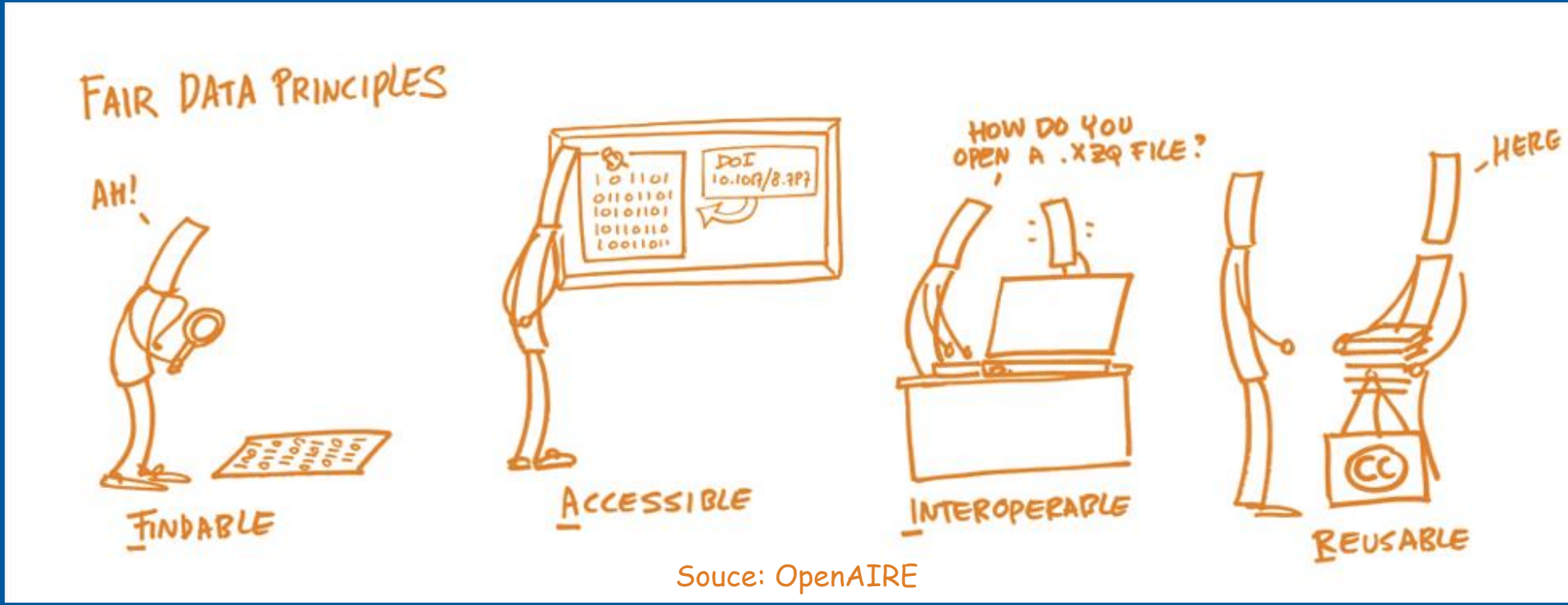
Wie wichtig ist die Dokumentation für die Nachnutzbarkeit von Forschungsdaten und wie werden meine Daten sichtbar? Beispiele von GFZ Data Services



Souce: OpenAIRE

FAIR im Long-Tail – Beispiele von GFZ Data Services

FAIR Principles

Findable

Accessible

Interoperable

Reusable

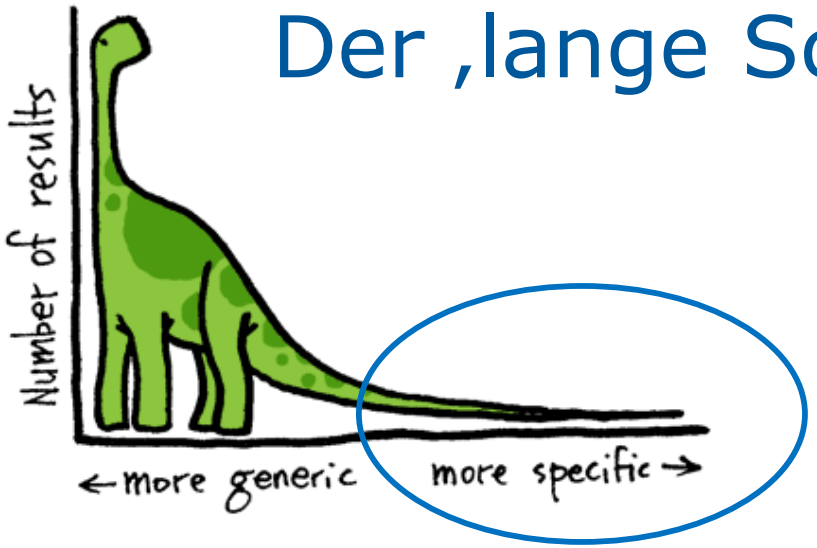
FAIR for Humans and Machines

"One of the grand challenges of data-intensive science, … is **to improve knowledge discovery through assisting both humans, and their computational agents**, in the discovery of, access to, and integration and analysis of, task-appropriate scientific data and other scholarly digital objects." (Wilkinson et al., 2015, https://doi.org/10.1038/sdata.2016.18)

– Reusable – Strategien im long-tail

– Findable: Nutzung von PIDs

Der ‚lange Schwanz' der Forschungsdaten



Als **'Long tail of research data'** werden Forschungsdaten bezeichnet, die in der Regel ein **geringes Datenvolumen** aufweisen, in **verschiedenen Datenformaten** vorliegen und in Bezug auf Archivierung und Management nur **schwer standardisierbar** sind.*

*Quelle: https://www.forschungsdaten.org/index.php/Long_tail_of_research_data

Data Publications – best practice for FAIR sharing data

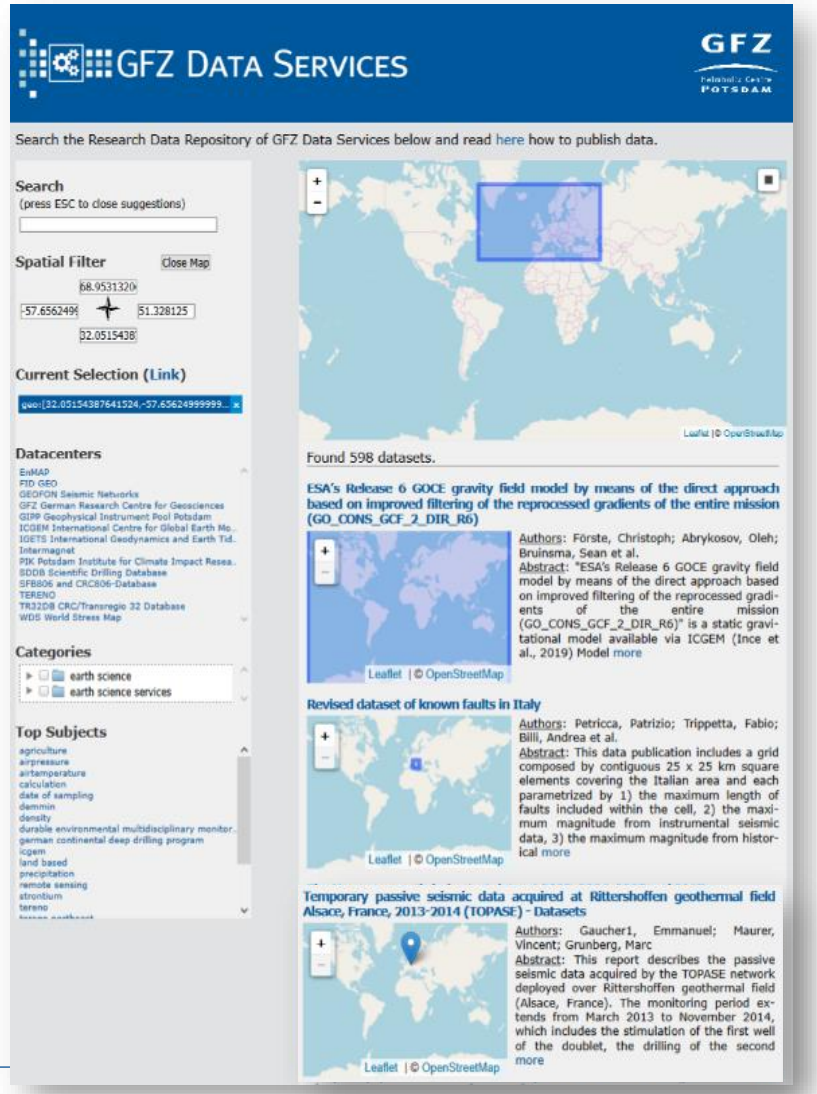
Publication of datasets as individual publications (with assigned persistent Identifier; DOI) **through data repositories** (ideally **domain repositories**)

- **Findable:** PID, standardised machine readable metadata, controlled vocabularies: discovery via portals and search engines,
- **Citable**: DOI-referenced datasets are citable just as journal articles (→ credit for researcher and institution)
- **Accessible:** via DOI/ PID, persistent data storage and access guaranteed by the publisher (= data repository)
- Interoperable: machine-readable data and metadata
- **Documented:** with metadata for discovery and reuse, licence

GFZ Data Services – domain repository

Profile

- Domain repository for the Geosciences
- DOIs for Data and Software
- Data: real-time data streams (observational data), tables, maps, model data, …
- Data curation by discipline scientists
- Data description templates, reports (internal review)



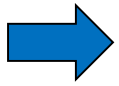
https://dataservices.gfz-potsdam.de

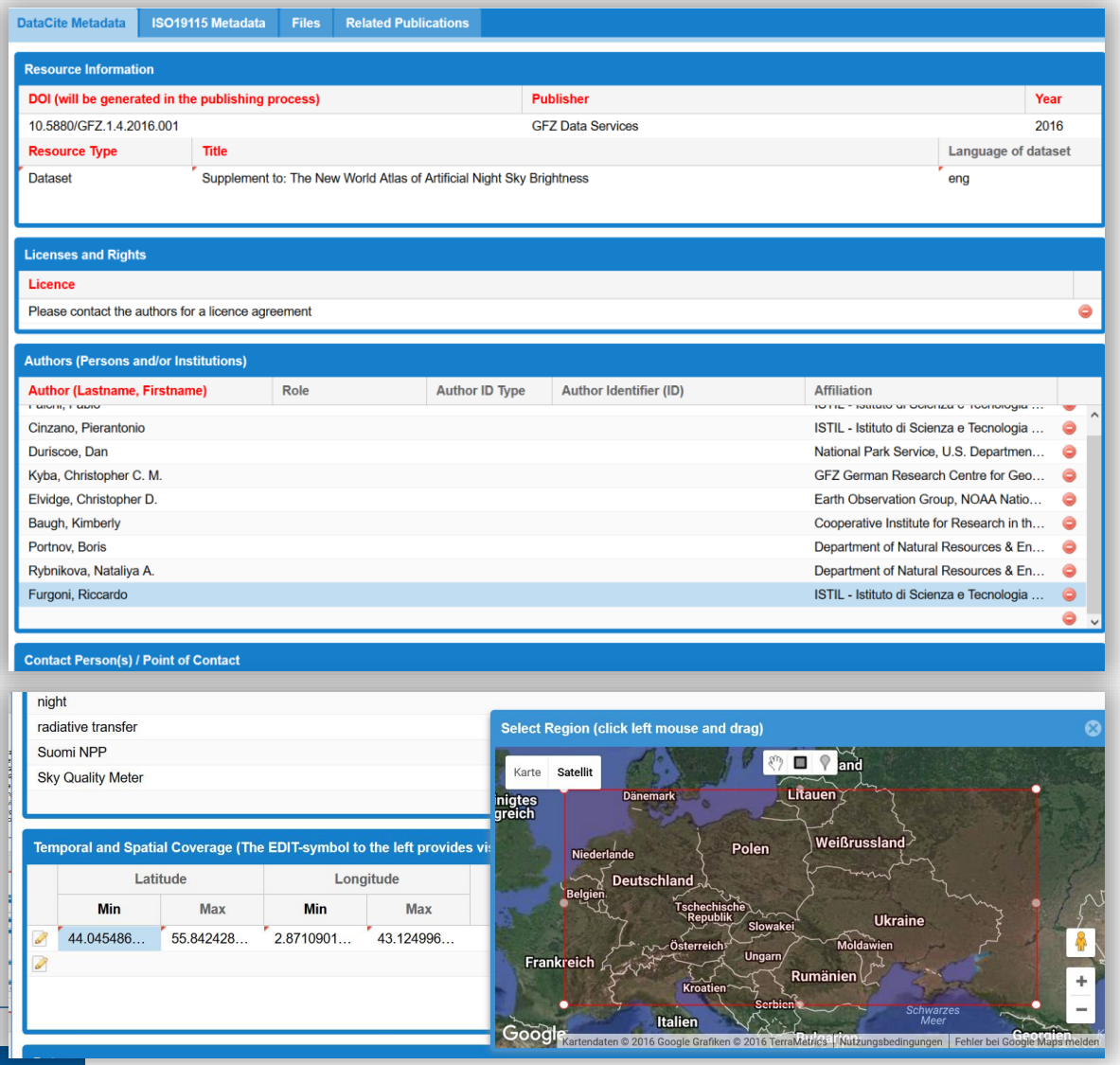
Technical „Highlights“

- International metadata standards (human & machine readable)
- Vocabularies for „rich“ metadata
- PIDs (DOI, ORCID, Fundref, IGSN)
- Open Licences for data and software
- OAI-PMH, Rest interface
- schema.org → Google Dataset Search

Tools 1: GFZ Metadata Editor (Java Script „translator“)

Input:
provided by researchers





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

DOI Landing Pages

„Special“ Features:

- Interactive map
- Searchable vocabulary lists

Standardised API

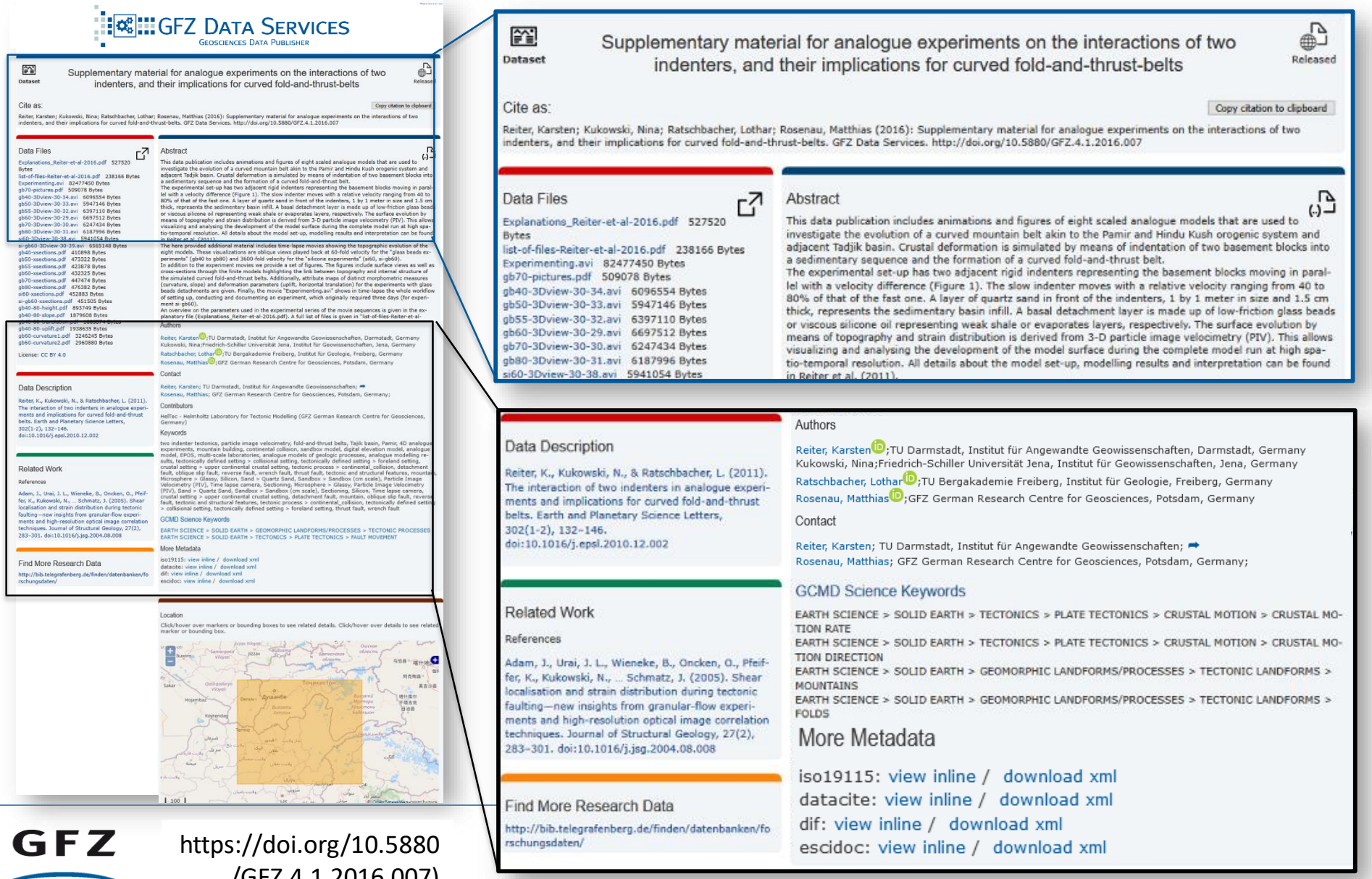
DataCite
FIND, ACCESS, AND REUSE DATA

B2FIND
EUDAT
EPOS
EUROPEANPLATEOBSERVINGSYSTEM

GFZ Helmholtz-Zentrum Potsdam
GFZ Data Services

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

Data access via DOI Landing Pages



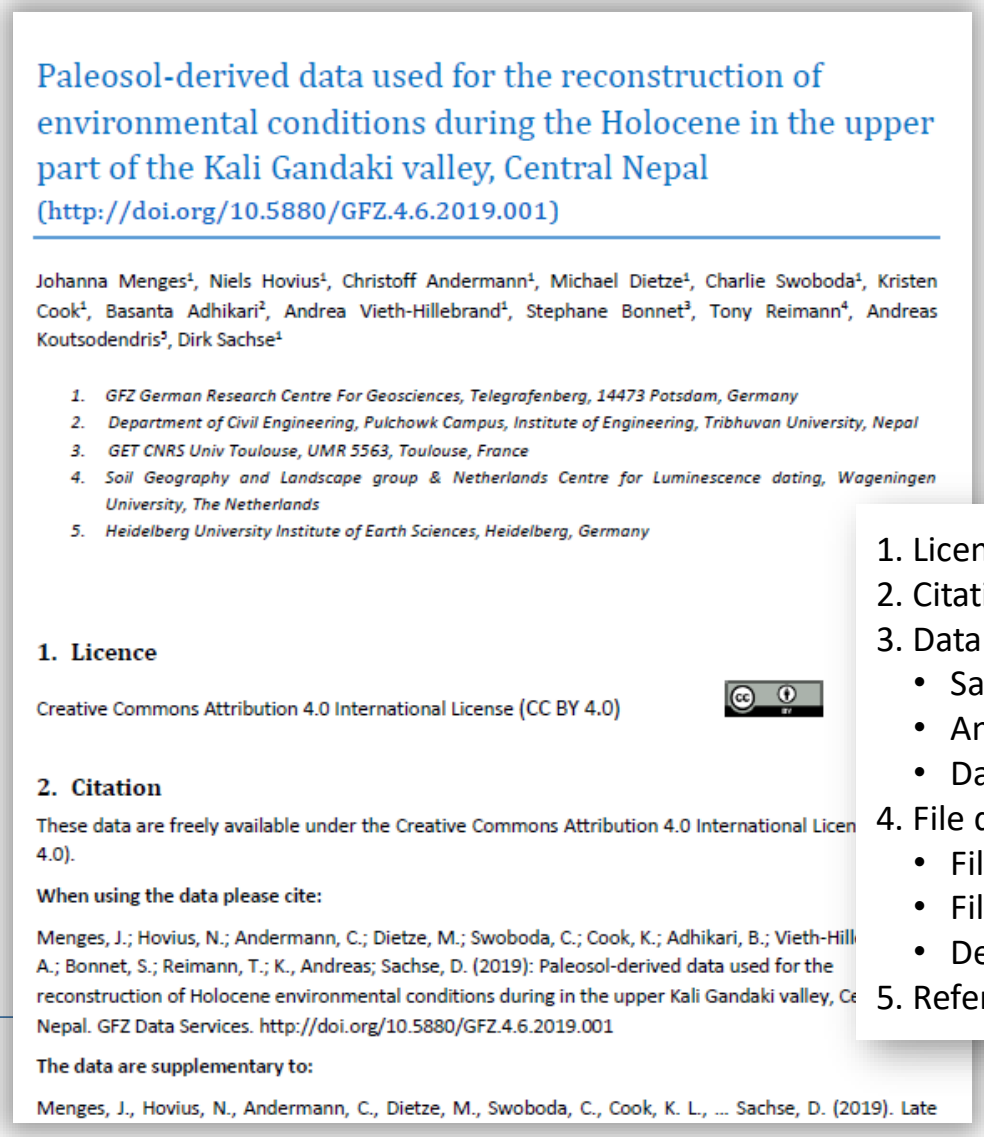
https://doi.org/10.5880/GFZ.4.1.2016.007)

- **Citation** information
- **ORCIDs** of authors
- **Controlled Vocabularies:** NASA GCMD Science Keywords, GeoSciML
- **Related References** to papers, reports, data, software, IGSNs (via PIDs)
- **Map**
- **XML metadata for download**: ISO19115, DataCite, Dublin Core

Tools 2: Data Description Templates

- 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

➢ Templates soon available via https://dataservices.gfz-potsdam.de (until then: via Email)



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

Example Data Description: Before the template

Stimulation data for each of the analyzed EGS projects.

The are provided in tabular form (CSV). The file names indicate 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 (-)

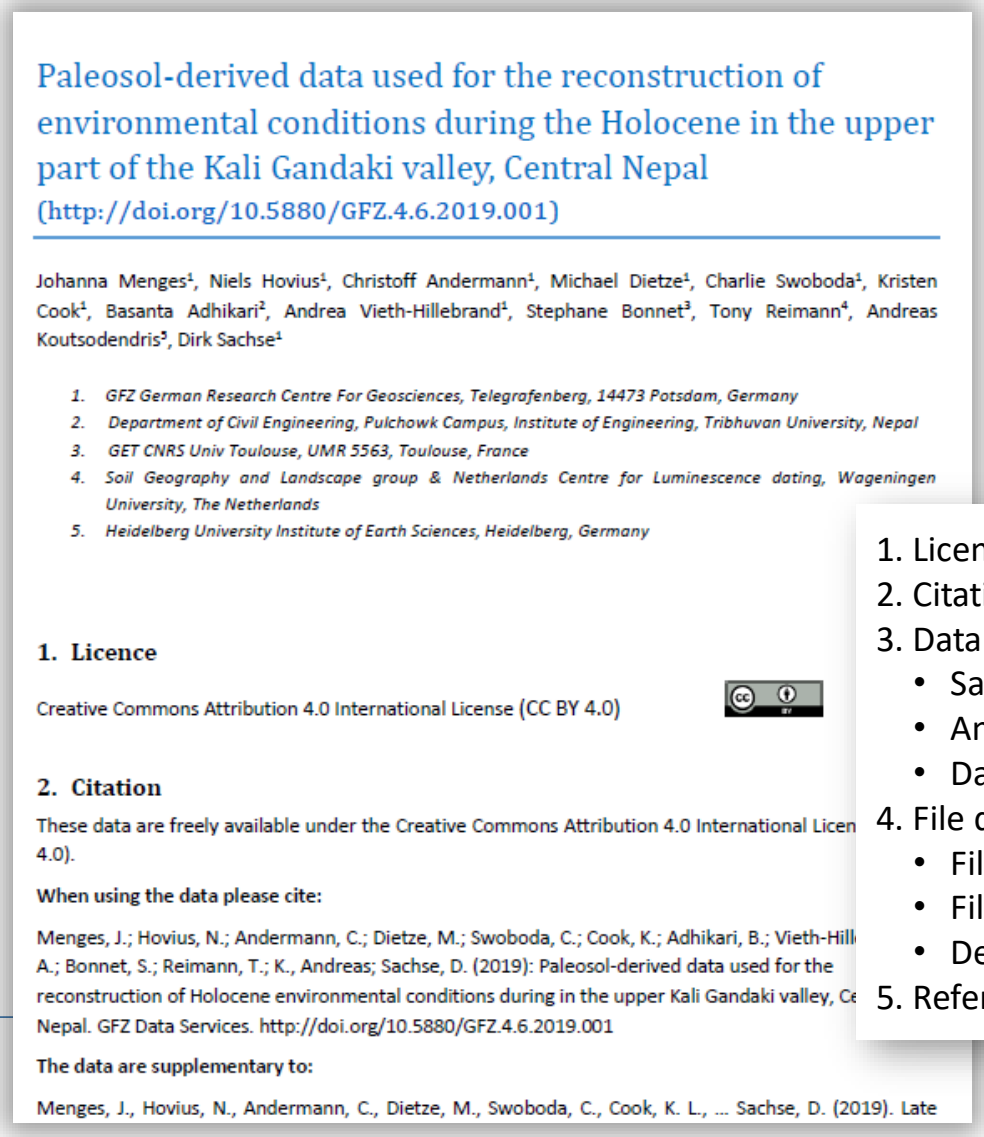
Example Data Description: after (= with the template)

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.

Tools 2: Data Description Templates

- Templates are an important tool for researchers and data curators to increase the quality of data descriptions
- Standardised layout and structure of data description



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

➢ Templates soon available via https://dataservices.gfz-potsdam.de (until then: via Email)

Part 2: Persistent Identifier in data publications

for data, software, cross-references to related work

PID for physical samples, cross references to samples underlying measurements

uniquely identifying persons

List of funders with DOIs

New PID for Institutions

PIDs and the provenance of research outcome

Data

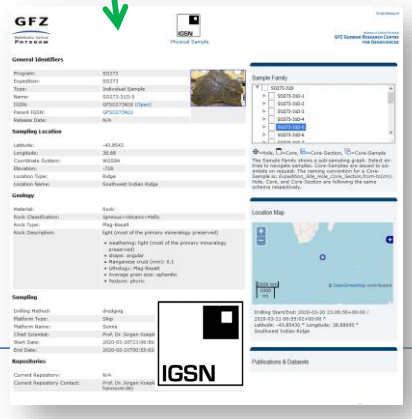
Table S3. Sierra Nevada analyses of plant samples

sample ID	IGSN	brief sample description	Element concentrations (µg/g) Al	Fe	Mn	Mg	Ca
fresh							
MW1							
MW2							
MW3							
MW4							
mean							
(2SE							

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
SN01	*	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	15.3	6.95

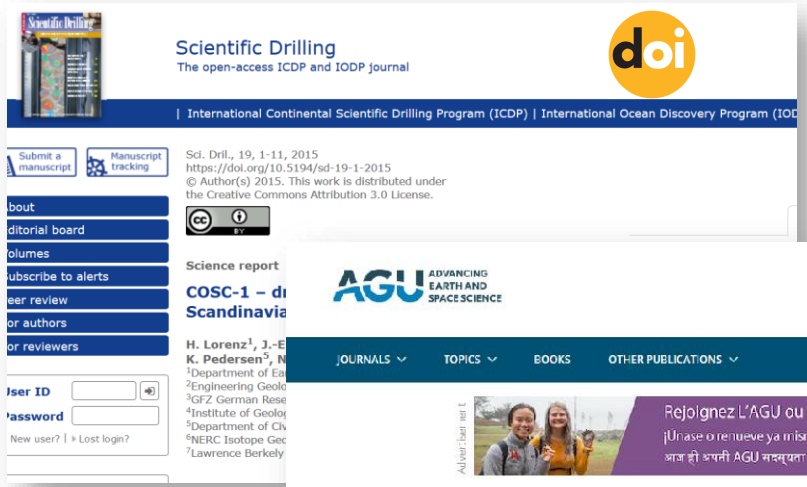
Sample



Sample Description

Data Publication

Papers



Report

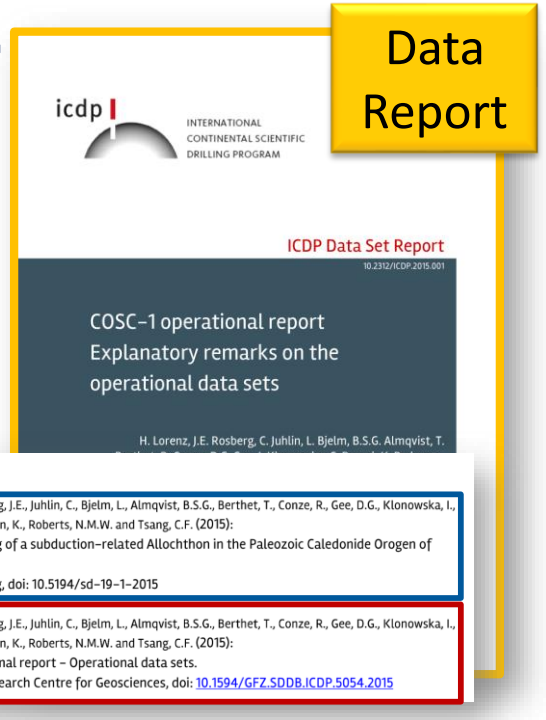
Linking papers, data, samples, …

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	

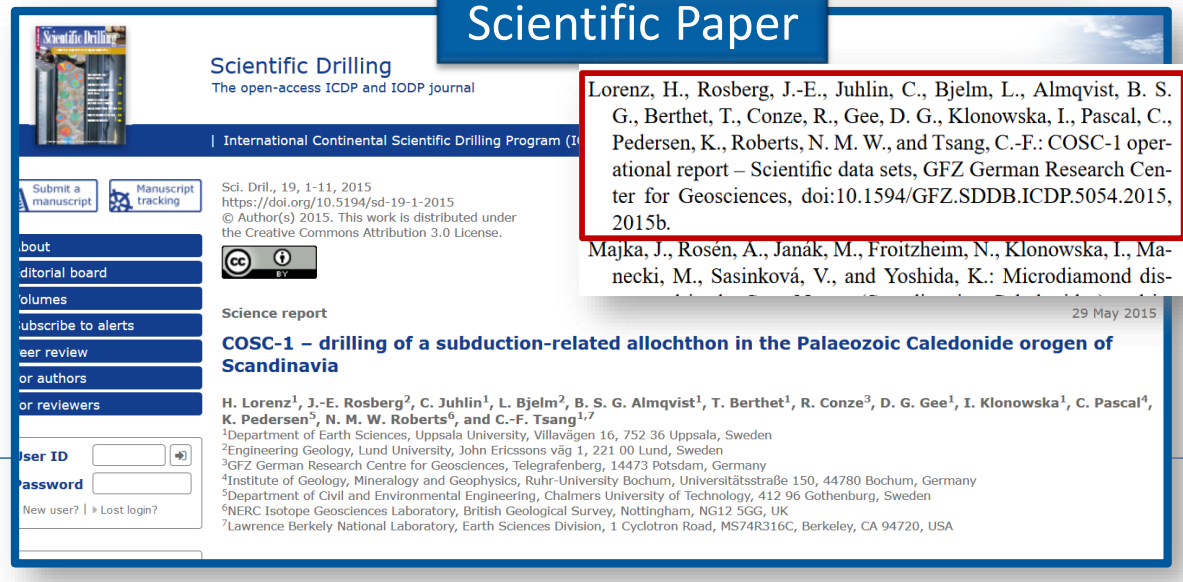
Data

Data Report

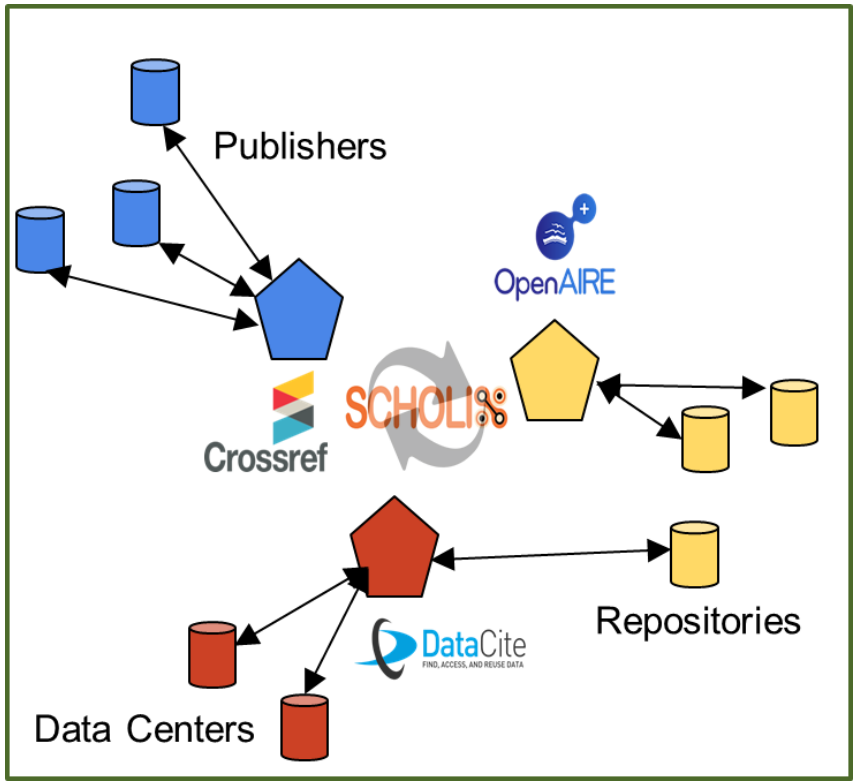


Sample

Scientific Paper



Scholix: linking data with papers



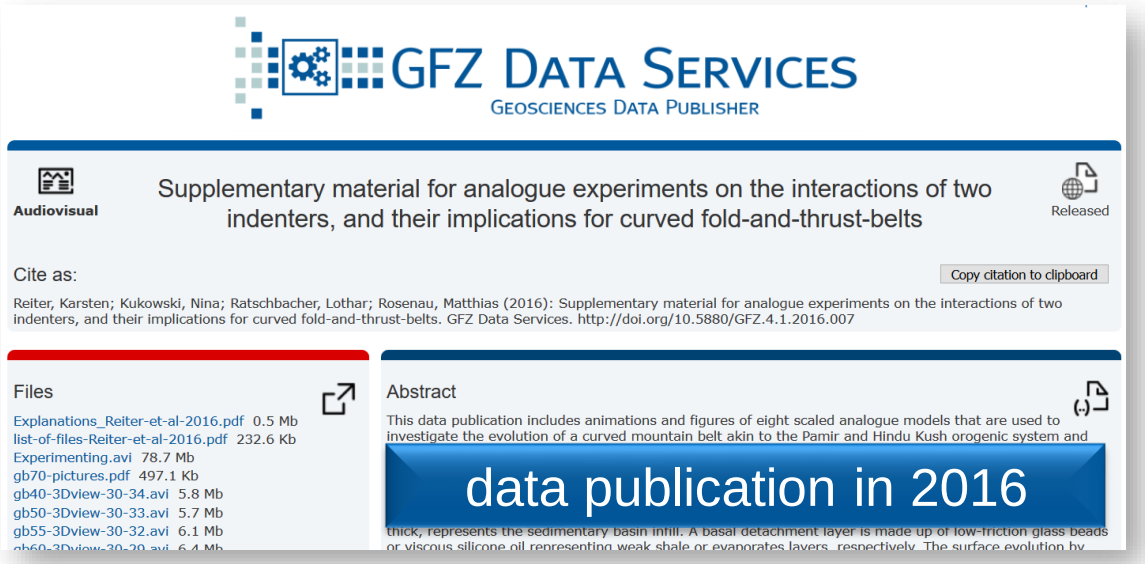
Credit: Wouter Haak, Elsevier

SCHOLIX

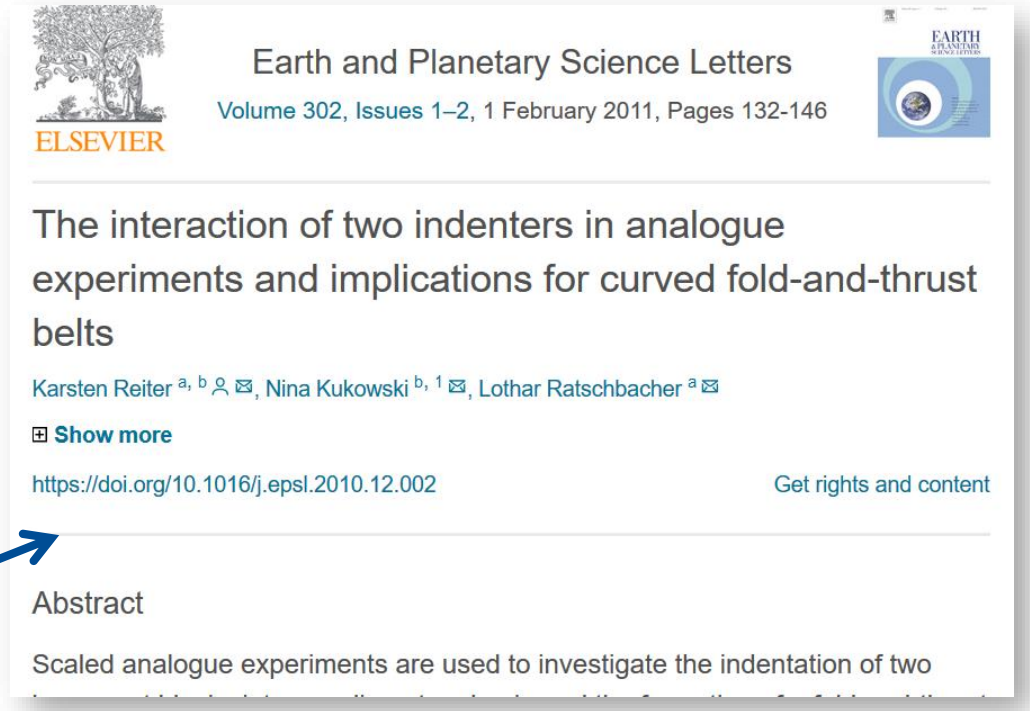
"interoperability framework for exchanging information about the links between scholarly literature and data"

Scholexplorer
The Data Literature Interlinking Service

The classical approach (before Scholix)

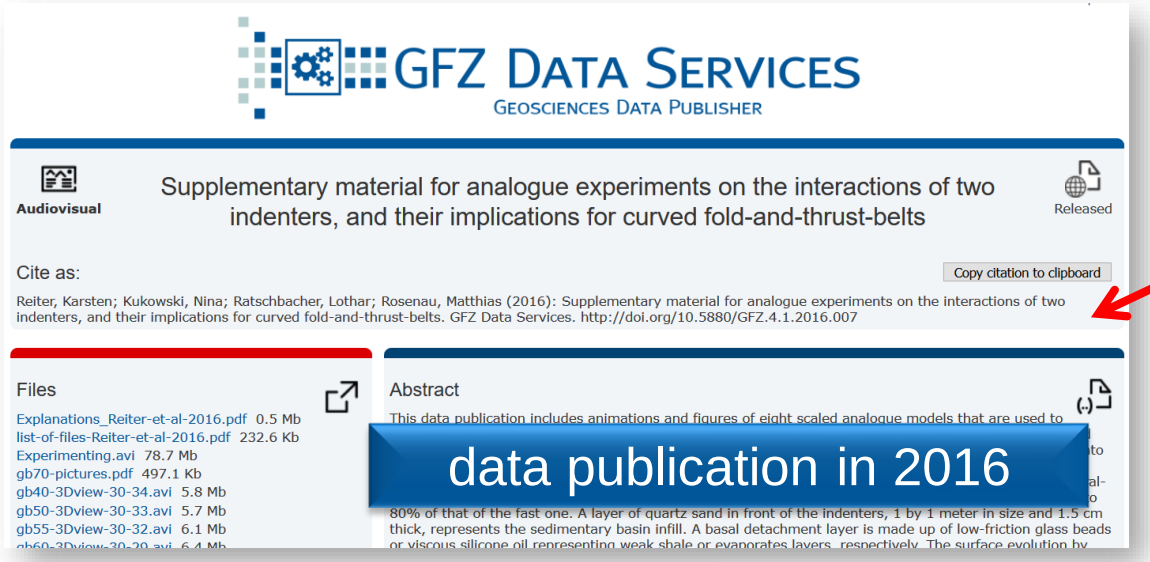


…supplement to an article from 2011

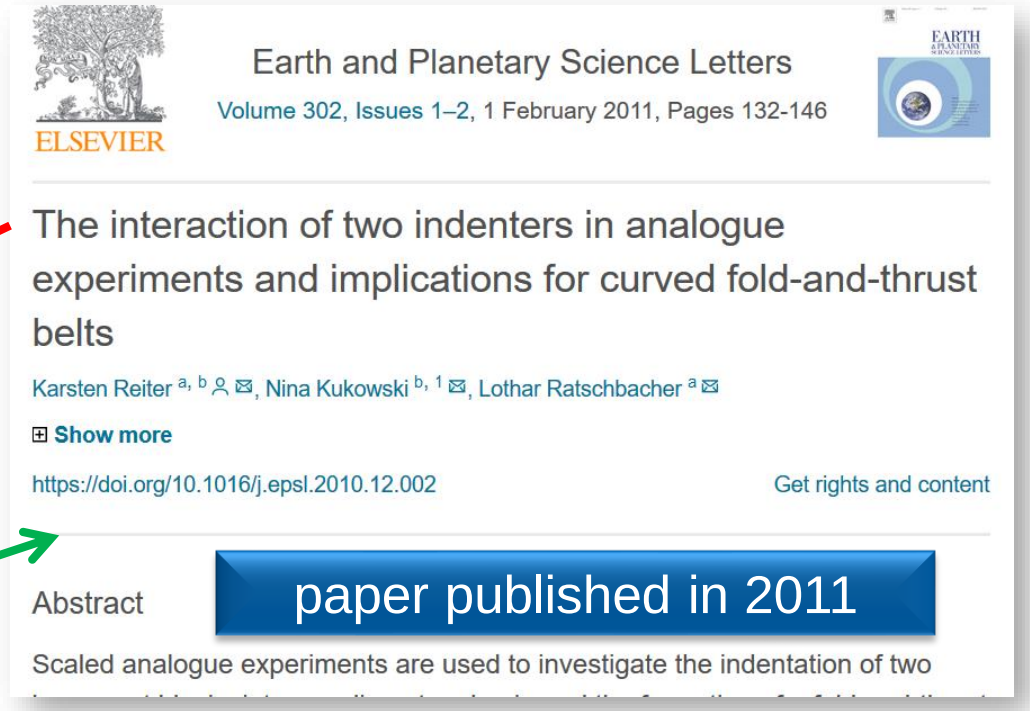


Data: http://doi.org/10.5880/GFZ.4.1.2016.007, Paper: https://doi.org/10.1016/j.epsl.2010.12.002

The classical approach (before Scholix)



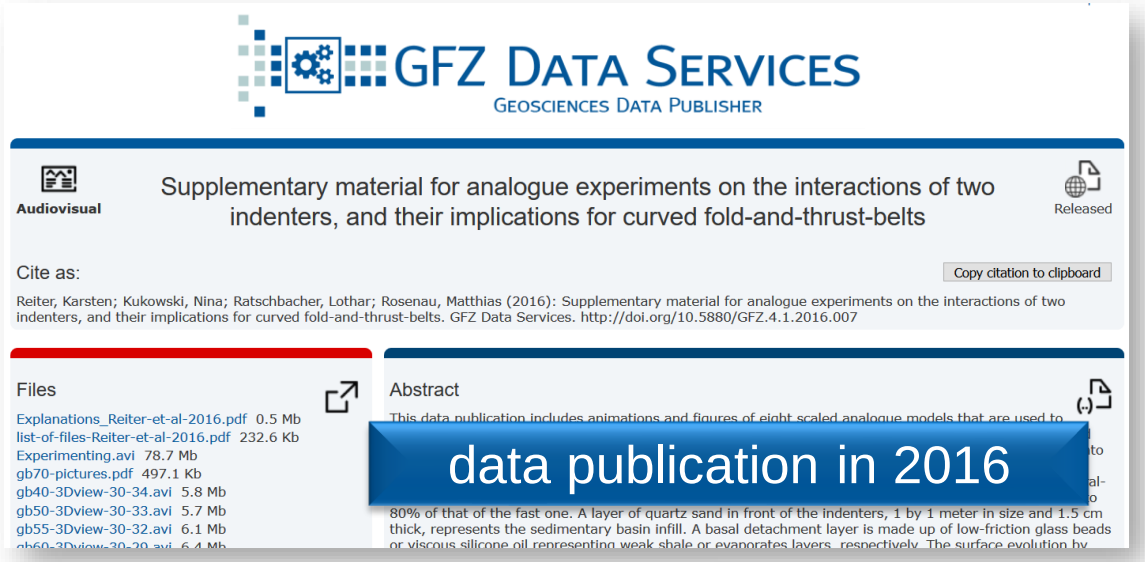
Link to paper in the metadata of the data publication

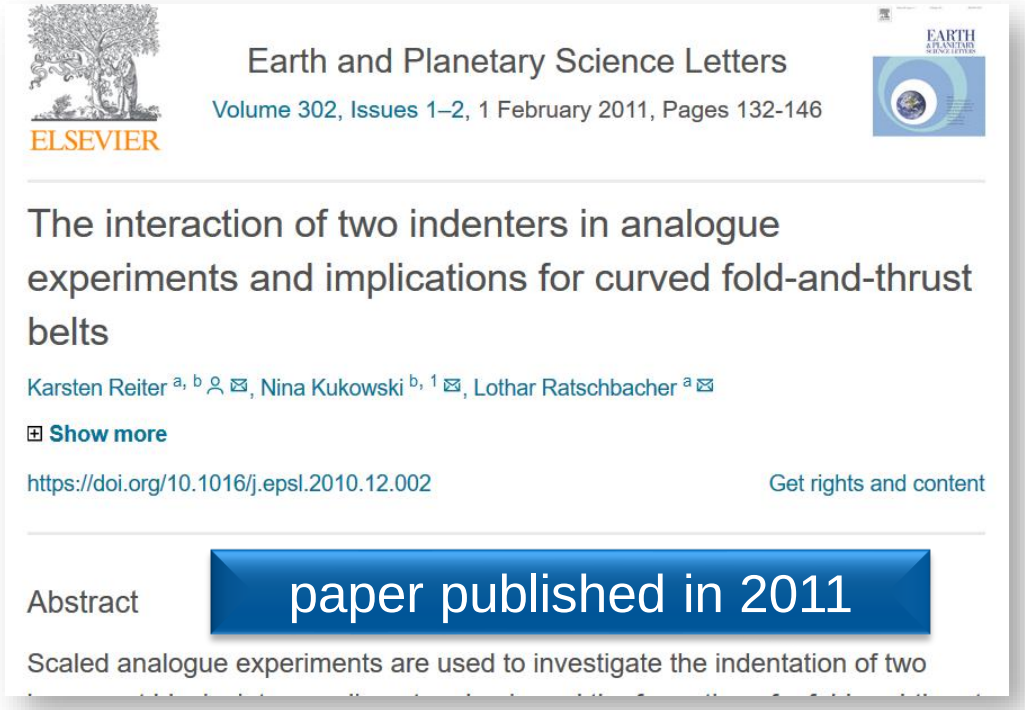


Is **not** possible to add the citation of the dataset to the paper after publishing

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

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

Vielen Dank für Ihre Aufmerksamkeit!

Kirsten Elger
kelger@gfz-potsdam.de