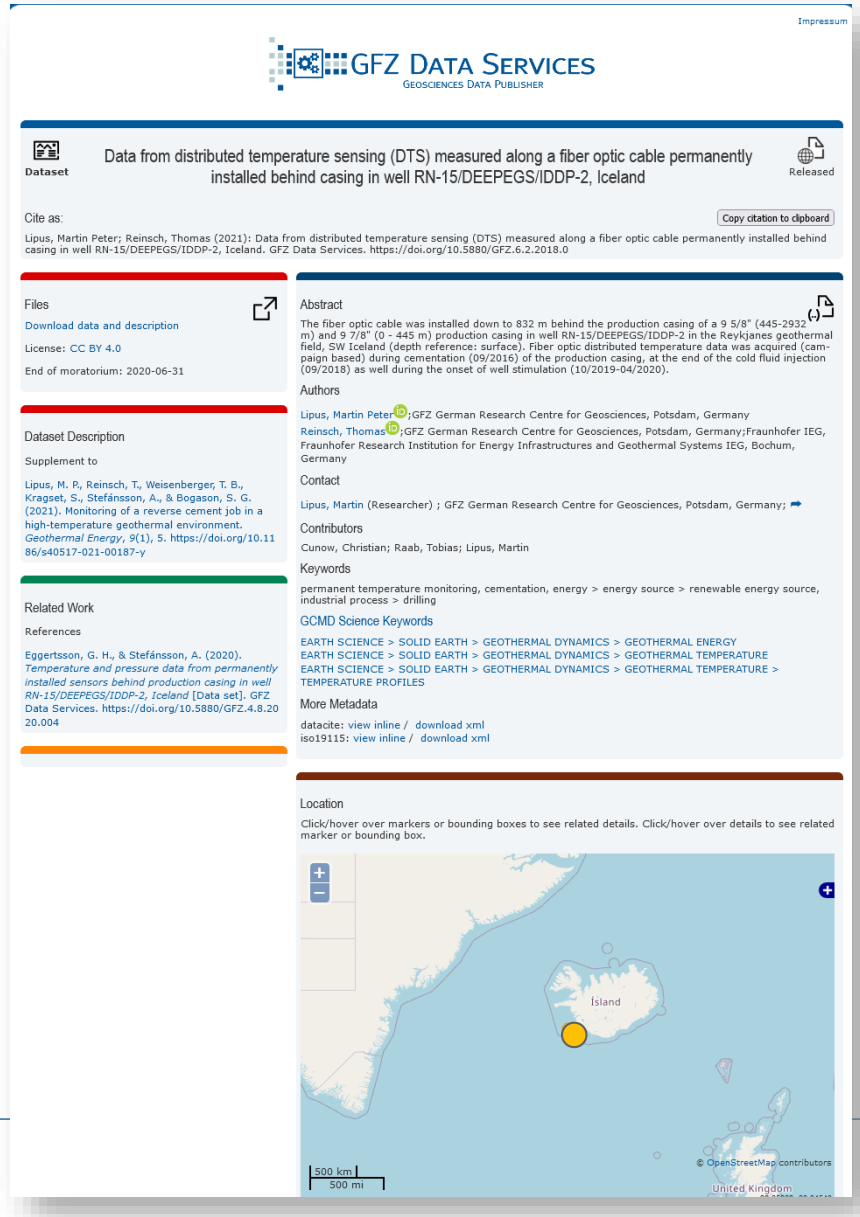
Metadata for data discovery: example DOI Landing Page

title citation

download data

key paper

related work



description/abstract

authors/ ORCID

keywords/control-led vocabularies

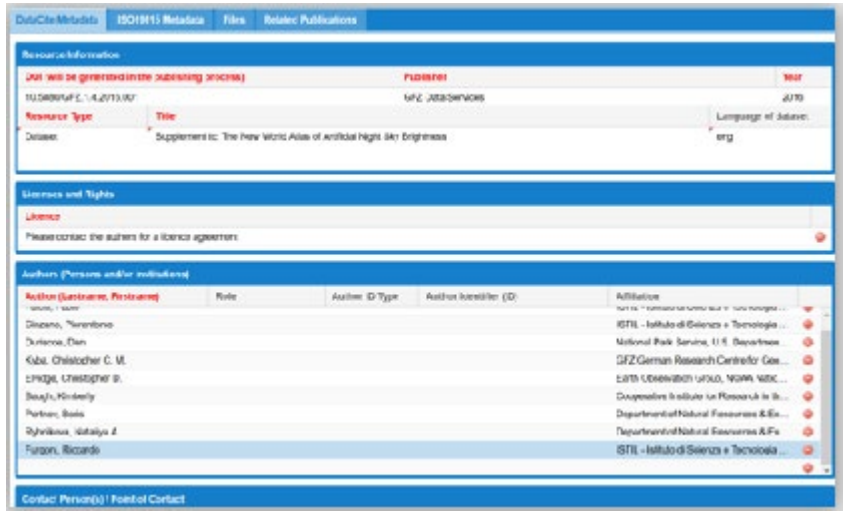
spatial coverage

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

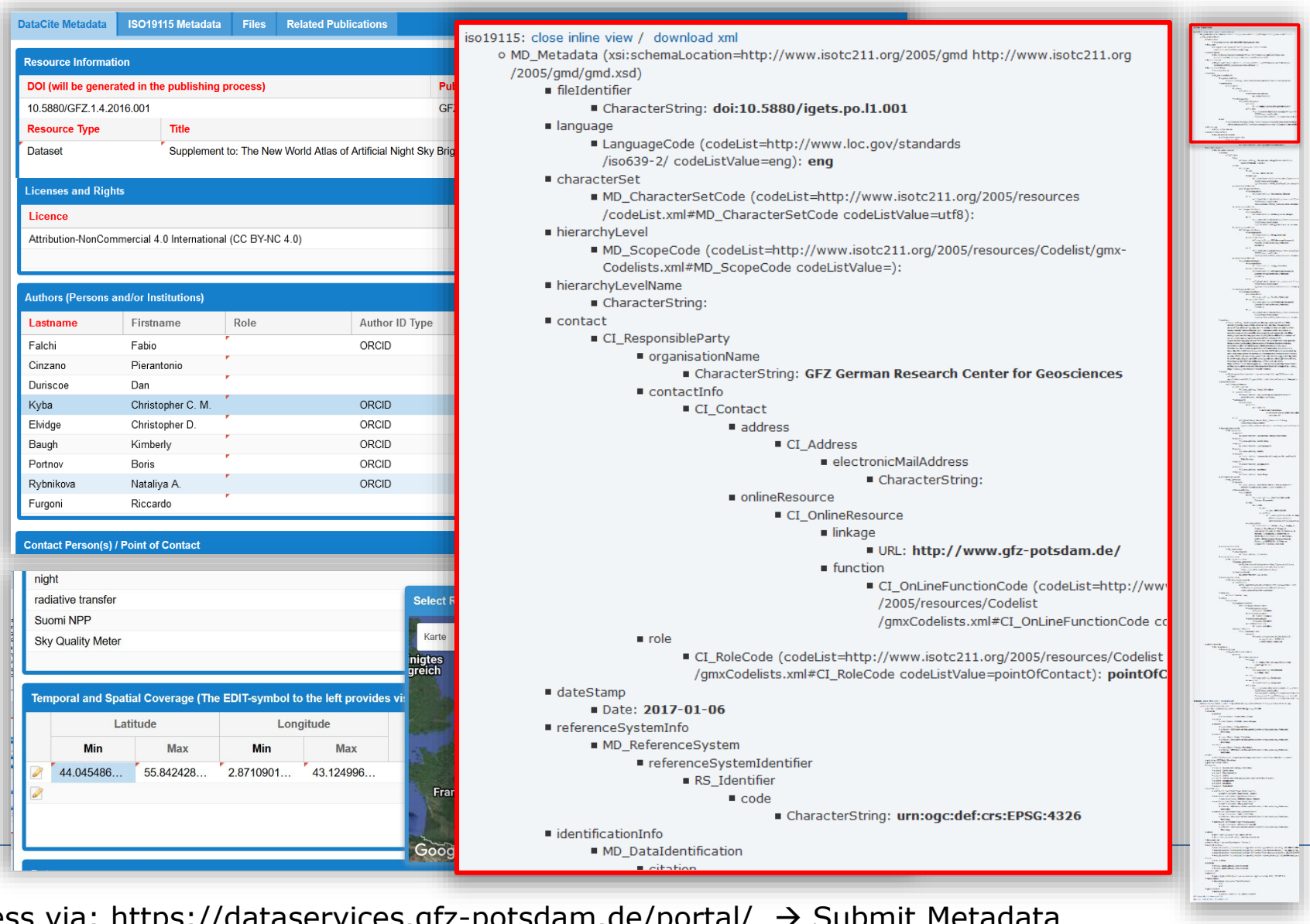
1. Discovery Metadata: via GFZ Metadata Editor



GFZ Metadata Editor (Java Script „translator“)

Input: provided by researchers

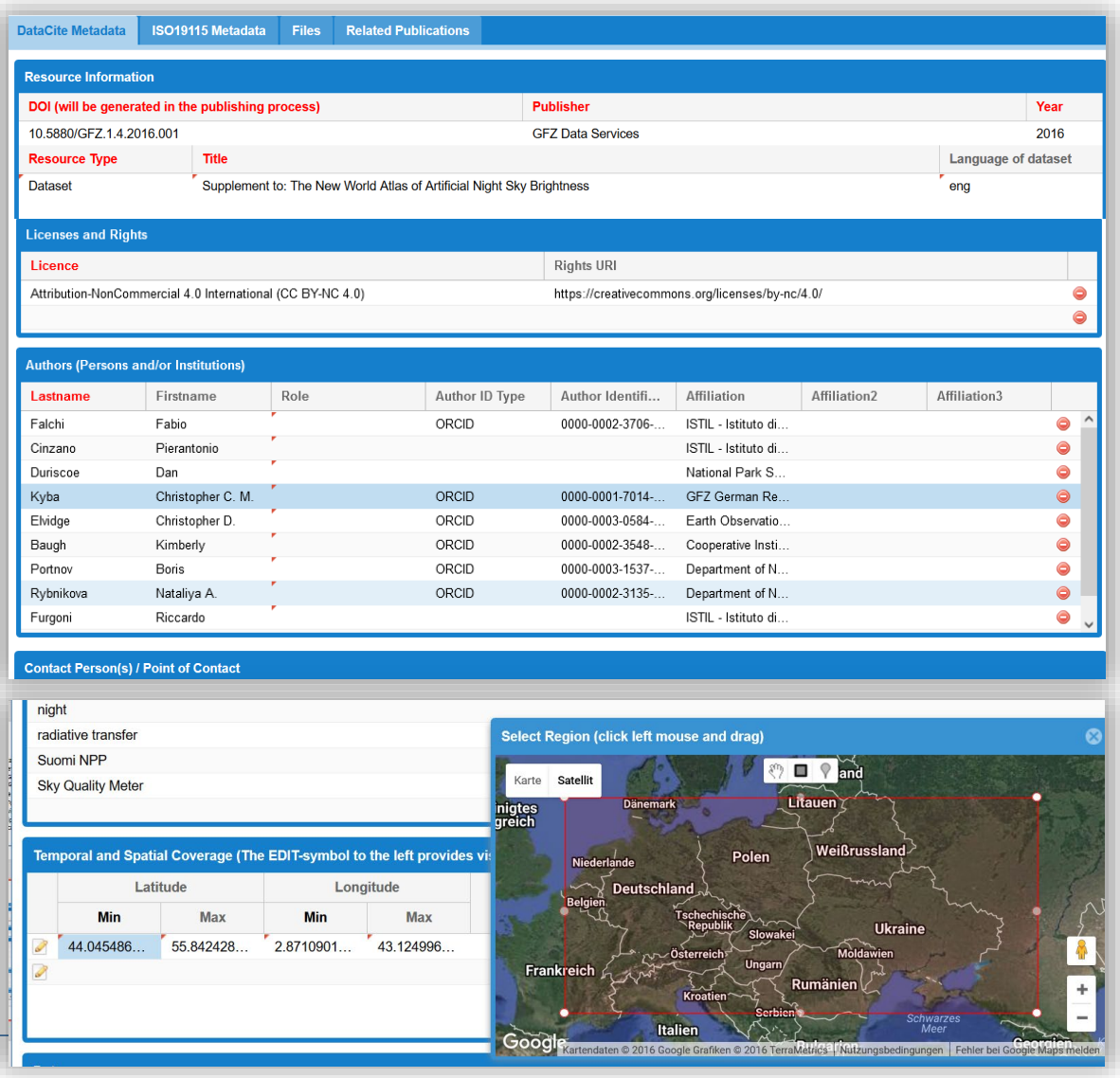
Output: XML (Extensible Markup Language): Metadata exchange format



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:

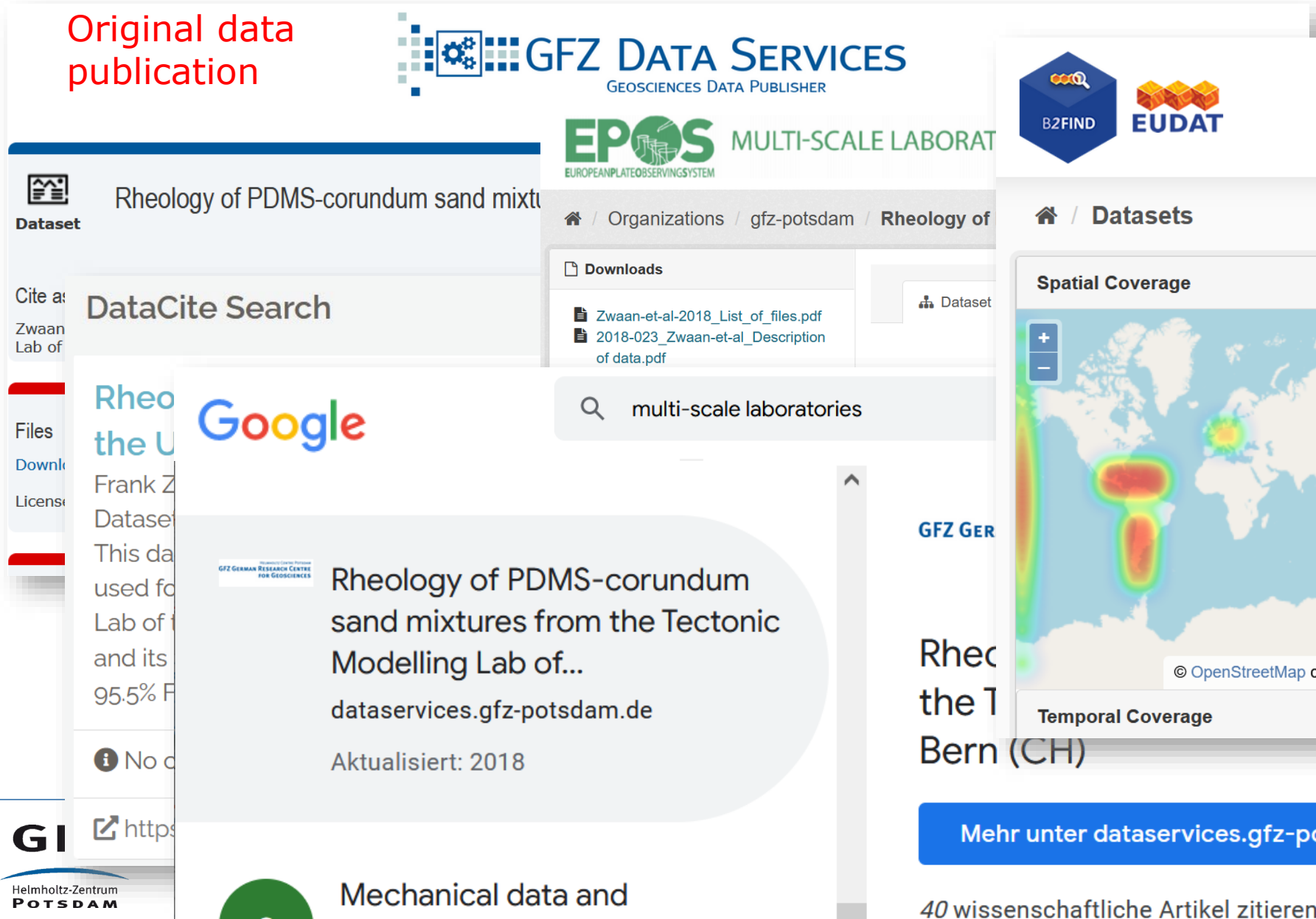
- Interactive map
- Searchable vocabulary lists

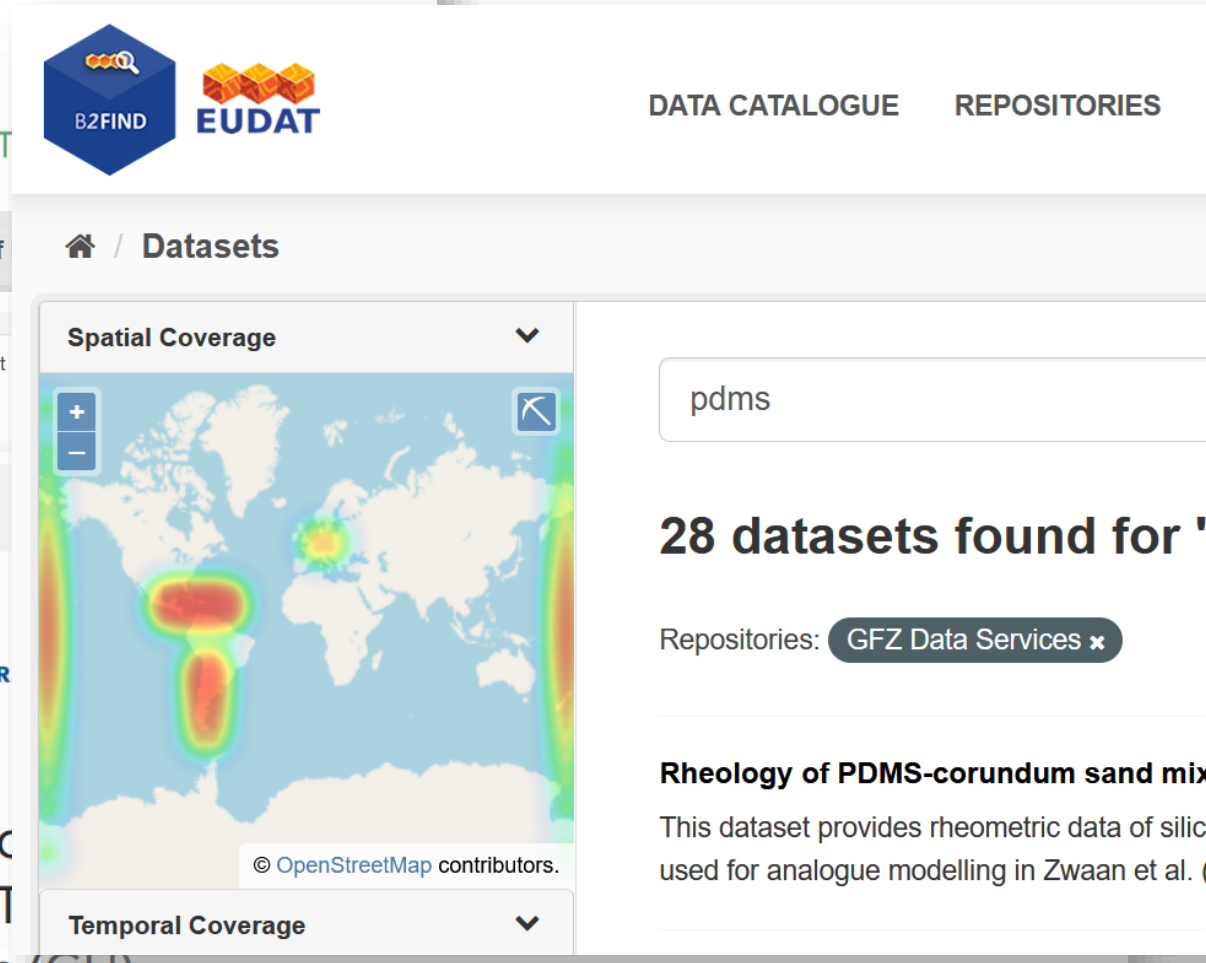
GFZ Helmholtz-Zentrum Potsdam

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

Metadata harvesting to other portals

Original data publication



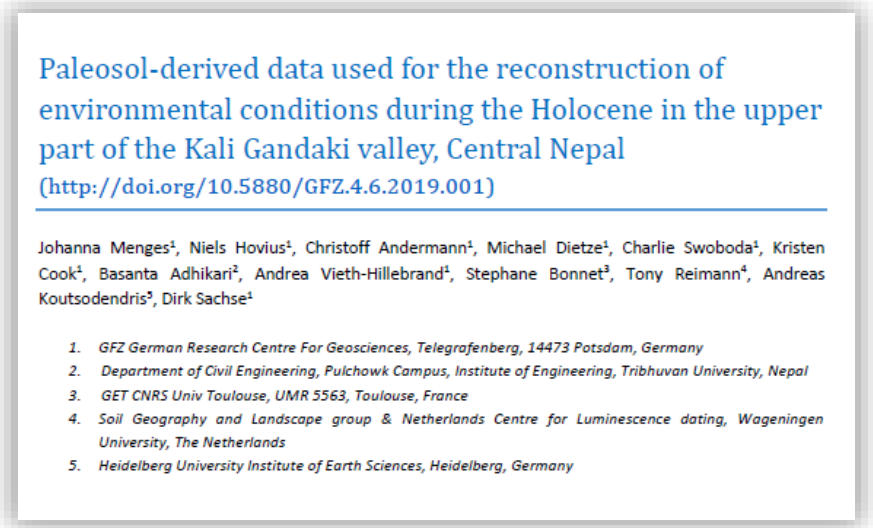


Helmholtz-Zentrum Potsdam

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]

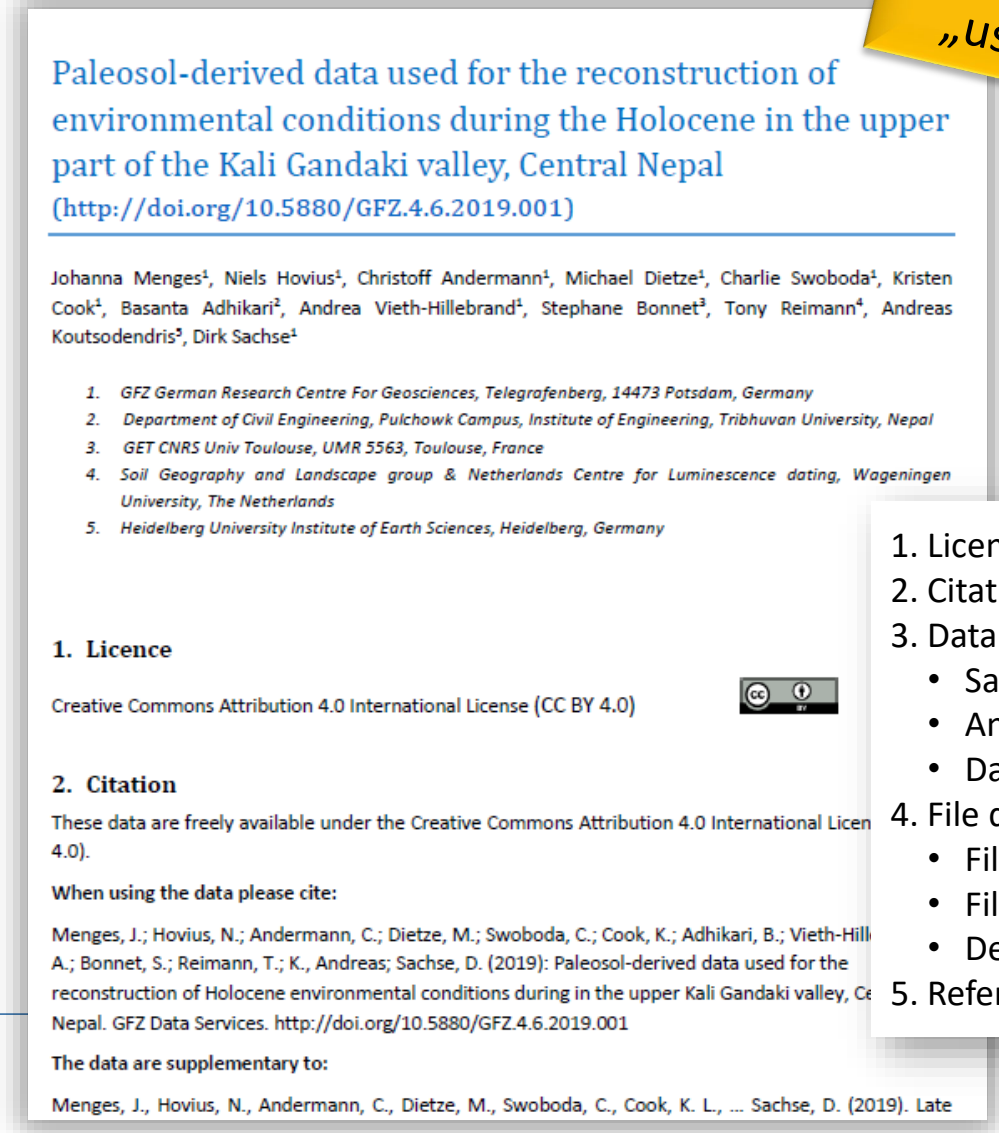
1. *GFZ German Research Centre For Geosciences, Telegrafenberg, 14473 Potsdam, Germany*
2. *Department of Civil Engineering, Pulchowk Campus, Institute of Engineering, Tribhuvan University, Nepal*
3. *GET CNRS Univ Toulouse, UMR 5563, Toulouse, France*
4. *Soil Geography and Landscape group & Netherlands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

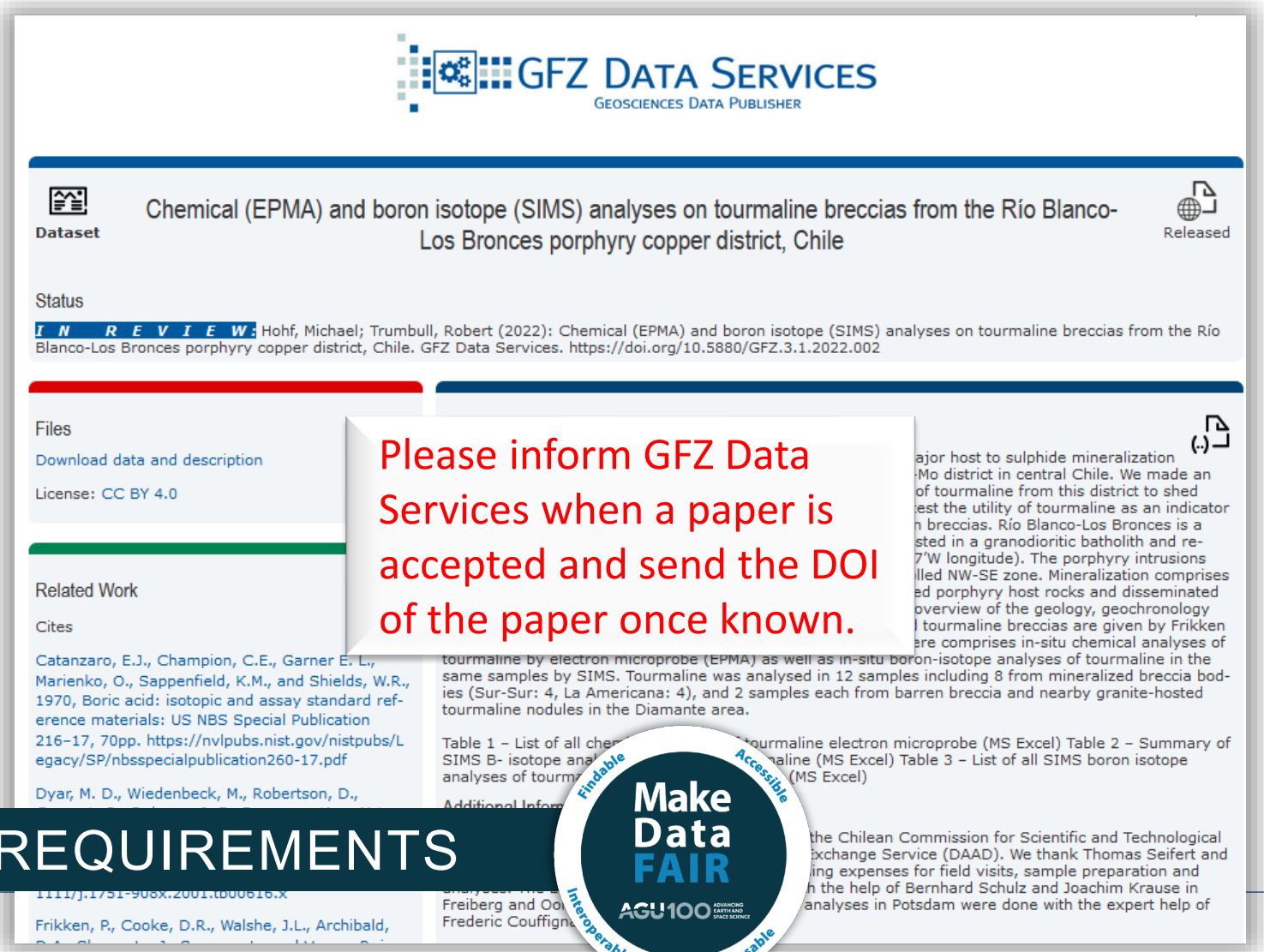
Tools for data publications by GFZ Data Services

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

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

Special Features: DOIs for software

Resource type →

Download static version →

Link to GitHub project page →

Open Source Software License →

Special Features: DOIs for software

Resource type →

Download static version →

Link to GitHub project page →

Open Source Software License →

Manual published as data report →

New Website launched in April 2022

- Responsive design
- Thematic search options
- Comprehensive **Publication instructions** (quick start guide, data file instructions, data description templates, …)
- **Portals**:
  - GFZ Data Services catalogue: DOI-referenced data
  - RI@GFZ: discovery portal for data and research infrastructure at GFZ
  - GFZ Sample Catalogue (IGSN)