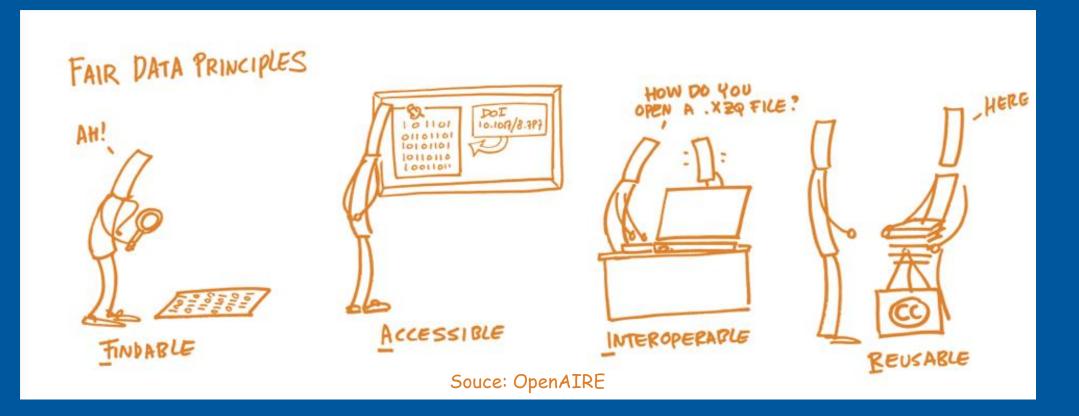
# Die Rolle von Fachrepositorien und Datenjournals bei der Umsetzung der FAIR Prinzipien



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 Principlies for Research Data Management





### 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

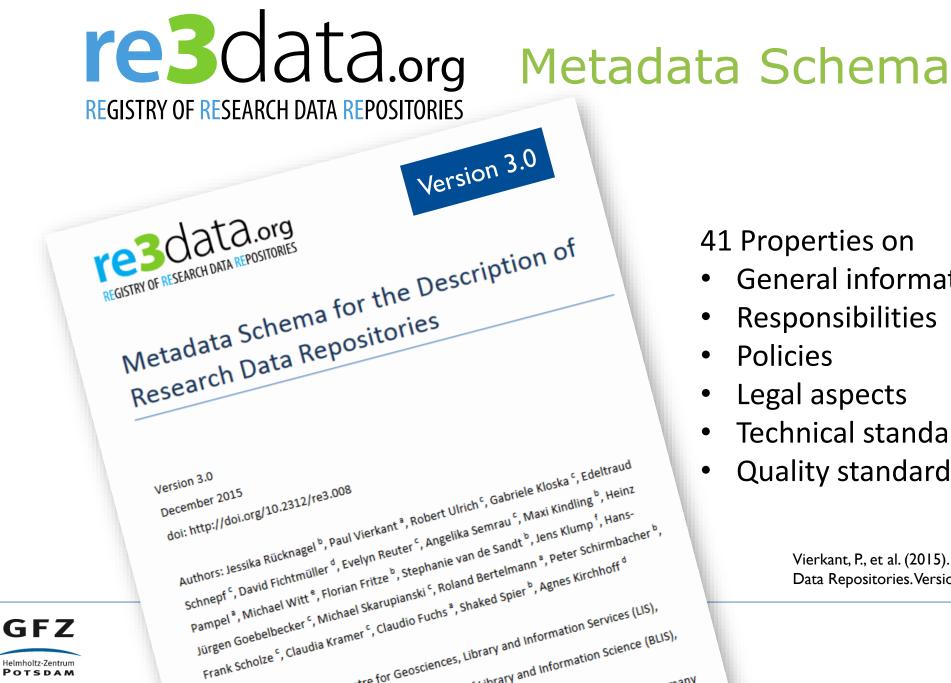




- **global registry** of research data repositories
- covers all academic disciplines
- presents repositories and portals for the permanent storage and access of research data sets to researchers, funding bodies, publishers and scholarly institutions.
- promotes a culture of sharing, increased access and better visibility of research data

2466 registered repositories (25 Feb 2020)





### 41 Properties on

- General information
- Responsibilities
- 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



Helmholtz-Zentrum Potsdam

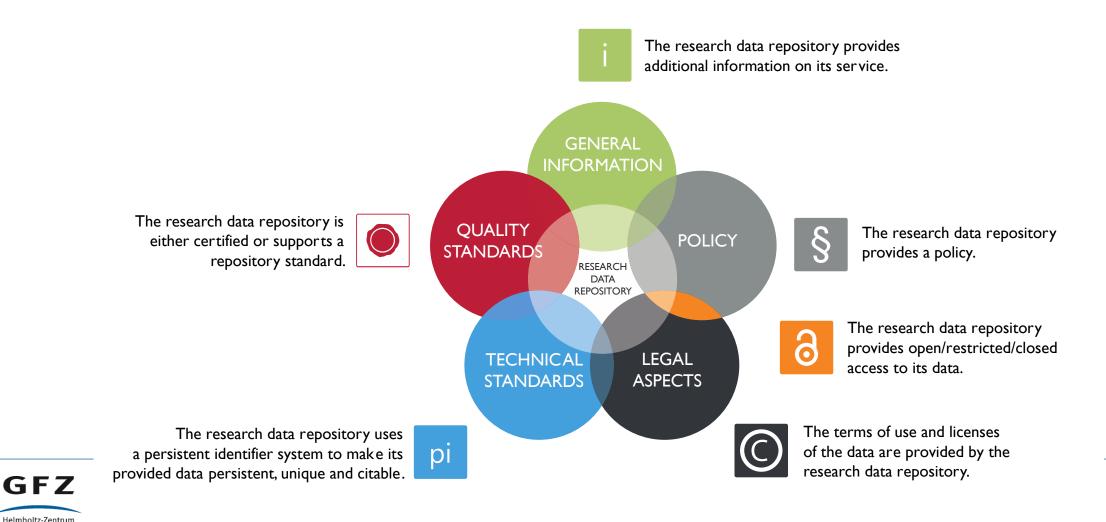
simple	Home Search Brows	e Suggest FAQ A	About Schema API	Contact Legal notice / Impre	ssum		
search box	Filter Subjects ⊞	Nordicana d			Q Searc		
filters	Content Types ⊞ Countries ⊞ Data access ⊞ Database access ⊞ Data licenses ⊞ Data upload ⊞	← Previous 1 Ne	$ext \rightarrow$	icons	Sort by		
		Found 1 result(s)					
	Data upload restrictions ⊞ Enhanced publication ⊞ Institution responsibility	Nordicana D collection Subject(s)	aphy) Natural Sciences				
	type ⊞ Institution type ⊞ Keywords ⊞	Content type(s)	Raw data Structured graphi Plain text	ics Images Scientific and statistical	data formats		
	Metadata standards ⊞ PID systems ⊞	Country Nordicana series D is a	Canada a formatted, online data report	t series archived at CEN. It is pro	duced only in		
Z	Provider types ⊞ Quality management ⊞ Repository languages ⊞ Repository types ⊞ Versioning ⊞	electronic form and is freely and openly accessible to CEN researchers and to other users. Each issu is published in French and in English, and is indexed via an assigned digital object identifier (DOI). A 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.					



POTSDAM

Icons

facilitating the selection process of appropriate research data repositories





News

Search



Repository Finder, a pilot project of the Enabling FAIR Data Project led by the American Geophysical Union (AGU) in partnership with DataCite and the Earth, space and environment sciences community, can help you find an appropriate repository to deposit your research data. The tool is hosted by DataCite and queries the re3data registry of research data repositories.

As part of the FAIRsFAIR project, which aims to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle, the Repository Finder is extended to query for repositories relevant to FAIRsFAIR Project.

#### Search re3data for a repository to upload your data



# 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





**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



"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."



Hanson et al. (2015, doi:10.1029/2015EO022207) and www.copdess.org

### 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 <u>re3data.org</u>. 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 <u>data availability</u>). In addition, Copernicus Publications provides with <u>Earth System Science Data</u> (<u>ESSD</u>) 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) <u>commitment statement</u>.

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.

### Advancing FAIR Data in Earth, Space, and Environmental Science

Inter

Make

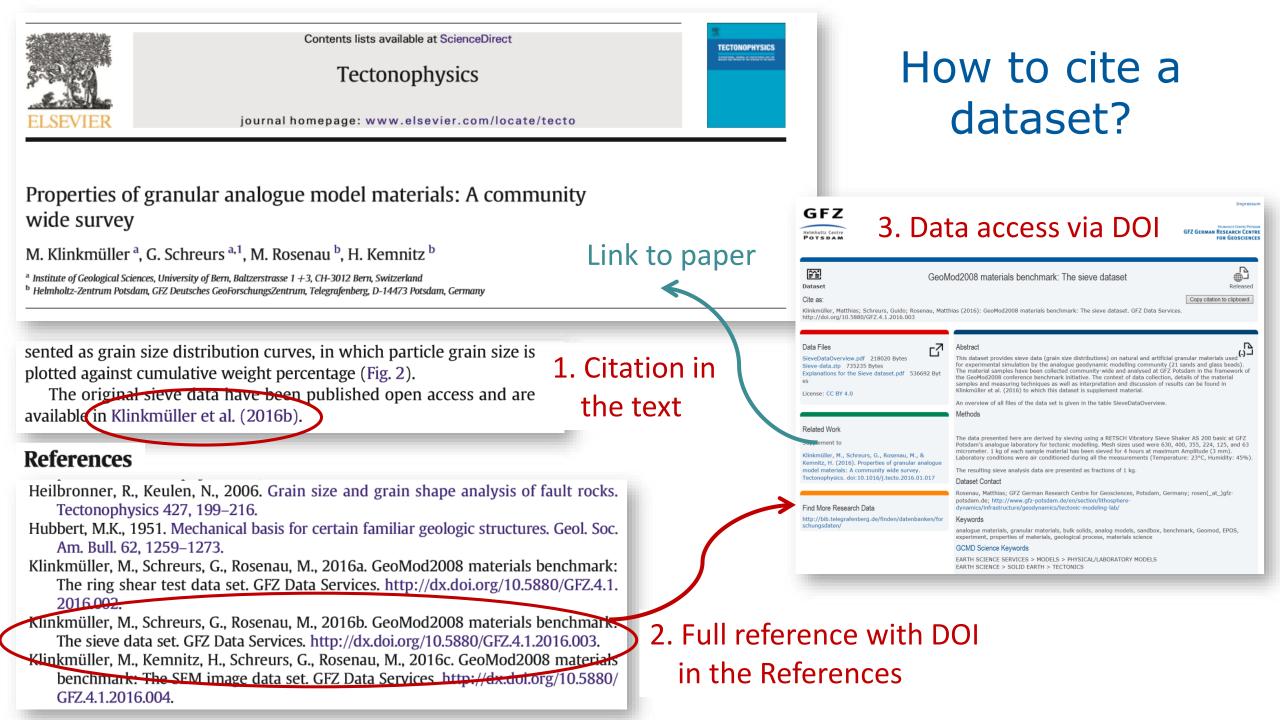
Data

AGUIOO ADVANCIN

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.



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



### 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	<ul> <li>For decoding and rendering files</li> <li>Long-term management of files</li> <li>Intellectual property rights attached to content</li> </ul>
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**

Unit

GFZ Helmholtz-Zentrum

Validation Text

integer val

# **Definition of data labels**

expedition number

Data Type Description

Numeric

Sites

1

2

Column Name

EXPEDITION

SITE

### DEVDWE

GI

Helmholtz-Zentrum POTSDAM

KEADIVIL				Nume	eric site number		Integer value	
		3	NAME				integer value	#
1		5	NAME	Text	sito no	sito an		#
atei Bearbeiten Format Ansicht ?					site name or locality		text string of max. 4 characters	0
VERTISSEMENT / WARNING					platform identifier, C=	Chikvu		#
lême si des efforts sont déployés p	our		LATFORM	Text	R=Drill Rig	oecific,	text string of max. 1	
Although efforts are made to ensure	th 5	i LA	TITUDE_DEG	Integer	decimal degrees of site		character integer value betwee	#
AUTEUR(S) / AUTHOR(S)	6	LA	TITUDE_MIN		decimal minutes of site		and 90	n 0 deg.
Centre d'études nordiques				Double	latitude (latitude of hole			B.
RÉSUMÉ / SUMMARY	7	LAT	TITUDE_DIR	Text			real value text string of max. 1	min.
Les données de ce numéro de Nordic	a				direction latitude		haracter	
					decimal degrees of site		acter	#
The datasets in this issue of Nord	. 8	LON	GITUDE_DEG	Integer	longitude (longitude of h			
CITATION DES DONNÉES / DATA CITAT	C			integer	n)		iteger value between	D
		LONG	GITUDE_MIN		decimal minutes of site	a	nd 180	
CEN 2014. Données environnemental			NODE_WIN	Double	longitude			deg.
CEN 2014. Environmental data from		LONG	TTUR -			re	al value	
CEN 2014. Environmental about		LONGITUDE_DIR	Text	direction		t string of max. 1	min.	
1	11		START	Date	direction site longitude	cha	aracter	
SITE(S)	12	DATE	END	Date	date of site start			m
				Date	date of site end		e in UTC	dd-mmm-yyyy
Nom / Name			83.09396			dat	e in UTC	dd minn-yyyy
Ellesmere Parks Canada (ELLEPAR	)		85.05550				1 A 844	dd-mmm-yyyy
Ellesmere Parks cam								
DESCRIPTION								
		(P)			(m): 0.20			
SITE: Ellesmere Parks Canada ( Profondeurs des / températures	de so.	1 (m) /	Ground temp	erature dep	th range (m). eren			
				_				
					round temperature (Degree	Celsiu	s (°C))	
TYPE: Température moyenne du so	ol (De	egré ce	lsius (°C)) /	Average gi	Conta Samp			
TYPE: Temperature								

### **Data Articles**/ **Reports**

Earth System Science Data EnMAP EnMAP-Flight-Cam F. Möller, A. Liebsche Yearly Oper of the CO<sub>2</sub> S Ketzin, Ger

highly variable between the disciplines but key information for data reuse

### Metadata for data discovery: example DOI Landing Page

### title citation

GFZ Helmholtz Centre Potsbam		GFZ GER	Impressun Neuwoltz Center Porson RMAN RESEARCH CENTRI FOR GEOSCIENCES
Dataset	COS	SC-1 operational report - Operational data sets	Released
	Roberts, Nick; Tsang,	ar; Bjelm, Leif; Almqvist, Bjarne; Berthet, Théo; Conze, Ronald; Gee, David G.; Klonowska, Iw Chinfu (2015): COSC-1 operational report - Operational data sets. GFZ German Research Cen DP.5054.2015	
Data Files This dataset contains files with r ccess. You may download or app following contacts:		Abstract The collisional Orogenv in the Scandinavian Caledonides (COSC) scientific drilling project for mountain building processes in a major mid-Paleozoic orogen in vestern Scandinavia and it modern analogues. The transport and emplacement of subduction-related highgrade contin transition (COTI complexes cont the Baltocendia paldrom and their influence on the units	s comparison with ent-ocean

 Lorenz, Henning Uppsala University, Department of Earth Sci ences, Geophysics henning.lorenz@geo.uu.se

http://cosc.icdp-online.org

Supporting information: Lorenz, H.; Rosberg, J. E.; et al. (2015): COSC-1 operational report Explanator y remarks on the operational data sets. Deutsches GeoForschungsZentrum GEZ doi:10.2312/ICDP.201

(R) All Data Sites 2427 Bytes Holes 15133 Bytes Core Runs 85575 Bytes Core Sections 300426 Bytes Core Boxes 59763 Bytes Core Overviews 61279327 Bytes (R) Lithological Descriptions (R) Sample Request (R) Core Samples take Mud Samples taken 20781 Bytes (R) Multi Sensor Core Logging (R) XRF logging Borehole Measurement Runs 12358 Bytes (R) Borehole Measurement Files (R) Composite Borehole Log Plots

License: CC BY 4.0

References

### related

### work

GFZ Helmholtz-Zentrum POTSDAM

### COSC Consortium download data files

Borehole Measurement Campaigns 4966 Bytes Drilling Time Breakdown per Day 11110 Bytes Drilling Time Breakdown of Tasks 102353 Bytes Drilling Technical Parameter 35538 Bytes Used Drill Bits 2981 Bytes

End of moratorium: /2017-03-01

#### Related Work IsReferencedBy

Lorenz, H.; Rosberg, J. E.; et al. (2015): Operational report about phase 1 of the collisional orogeny in the scandinavian caledonides scientific drilling project (COSC-1). Deutsches GeoForschungsZentrum GFZ doi:10.2312/ICDP.2015.002

Lorenz, H.; Rosberg, J.-E.; et al. (2015): COSC-1 drilling of a subduction-related allochthon in the Palaeozoic Caledonide orogen of Scandinavia. Scientific Drilling doi:10.5194/sd-19-1-2015

and basement will be studied in a section provided by two fully cored 2.5 km deep drill holes. This operational report concerns the first drill hole, COSC-1 (ICDP 5054-1-A), drilled from early May to late August 2014

COSC-1 is located in the vicinity of the abandoned Froå mine, close to the town of Åre in Jämtland, Sweden and was planned to sample a thick section of the Seve Nappe and to penetrate its basal thrust zone into the underlying lower grade metamorphosed allochthon. Despite substantial technical problems, the drill hole reached 2502.8 m driller's depth and nearly 100 % core recovery was achieved. Surprising was the homogeneity of the Seve Nappe rocks, the unexpected thickness of its basal thrust zone (> 500 m) and that the drill hole, therefore, did not penetrate the bottom of the thrust zone. However, lower grade metasedimentary rocks were encountered in the lowermost part of the drill hole together with tens of metres thick mylonites that are, unexpectedly, rich in large garnets,

The drill core was documented on-site and XRF scanned off site. During various stages of the drilling, the borehole was documented by comprehensive downhole logging. This operational report provides an overview over the COSC-1 operations from drilling preparations to the sampling party and describes the available datasets and sample material.

#### Keywords

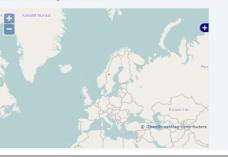
SOLID EARTH, ROCKS/MINERALS/CRYSTALS, geoscientificInformation, caledonides, COSC, deep hydrosphere, dynamics, europe, heat flow, himalaya, ICDP-2011/03, microbiology, norway, orogen, scandes scandinavia, seismic, sweden, earth science

GCMD Science Keywords EARTH SCIENCE > SOLID EARTH > ROCKS/MINERALS/CRYSTALS > METAMORPHIC ROCKS

ROCK FORMATION More Metadata iso19115: view inline / download xml datacite: view inline / download xml dif: view inline / download xml

escidoc: view inline / download xml

Location Latitude: 63,4063 Longitude: 13,203057



### description/ abstract

### **Keywords**

#### More Metadata

iso19115: view inline / download xml datacite: view inline / download xml dif: view inline / download xml escidoc: view inline / download xml

## spatial

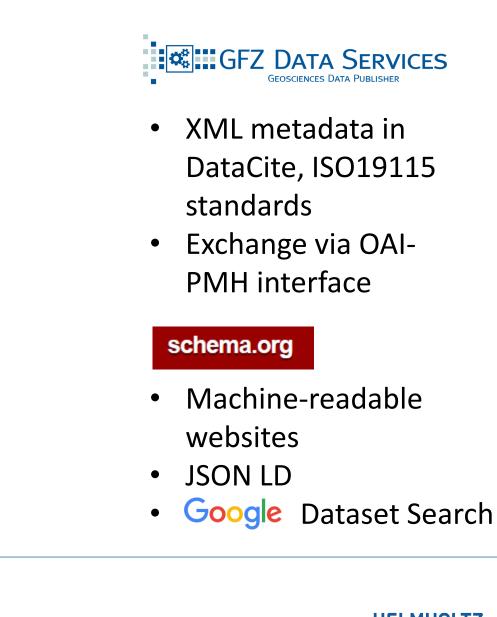
### coverage

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

XML metadata following international standards (DataCite, ISO19115)

### Machine-readable metadata standards





GFZ Helmholtz-Zentrum

# Controlled Vocabularies in the Metadata Editor



NASA Global Change Master Directory (GCMD) Keywords Example: EARTH SCIENCE > SOLID EARTH > TECTONICS > VOLCANIC ACTIVITY > ERUPTION DYNAMICS

ita		Thesaurus	
	Description	Filter on keyword: tecto	10 INSP/0
prope	Here we report the ray a regional structure (le we report results com of Calcare Massiccio smaller cores, 38 mm divided the rock samp hangingwall (HW) ma erties is not recomm Note Ma © cal	Name       Image: Solid Earth Science Keywords       Image: Solid Earth       Image:	
liting	e Keywords is not recommeded o CE > SOLID EARTH > 1	Image: Construction       Image: Construction<	GeoSciML: Geoscience Vocabularies for Linked Data

# Persistent Identifier in data publications



for data, software, cross-references to related work



List of funders with DOIs

ORCID

GF7

Helmholtz-Zentrum

Connecting Research and Researchers

# uniquely identifying persons

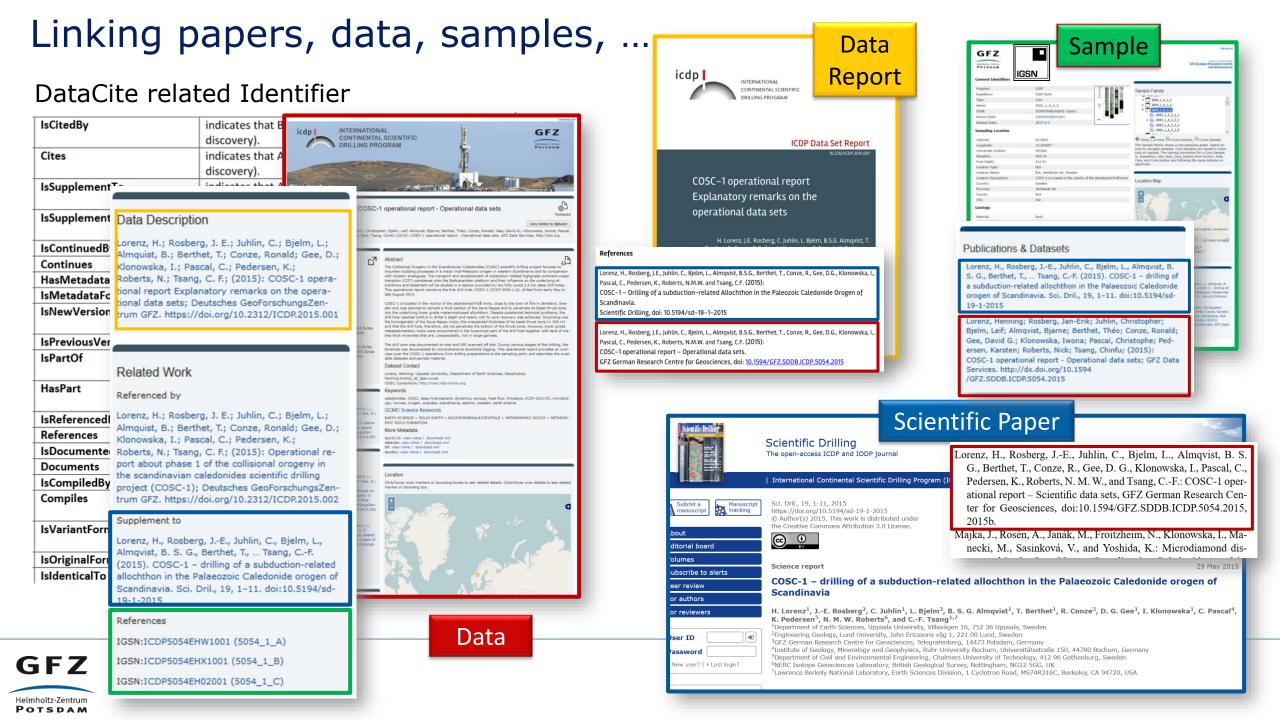


PID for physical samples, cross references to samples underlying measurements

### **ROR** New PID for Institutions







# 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

DataCite Metadata						GI	FZ Data Services
Resource Information							
DOI (will be generated in th	e publishing process)			Yoar		(	
10.5880/GFZ.4.1.2019.005				2019		Supplementary material for	analogue experiments on the interactions of two indenters, and their
Resource Type	Title			Version Language of dataset		suppremental) material for	nplications for curved fold-and-thrust-belts
Dataset	Supplement to 'Synchronization of	great subduction megathrust earthqua	akes: Insights from scale model analysis"	eng			
Licenses and Rights						Cite as: Reiter, Karsten; Kukowski, Nina; Ratschbacher, Lothar indenters, and their implications for curved fold-and-th	Copy citation to clipboard ; Rosenau, Matthias (2016): Supplementary material for analogue experiments on the interactions of two hrust-belts. GFZ Data Services. http://doi.org/10.5880/GFZ.4.1.2016.007
Licence							
CC BY 4.0				<b>@</b>	<b>N</b>	Files	Abstract
Authors (Persons and/or Ins	titutions)					Explanations_Reiter-et-al-2016.pdf 0.5 Mb	This data publication includes animations and figures of eight scaled analogue models that are used to
Lastname	Firstname	Role	Author ID Type Author Identifier (ID)	Affiliation		list-of-files-Reiter-et-al-2016.pdf 232.6 Kb Experimenting.avi 78.7 Mb	investigate the evolution of a curved mountain belt akin to the Pamir and Hindu Kush orogenic system and adjacent Tadjik basin. Crustal deformation is simulated by means of indentation of two basement blocks into a sedimentary sequence and the formation of a curved fold-and-thrust belt.
Rosenau	Matthias		ORCID 0000-0003-1134-5381	GFZ German Research Centre for G		gb70-pictures.pdf 497.1 Kb gb40-3Dview-30-34.avi 5.8 Mb	The experimental set-up has two adjacent rigid indenters representing the basement blocks moving in parallel with a velocity difference (Figure 1). The slow indenter moves with a relative velocity ranging from 40
lorenko	Illia			Università della Svizzera Italiana, L 🤤		gb50-3Dview-30-33.avi 5.7 Mb gb55-3Dview-30-32.avi 6.1 Mb	to 80% of that of the fast one. A layer of quartz sand in front of the indenters, 1 by 1 meter in size and 1.5 cm thick, represents the sedimentary basin infill. A basal detachment layer is made up of low-friction glass
Corbi	Fabio		ORCID 0000-0003-2662-3065	Roma Tre University, Rome, Italy 🤤		gb60-3Dview-30-29.avi 6.4 Mb	beads or viscous silicone oil representing weak shale or evaporates layers, respectively. The surface evolution by means of topography and strain distribution is derived from 3-D particle image velocimetry (PIV). This
Rudolf	Michael		ORCID 0000-0002-5077-5221	GFZ German Research Centre for G 🤤		gb70-3Dview-30-30.avi 6.0 Mb gb80-3Dview-30-31.avi 5.9 Mb	allows visualizing and analysing the development of the model surface during the complete model run at high
Kornhuber	Ralf			Freie University Berlin, Berlin, Germ 🤤		si60-3Dview-30-38.avi 5.7 Mb	spatio-temporal resolution. All details about the model set-up, modelling results and interpretation can be found in Reiter et al. (2011).
Oncken	Onno		ORCID 0000-0002-2894-480X	GFZ German Research Centre for G 🤤			
				٢		Dataset Description	Keywords
						Supplement to	two indenter tectonics, particle image velocimetry, fold-and-thrust belts, Tajik basin, Pamir, 4D analogue experiments, mountain building, continental collision, sandbox model, digital elevation model, analogue
						Reiter, K., Kukowski, N., & Ratschbacher, L. (2011).	model, EPOS, multi-scale laboratories, analogue models of geologic processes, analogue modelling results, Microsphere > Glassy, Particle Image Velocimetry (PIV), Sand > Quartz Sand, Sandbox > Sandbox (cm
Contact Person(s) / Point of	Contact					The interaction of two indenters in analogue	scale), Sectioning, Silicon, Time lapse camera, crustal setting > upper continental crustal setting,
						experiments and implications for curved fold-and- thrust belts. Earth and Planetary Science Letters,	detachment fault, mountain, oblique slip fault, reverse fault, tectonic and structural features, tectonic proces: > continental_collision, tectonically defined setting > collisional setting, tectonically defined setting >
Author (Lastname, Firstnam	,	Emall	Website	Affiliation		302(1-2), 132-146. doi:10.1016/j.epsl.2010.12.00	foreland setting, thrust fault, wrench fault GCMD Science Keywords
Rosenau, Matthias	Lab Manager	rosen@gfz-potsdam.de	http://www.gfz-potsdam.de/en/section/litho.	GFZ German Research Centre for Geoscie		2	
				9			EARTH SCIENCE > SOLID EARTH > GEOMORPHIC LANDFORMS/PROCESSES > TECTONIC PROCESSES EARTH SCIENCE > SOLID EARTH > TECTONICS > PLATE TECTONICS > FAULT MOVEMENT
Contributors (Persons and/c	ar Institutions)					Related Work	More Metadata
Lastname	Firstname	Role	Contributor ID Contributor Identifier (ID)	Affiliation		References	iso19115: view inline / download xml datacite: view inline / download xml
HelTec - Helmholtz Laborator		HostingInstitution		GFZ German Research Centre for G		Adam, J., Urai, J. L., Wieneke, B., Oncken, O.,	dif: view inline / download xml
Herrec - Heimholtz Laborator	y for f	Hostinginstitution		Grz Gemian Research Centre for G 🥥		Pfeiffer, K., Kukowski, N., Schmatz, J. (2005). Shear localisation and strain distribution during	escidoc: view inline / download xml
						tectonic faulting-new insights from granular-flow	
						experiments and high-resolution optical image correlation techniques. Journal of Structural	
						Geology, 27(2), 283-301. doi:10.1016/j.jsg.2004.08.008	Location
							Click/hover over markers or bounding boxes to see related details. Click/hover over details to see related marker or bounding box.
							Number ) preak vinoyant (Ayydang) 6 Ou 5 Cor
						Find More Research Data	Navony Samorqond Jizzax Gameencraal Outxaan Obsacmb 自给且 克州
						http://bib.telegrafenberg.de/finden/datenbanken/fo rschungsdaten/	Buxoro Viloyati Bunormu Bunormu
FZ							menabat Gvd Hoxuaxou 喀什地区
							Dashaadarvo mooeu

Helmholtz-Zentrum

#### HELMHOLTZ

器什库尔

Точикистон

Hojambaz Denovo Душанбе

# GFZ Metadata Editor (Java Script "translator")

# **Input:** by scientists

### "Special" Features:

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

DOI (will be generated in		process)		Publisher		Year
10.5880/GFZ.1.4.2016.001				GFZ Data Services		2016
Resource Type	Title				Language	of datase
Dataset	Supplemen	nt to: The New Wo	rld Atlas of Artificial Night Sk	y Brightness	eng	
Licenses and Rights						
Licence						
Please contact the authors	for a licence ag	greement				
Authors (Persons and/or I	nstitutions)					
Author (Lastname, Firstn	ame)	Role	Author ID Typ	e Author Identifier (ID)	Affiliation	
Cinzano, Pierantonio					ISTIL - Istituto di Scienza e Tecnolo	-
Duriscoe, Dan					National Park Service, U.S. Depart	
Kyba, Christopher C. M.					GFZ German Research Centre for	
Elvidge, Christopher D.					Earth Observation Group, NOAA N	
Baugh, Kimberly					Cooperative Institute for Research	
Portnov, Boris					Department of Natural Resources a	
Rybnikova, Nataliya A.					Department of Natural Resources	
Furgoni, Riccardo					ISTIL - Istituto di Scienza e Tecnolo	ogia (
Contact Person(s) / Point	of Contact					
night						
radiative transfer			Se	lect Region (click left mouse and d	drag)	
Suomi NPP				1 a martine	V D V and	
Sky Quality Meter			nig grei	Karte Satellit tes Dänemark ich	Litauen	
Temporal and Spatial C	overage (The	EDIT-symbol to	o the left provides vi	Medenande	Polen Weißrussland	
Latitud	e	Lon	gitude	Belgien Deutschland	the second state	
Min	Max	Min	Max	Tschechisch Republik	he Slowakei Ukraine	1
	5.842428	2.8710901	43.124996	Österreich	Moldawien	1
44.045486 5			100	Frankreich	Ungarn	

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









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<sup>1</sup>, Author-2<sup>2</sup> ...
- 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. <u>International Geo Sample Numbers?</u> 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

 $\rightarrow$  Internal part of metadata

 $\rightarrow$  transferred to other portals as machine-readable XML

 $\rightarrow$  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<sup>3</sup>) - 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 (<u>https://github.com/PennStateRockandSedimentMechanics/xlook</u>).

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

# 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, 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
				Python script for analysing and plotting friction and time series data (Mohr plot,
Scripts	978 kb	RSTanalysis	.py	histograms, shear curves)
		VSTanalysis	.py	Python script for analysing and plotting VST data
		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
Data files	2932 kb	426-01_GFZ_quartzsandG23_peak	.txt	Pairs of normal stress and corresponding shear strength for peak friction
butumes		426-01_GFZ_quartzsandG23_reactivation	.txt	Pairs of normal stress and corresponding shear strength for reactivation friction
				Visualization of time series data (shear curves): Shear stress vs. shear displacement
		426-01_GFZ_quartzsandG23_ts	.pdf	for 18 measurements
				Table of time series data for 18 measurements of shear stress (Pa, collumns 2-19) a
		426-01_GFZ_quartzsandG23_ts	.txt	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	ху	Value of XY in units xy



### 5. References

#### 4. References

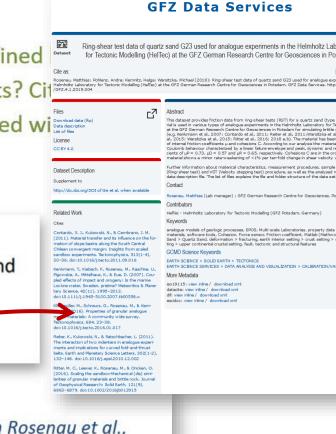
References for dataset and those cited above. If the data follows a standard format defined community, please cite the standard. Does this data incorporate other authors' datasets? Cites should contain a DOI (or URL for reports or websites). The DOI numbers shall be provided with the standard with the standard 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

				XRF analyses: Raw data (major oxides)								
sample ID	<b>IGSN<sup>†</sup></b>	brief sample description	mean depth	$SiO_2$	<b>TiO</b> <sub>2</sub>	$Al_2O_3$	$\mathbf{Fe}_2\mathbf{O}_3$	MnO	MgO	CaO	Na <sub>2</sub> O	
			(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	

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



#### 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, Al....



# 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	



#### iso19115: view inline / download xml GFZ Rataballa Castra COSC-1 operational report - Operational data sets datacite: view inline / download xml dif: view inline / download xml escidoc; view inline / download yml WSM Technical Report 17-02 Manual of the Matlab Script Stress2Grid Moritz Ziegler and Oliver Heidbach icdp cosc-1 operational report Explanatory remarks on the scientific data sets Operational Report about Phase 1 of the Collisional Orogeny in the Scandinavian Caledonides scientific drilling project (COSC-1)

HELMHOLTZ



#### 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





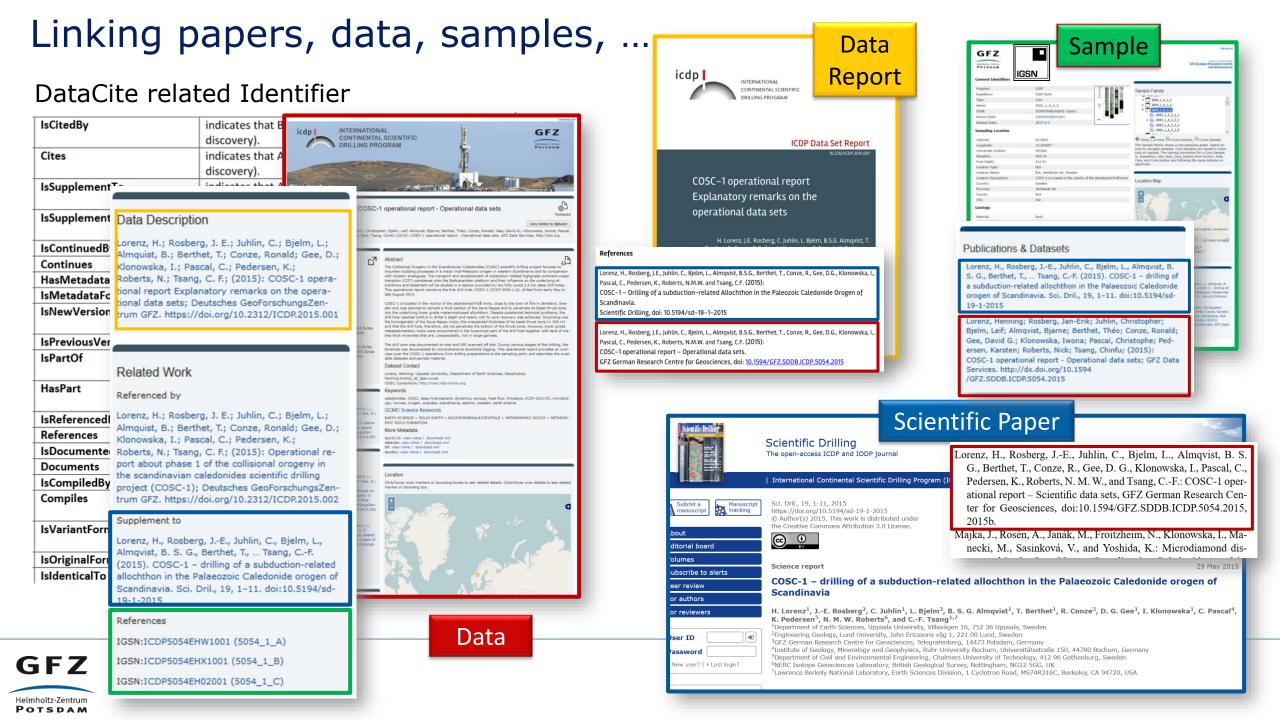
- First data journal, launched in 2008
- *"*international*,* interdisciplinary journal for the publication of articles on original research data"
- No interpretation of the data!
- In 2020:
  - more than 500 peer-reviewed descriptions of easilyand freely-accessible data products from
  - more than 4000 data providers
  - archiving their products at more than 100 data centres
  - IF = 10.951

#### Break – Questions?

Next: Scholix: Incresasing the transparency of research products



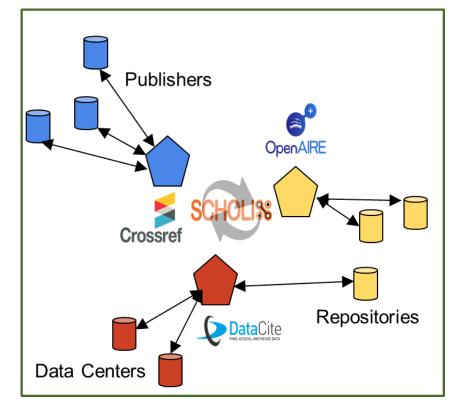
HELMHOLTZ





# Scholix: linking data with papers





Credit: Wouter Haak, Elsevier

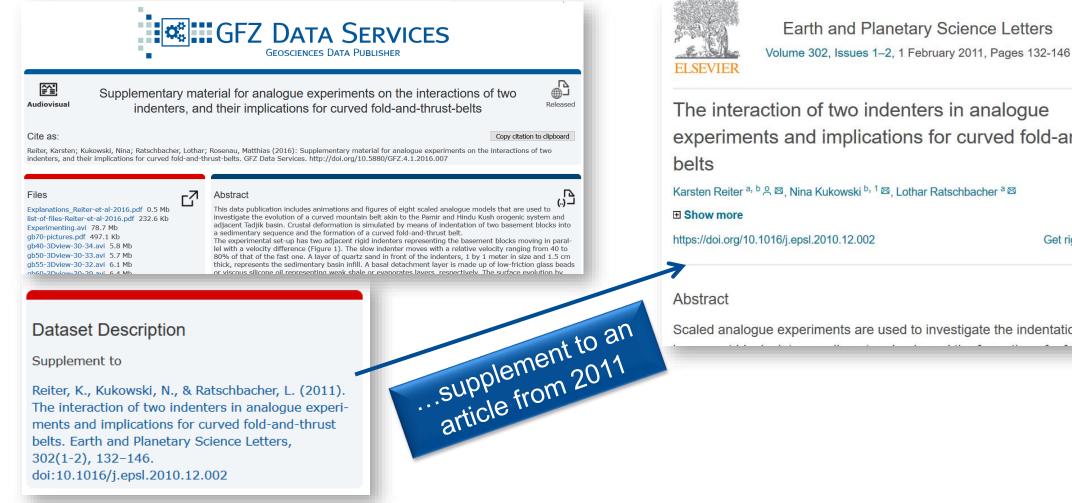
# SCHOLI%

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





# The classical approach (before Scholix)



The interaction of two indenters in analogue experiments and implications for curved fold-and-thrust

Karsten Reiter <sup>a, b</sup> <sup>A</sup> <sup>⊠</sup>. Nina Kukowski <sup>b, 1</sup> <sup>⊠</sup>. Lothar Ratschbacher <sup>a</sup> <sup>⊠</sup>

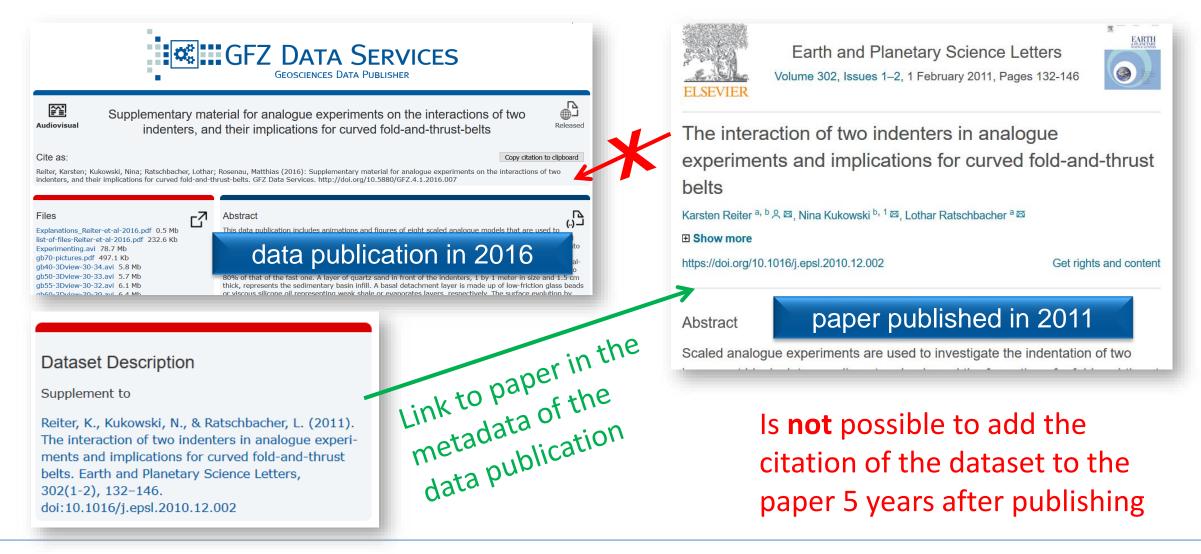
Get rights and content

Scaled analogue experiments are used to investigate the indentation of two



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

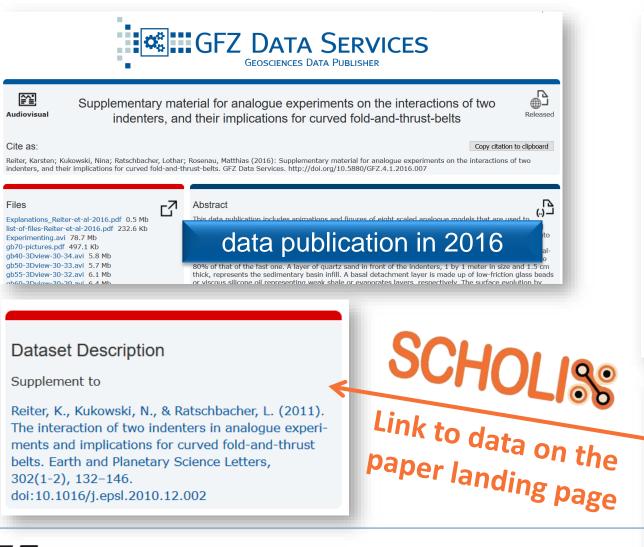
# The classical approach (before Scholix)





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





Earth and Planetary Science Letters Volume 302, Issues 1–2, 1 February 2011, Pages 132-146



The interaction of two indenters in analogue experiments and implications for curved fold-and-thrust belts

Karsten Reiter <sup>a, b</sup> <sup>A</sup> <sup>⊠</sup>, Nina Kukowski <sup>b, 1</sup> <sup>⊠</sup>, Lothar Ratschbacher <sup>a</sup> <sup>⊠</sup>

**B** Show more

https://doi.org/10.1016/j.epsl.2010.12.002

Get rights and content

GFZ

Helmholtz Centre

Abstract

#### paper published in 2011

new

Scaled analogue experiments are used to investigate the indentation of two

#### Research data for this article

GFZ Data Services Geosciences data



Data associated with the article:

Supplementary material for analogue experiments on the interactions of two indenters, and their implications for curved fold-and-thrust-belts *¬* 



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