LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

TIB

DOIs for Research Data

Open Science Days 2017, 16.-17. Oktober 2017, Berlin

Britta Dreyer, Technische Informationsbibliothek (TIB)
http://orcid.org/0000-0002-0687-5460

Scope

1. DataCite Services
2. Data Citation
3. Connecting Scholarly Output
4. Scholarly Link Exchange (Scholix)
5. Event Data
6. ORCID Integration

History

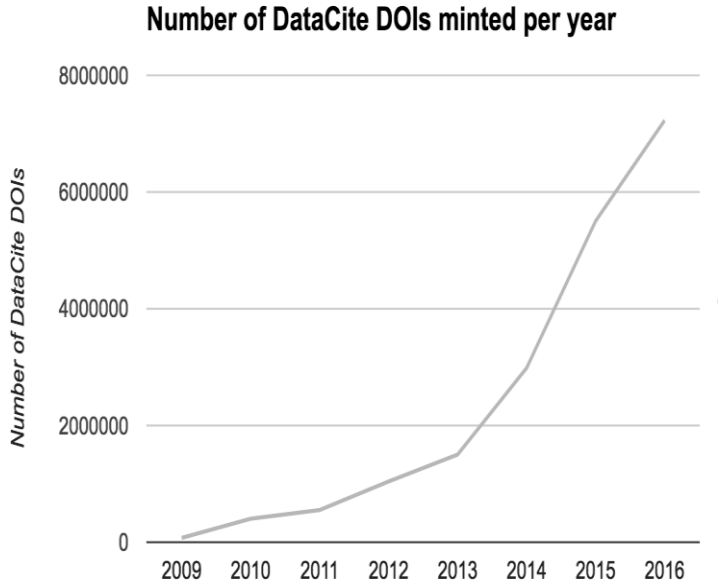
- In 2004 TIB became the global agent for the registration of data DOIs
- The first data set with a DOI from the World Data Center for Climate (WDCC) at DKRZ available on the Internet 2004-03-18:

DOI: 10.1594/WDCC/EH4_OPYC_SRES_A2

EH4_OPYC_SRES_A2_ACLCAC10	dataset
EH4_OPYC_SRES_A2_ACLCAC100	dataset
EH4_OPYC_SRES_A2_ACLCAC1000	dataset
EH4_OPYC_SRES_A2_ACLCAC150	dataset
EH4_OPYC_SRES_A2_ACLCAC200	dataset

DataCite - A quick snapshot

- German charitable association founded 2009
- 50 members worldwide
- > 1300 data centres
- > 10 million DOIs created
- More than 8 million resolutions/month



DataCite
FIND, ACCESS, AND REUSE DATA

DataCite - Mission

DataCite is the leading global provider of DOIs for research data, enabling users to register, find, use, connect and track research data.

DataCite
FIND, ACCESS, AND REUSE DATA

Register DOI

All Data Centers :

Resource Types	
Dataset	3,659,728
Text	2,307,327
Image	963,830
Collection	434,880
Other	352,965
Physical object	72,587
Software	42,034
Audiovisual	26,374
Event	7,667
Film	1,540
Sound	1,210
Model	815
Interactive resource	621
Workflow	270
Service	38

Total Works: 9,143,720

Max Planck Gesellschaft:

Resource Types	
Text	1,522
Dataset	350
Collection	53
Image	17
Software	12
Audiovisual	7
Physical object	5
Workflow	4
Other	3

Total Works: 2,067

https://search.datacite.org

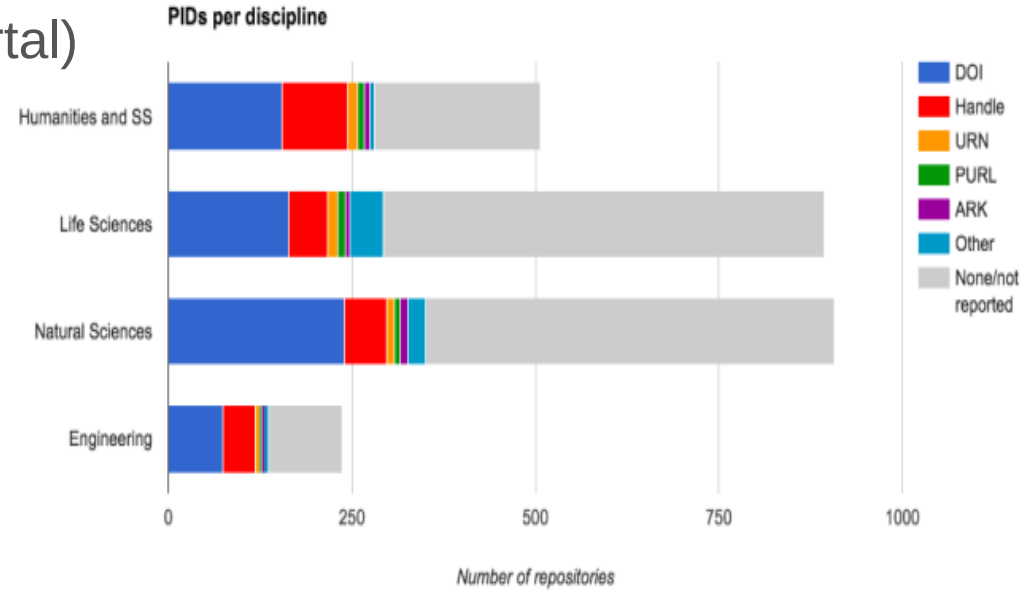
DataCite
FIND, ACCESS, AND REUSE DATA

Find a Repository for your Research Data

- 750 data centres in the re3data registry assign DataCite DOIs to data.

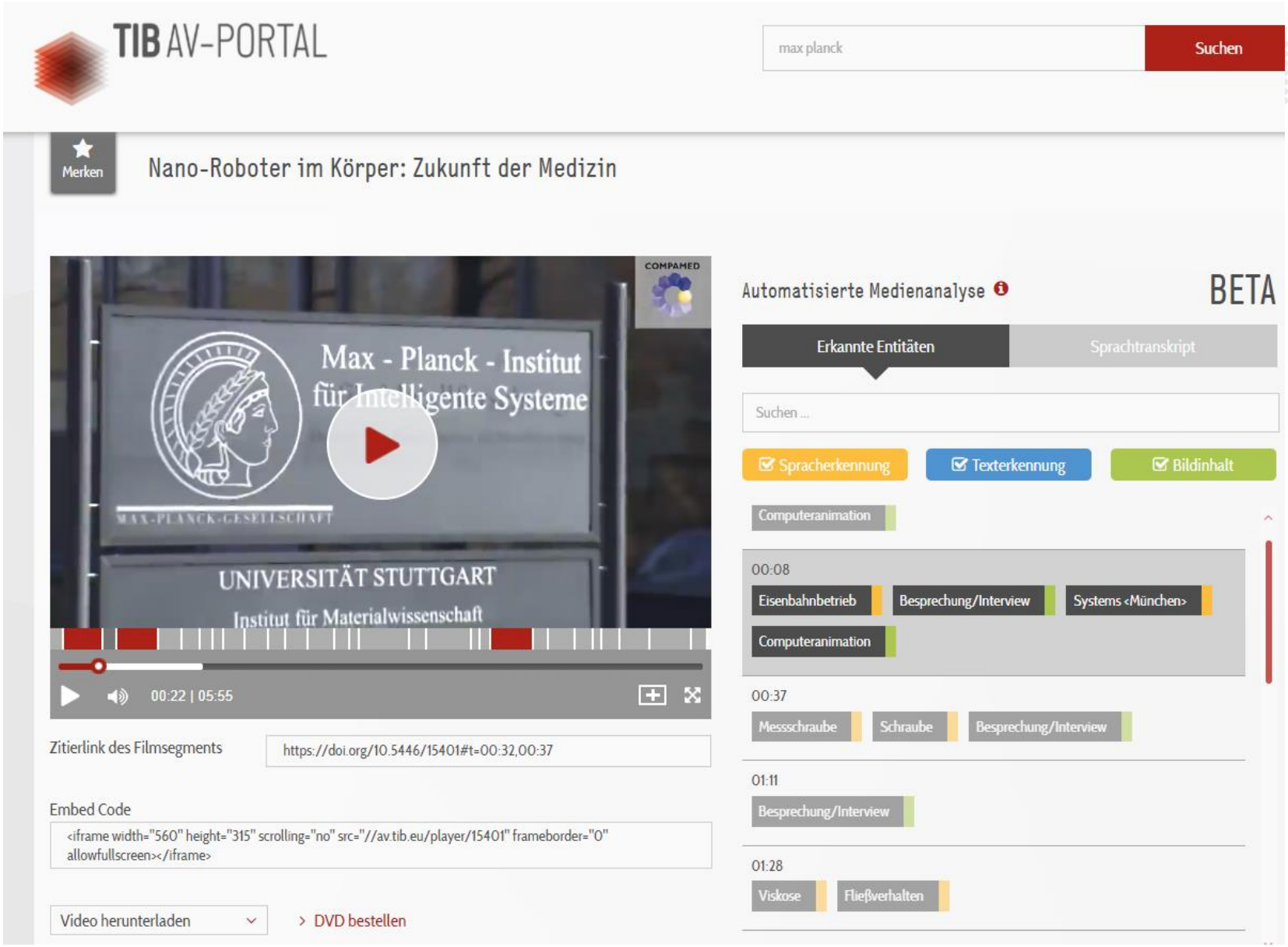
Types:

- Multidisciplinary (e.g. Figshare)
- Discipline specific (e.g. PANGAEA)
- Institution specific (e.g. HEP Data from CERN)
- Resource specific (TIB AV-Portal)



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

https://www.datacite.org/re3data.html


TIB AV-PORTAL
max planck
Suchen
TIB
Merken
Nano-Roboter im Körper: Zukunft der Medizin
COMPAMED
Max - Planck - Institut
für Intelligente Systeme
MAX-PLANCK-GESELLSCHAFT
UNIVERSITÄT STUTTGART
Institut für Materialwissenschaft
00:22 | 05:55
Zitierlink des Filmsegments
https://doi.org/10.5446/15401#t=00:32,00:37
Embed Code
<iframe width="560" height="315" scrolling="no" src="//av.tib.eu/player/15401" frameborder="0" allowfullscreen></iframe>
Video herunterladen
DVD bestellen
Automatisierte Medienanalyse
BETA
Erkannte Entitäten
Sprachtranskript
Suchen ...
Spracherkennung
Texterkennung
Bildinhalt
Computeranimation
00:08
Eisenbahnbetrieb
Besprechung/Interview
Systems <München>
Computeranimation
00:37
Messschraube
Schraube
Besprechung/Interview
01:11
Besprechung/Interview
01:28
Viskose
Fließverhalten

Granularity – Recommendations (International DOI Foundation)

“A DOI name can be assigned to any object, regardless of the extent to which that object might be a component part of some larger entity.

DOI names can be assigned at any desired degree of precision and granularity that a registrant deems to be appropriate.”

DOI Handbook (https://www.doi.org/doi_handbook/2_Numbering.html#2.3.2)

Granularity – Recommendations DataCite/da|ra

1. **Citation:** The current citation and research practices among the client's user community: what is likely to be cited?
2. **The use of data:** The needs of various stakeholders: how will funders/publishers/ administrators etc. use the data?
3. **The type of resource**: for example a complex dataset may require a more granular identifier structure than a document or image file.
4. **Sustainability:** The client must be able to maintain each item with a DOI name in accordance with DataCite client responsibilities.

Recommendations DataCite/da|ra (2)

If necessary:

- Several DOI names for the different granularity level

and

- Connect Metadata record element <relatedidentifier> with *“isPartOf/hasPart"*

Rauber et al, Data Citation of Evolving Data: Recommendations of the Working Group on Data Citation (WGDC), 2017, RDA, https://b2share.eudat.eu/records/ead2dc65f599497f81cf403b97fcfcb0

DataCite
FIND, ACCESS, AND REUSE DATA

Granularity Example

PANGAEA®
Data Publisher for Earth & Environmental Science

Data Description

Citation: **WOCE Hydrographic Programme, WHP (2002):** Hydrochemistry measured on water bottle samples during Ryofu Maru cruise 49RY9407_1 on section P09. doi:10.1594/PANGAEA.837292

Related to: **WOCE (2002):** World Ocean Circulation Experiment, Global Data, Version 3.0. *WOCE International Project Office, WOCE Report, Southampton, UK; p U.S. National Oceanographic Data Center, Silver Spring*, **180/02**, DVD-ROM

Further details: WHP cruise summary information of section P09 (WOCE)

Project(s): **World Ocean Circulation Experiment** (WOCE)

Coverage: *Median Latitude:* 25.122751 * *Median Longitude:* 137.036789 * *South-bound Latitude:* 13.990700 * *West-bound Longitude:* 136.954700 * *North-bound L* 34.250000 * *East-bound Longitude:* 137.448200

Date/Time Start: 1994-07-08T16:58:00 * *Date/Time End:* 1994-07-26T21:28:00

Minimum DEPTH, water: 0.00 m * *Maximum DEPTH, water:* 5589.70 m

Event(s): **49RY9407_1/1-1** * *Latitude:* 34.250000 * *Longitude:* 137.002300 * *Date/Time:* 1994-07-08T16:58:00 * *Elevation:* -150.0 m * *Campaign:* 49RY9407_1 Ryofu Maru * *Device:* CTD/Rosette (CTD-RO) * *Comment:* Section P09

49RY9407_1/10-1 * *Latitude:* 32.816500 * *Longitude:* 137.000300 * *Date/Time:* 1994-07-10T02:30:00 * *Elevation:* -3895.0 m * *Campaign:* 49RY9407_ Ryofu Maru * *Device:* CTD/Rosette (CTD-RO) * *Comment:* Section P09

49RY9407_1/10-2 * *Latitude:* 32.806800 * *Longitude:* 137.077700 * *Date/Time:* 1994-07-10T04:16:00 * *Elevation:* -4015.0 m * *Campaign:* 49RY9407_ Section P09

DataCite
FIND, ACCESS, AND REUSE DATA

DataCite Metadata Schema

- relatively simple schema, maintained by DataCite members
- flexible with regards to resource type
- support for collections
- multiple relation types for related content

Findable, Accessible, Interoperable, and Re-usable = FAIR

DataCite
FIND, ACCESS, AND REUSE DATA

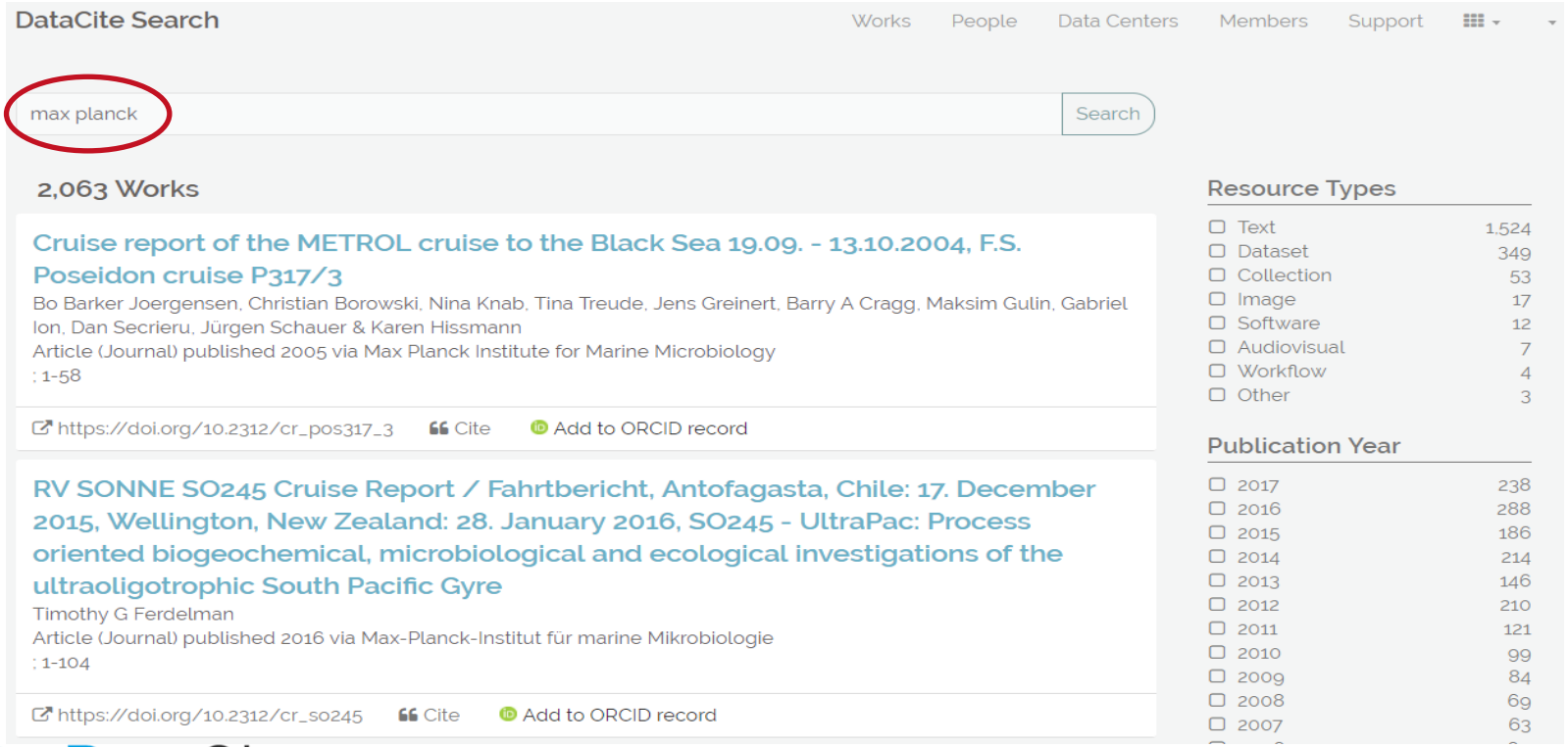
http://schema.datacite.org/

https://www.force11.org/group/fairgroup/fairprinciples

Find Resources

Find resources with DOIs and associated information.

- DataCite API
- DataCite Search
- DataCite OAI-PMH



DataCite
FIND, ACCESS, AND REUSE DATA

MPG Research Data Publications

Data Centers

TIB Hannover	697
Max Planck Digital Library	674
World Data Center for Climate	105
Joint Research Centre's Institute of Energy	79
ZENODO - Research. Shared.	62
Max Planck Institute for Human Development	59
University of Cambridge	32
PANGAEA - Data Publisher for Earth & Environmental Science	31
figshare Academic Research System	30
TIB KMO / FLOWWORKS GmbH	19
GESIS Leibniz Institute for the Social Sciences	18
Coherent X-ray Imaging Data Bank	15
Columbia University Libraries/Information Services (CUL/IS)	15
Humboldt-Universität zu Berlin Arbeitsgruppe Elektronisches Publizieren am Computer- und Medienservice (CMS)	15
ResearchGate	15

Use Resources

Access to the content that was registered, and information how it can be used.

- License information in metadata
- Most of the content available without restrictions
- Directly access content via Content Resolver Service


```
- <rightsList>
    <rights rightsURI="http://creativecommons.org/licenses/by/3.0/">Creative Commons Attribution 3.0 Unported (CC-BY)</rights>
  </rightsList>
```

https://support.datacite.org/docs/datacite-content-resolver

Citation Recommendations

1. All datasets intended for citation must have a **globally unique persistent identifier** that can be expressed as unambiguous URL.

2. Persistent identifiers for datasets must support **multiple levels of granularity**, where appropriate.

3. This persistent identifier expressed as URL must **resolve to a landing page** specific for that dataset.

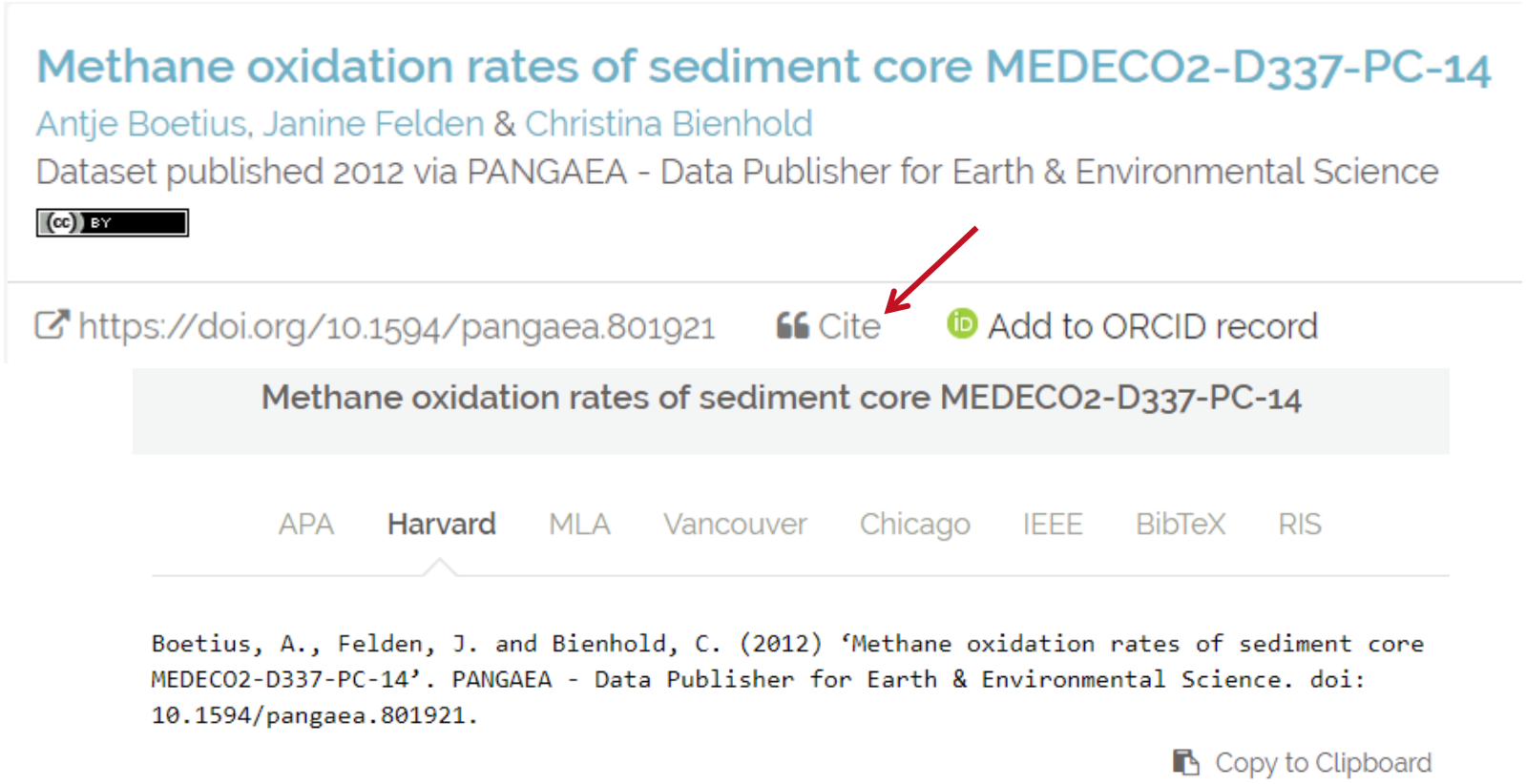
4. The persistent identifier must be embedded in the landing page in **machine-readable format**.

5. The repository must provide **documentation and support** for data citation.

A Data Citation Roadmap for Scholarly Data Repositories (2017): M. Fenner and Mercè Crosas. LIBER Webinar

Cite Resources

1. DataCite Search



https://search.datacite.org/works?query=max+planck&resource-type-id=dataset&data-center-id=tib.mpdl#

DOI Landing Page - Best Practice

Deep-water amphipods from mooring time-series FEVI7 in 800 m depth at AWI HAUSGARTEN

Angelina Kraft, Eduard Bauerfeind, Eva-Maria Nöthig, Michael Klages, Agnieszka Beszczynska-Möller & Ulrich Bathmann

Dataset published 2013 via PANGAEA - Data Publisher for Earth & Environmental Science

Is part of (1)

References (1)

https://doi.org/10.1594/PANGAEA.809438 Cite

PANGAEA.
Data Publisher for Earth & Environmental Science

Citation:

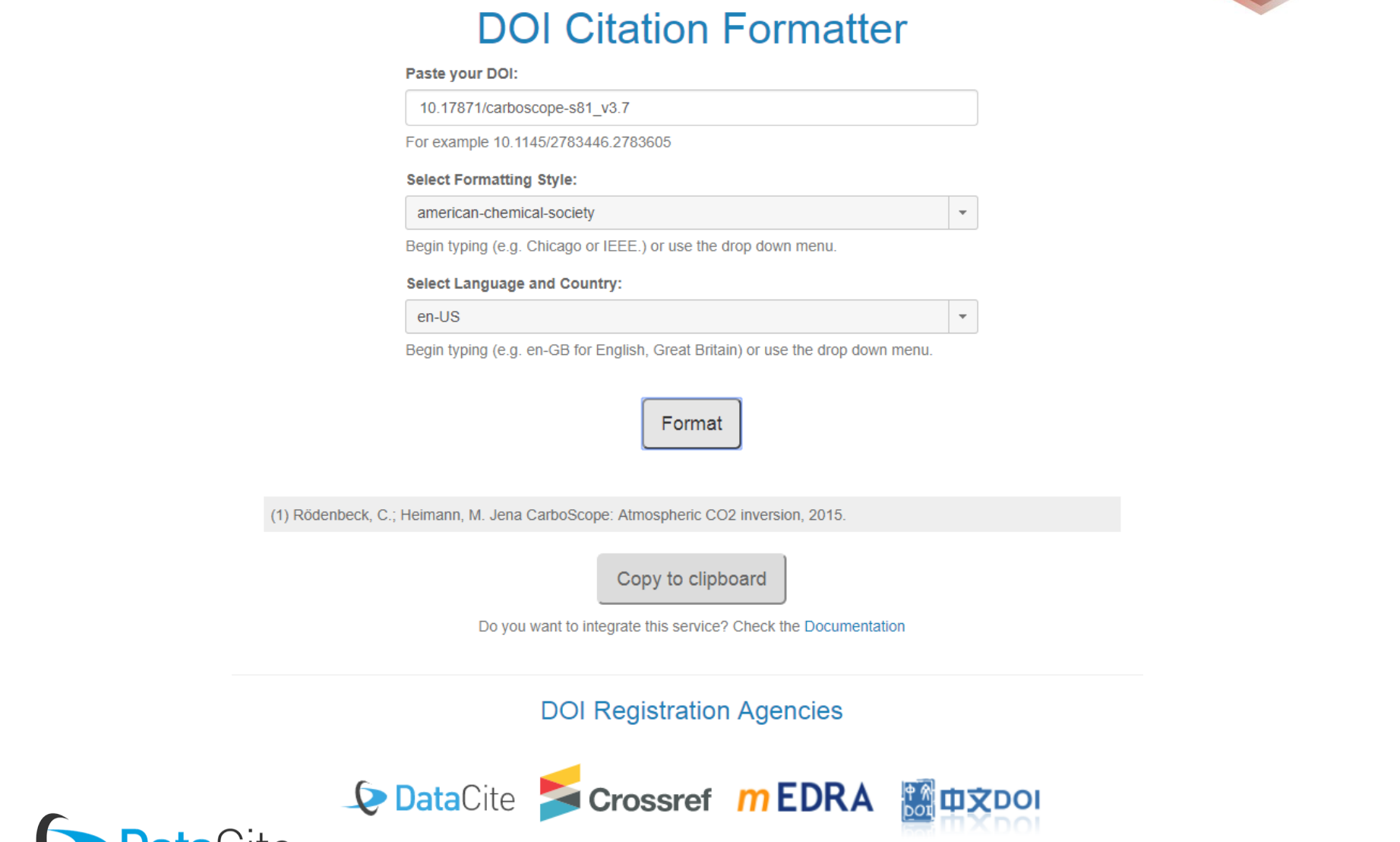
Kraft, Angelina; Bauerfeind, Eduard; Nöthig, Eva-Maria; Klages, Michael; Beszczynska-Möller, Agnieszka; Bathmann, Ulrich (2013): Deep-water amphipods from mooring time-series FEVI7 in 800 m depth at AWI HAUSGARTEN. doi:10.1594/PANGAEA.809438,

In supplement to: **Kraft, Angelina; Bauerfeind, Eduard; Nöthig, Eva-Maria; Klages, Michael; Beszczynska-Möller, Agnieszka; Bathmann, Ulrich (2013):** Amphipods in sediment traps of the eastern Fram Strait with focus on the life-history of the lysianassoid Cyclocaris guilelmi. *Deep Sea Research Part I: Oceanographic Research Papers*, **73**, 62-72, doi:10.1016/j.dsr.2012.11.012

Always quote above citation when using data! You can download the citation in several formats below.

RIS Citation | BibTeX Citation | Text Citation | Facebook | Twitter | Google+ | Show Map | Google Earth

Cite Resources – Citation Formatter



DataCite
FIND, ACCESS, AND REUSE DATA

https://citation.crosscite.org/

Connecting scholarly output

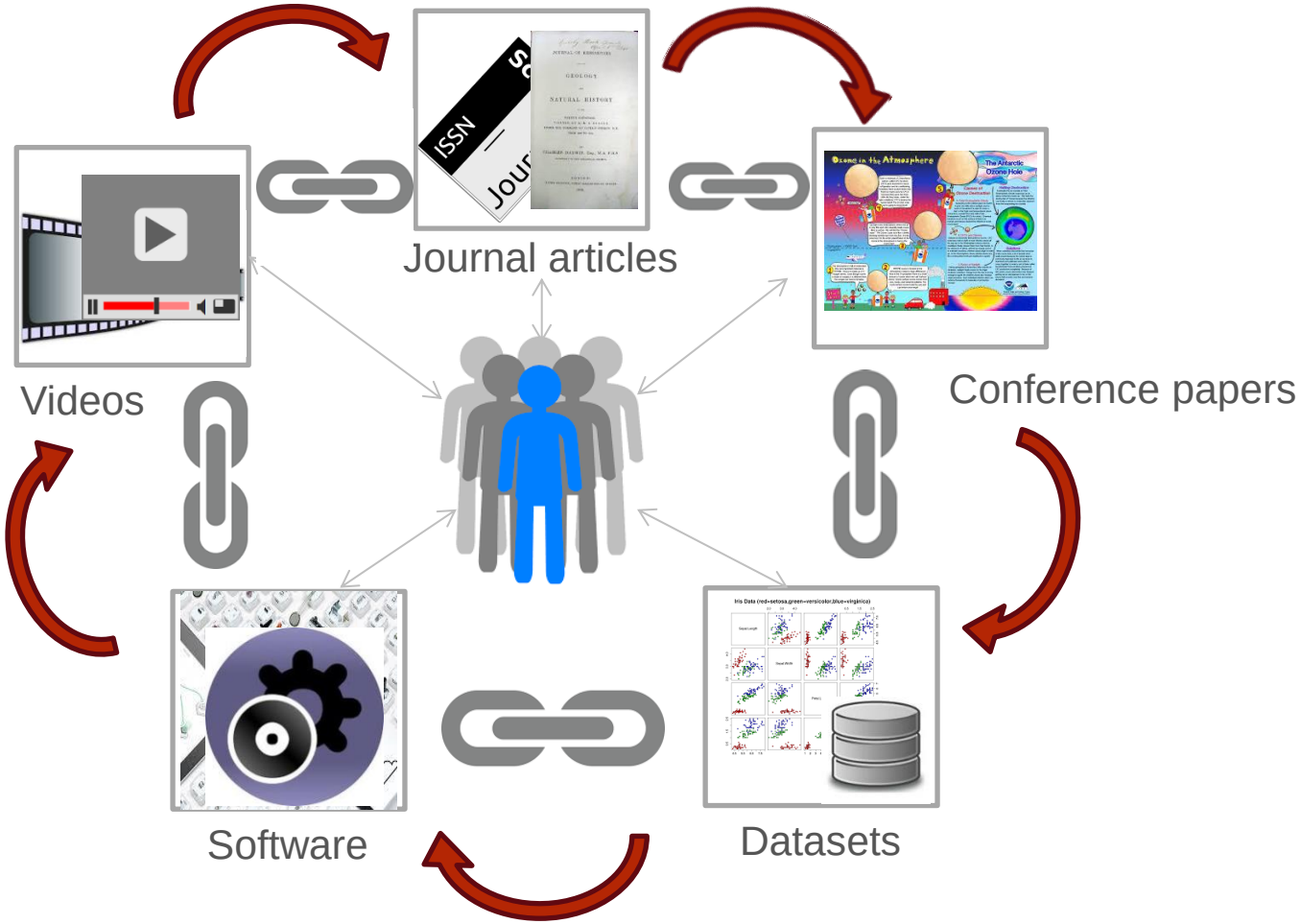
Connect resources, which have a DataCite DOI, to other resources - for example:

- ✓New versions of the same dataset,
- ✓Collections of related datasets,
- ✓or articles citing the dataset. …….AND

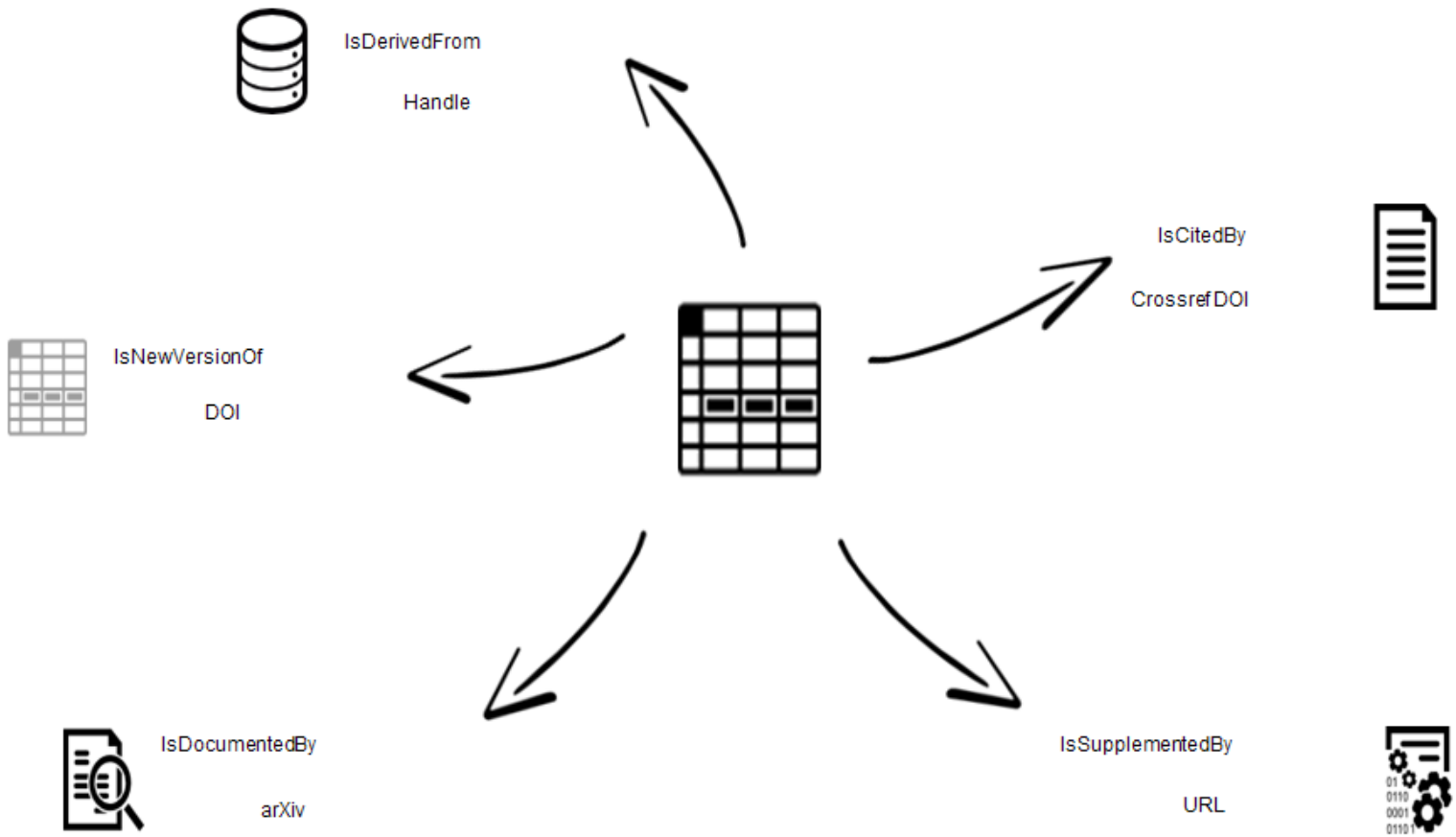
→linking these resources to the people and organizations (coming next) who have contributed to their generation.

DataCite
FIND, ACCESS, AND REUSE DATA

Seamless Integration across the research life cycle

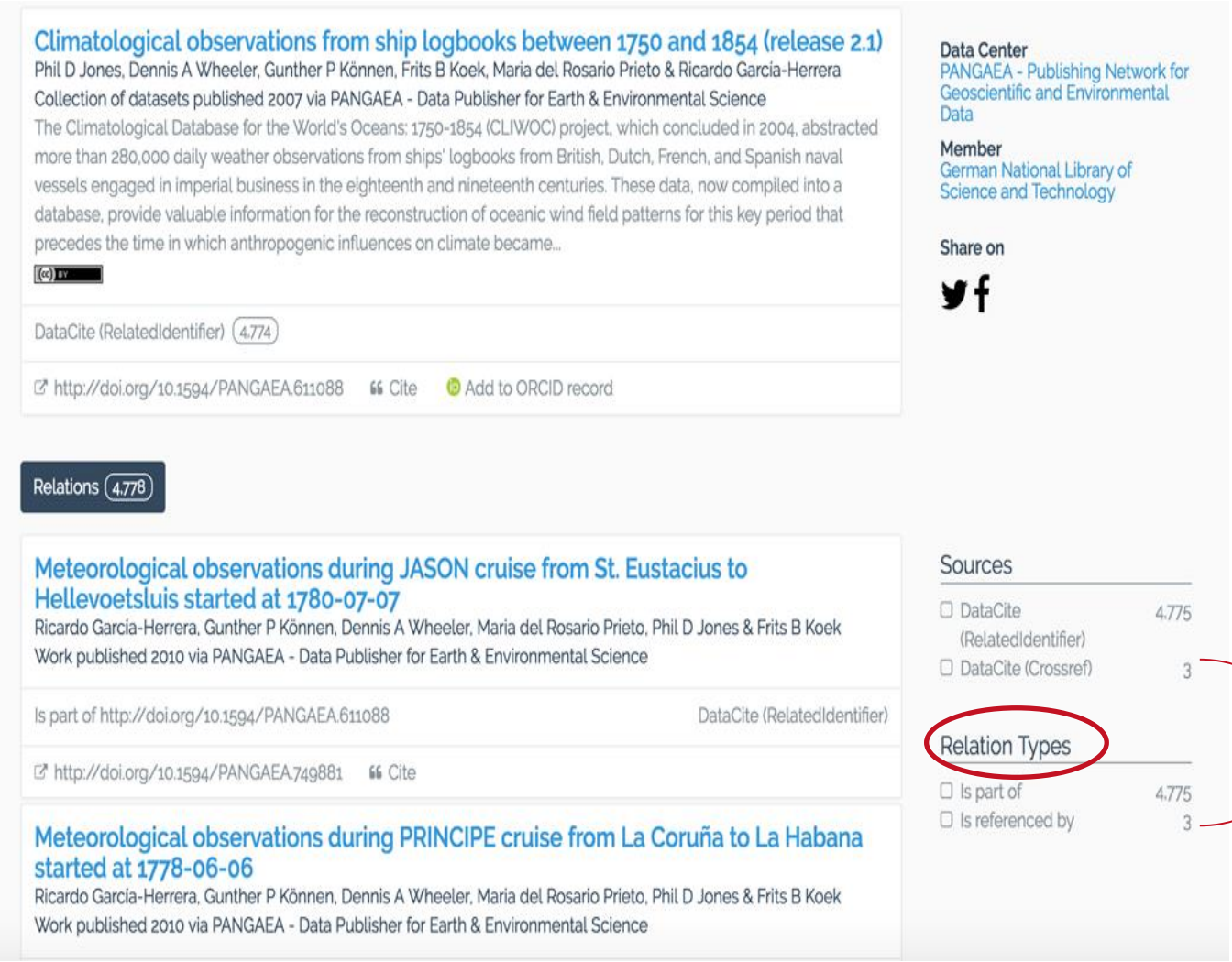


Relation Types



DataCite
FIND, ACCESS, AND REUSE DATA

Related Resources



3 article reference lists

Scholerly Link Exchange (Scholix)

- RDA/WDS Scholix Working Group 2016
- Group aims to enable a comprehensive global view of the links between scholarly literature and data, and doing this by establishing:
  1. An interoperability framework with guidelines and standards
  2. Enabling infrastructure
  3. Outreach and support for communities of practice
- Guidelines finalized by the end of 2017

Scholix is supported by the following organizations:

Minimal Information – Heavy Use of PIDs and their Metadata

- Standardized information exchange will potentially include all data centers and publishers

Link Information Package		
Link Publication Date (1) **Link Provider (1..N)** **Relationship Type (1)** License URL (0..1)		
Source Object		Target Object
Object Identifier (1) **Object Type (1)** Object Title (0..1) Object Publisher (0..1) Object Creator (0..N) Object Publication Date (0..1)		**Object Identifier (1)** **Object Type (1)** Object Title (0..1) Object Publisher (0..1) Object Creator (0..N) Object Publication Date (0..1)

- The initial group of Scholix hubs includes:

1. Crossref, working with publishers
2. DataCite, working with data centers
3. OpenAIRE, working with institutional repositories

http://www.scholix.org/

Linking Data and Articles
Research - Conceptual Model

- **Linkage as Triples**. In the form *subject-predicate-object*, consistent with the Resource Description Framework (RDF) data model.
- **Describing the relation**. Additional information such as relation type (e.g. A *is new version of* B) and provenance.
- **Persistent Identifiers as HTTP URIs**. This makes them actionable, and compatible with the RDF data model.
- **Centralized infrastructure for persistent identifier linking**. Provided for example by ORCID and DataCite, facilitating discovery.

Statistics of provided Links

Content provider	Contributed links	Referred objects	Referred publications	Referred datasets	Referred objects of unknown typology
	36	33	18	18	0
OpenAIRE	29284	17535	14125	14642	517
RCSB	175648	131786	87713	87824	111
Pangaea	856181	238244	112847	525880	217454
Datasets in Datacite	33762553	2998094	886400	31408133	1468020
Cambridge Crystallographic Data Centre	1276152	906001	634785	638885	2482
3TU.Datacentrum	432	351	0	216	216
ICPSR	266804	70110	133402	133402	0
IEDA	1474	921	603	794	77
Thomson Reuters	48592	28867	23714	24878	0
PubMed	1032816	508456	516408	516408	0
Springer Nature	56510	35289	28237	28255	18
Elsevier	138972	90007	69486	69486	0
Australian National Data Service	19552	12411	9775	9777	0
IEEE	94	59	47	47	0
Crossref	0	392837	0	0	0

Track Re-Use

Data Citations

Started

- DOI resolutions
- Repository usage stats

Planned

- Wikipedia
- Twitter

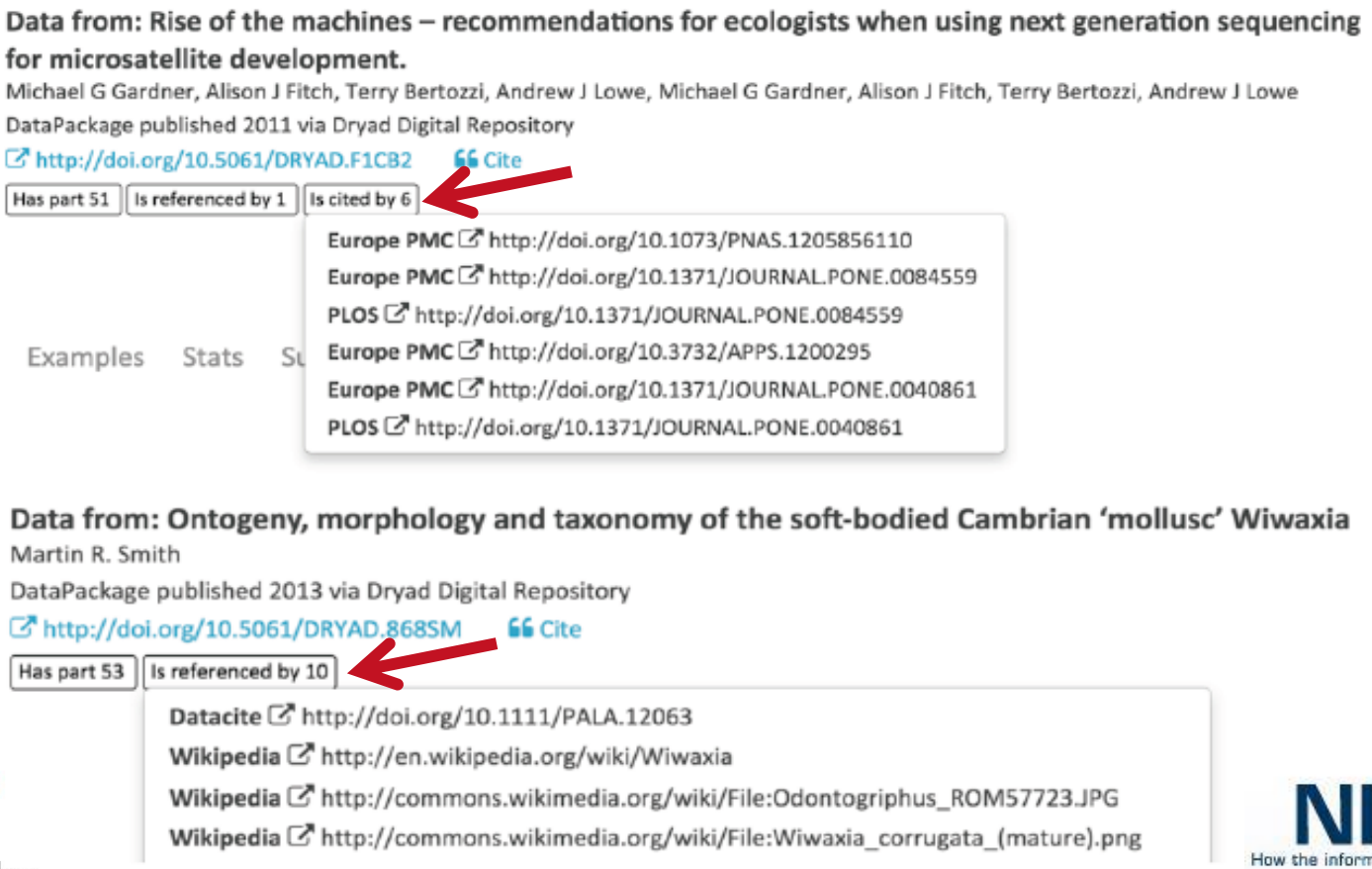
→ **Forward Data Citations to Data Centers via:**

- DataCite Search
- Notifications

DataCite
FIND, ACCESS, AND REUSE DATA

Data Level Metrics (DLM)

- Collects events found via the relatedIdentifier and nameIdentifier attributes of DataCite Metadata



Making Data Count: Promoting a New Normal

….will develop and deploy the social and technical infrastructure necessary to elevate data to a first-class research output.

1. Develop and publish a **COUNTER code** of practice recommendations for how data usage is measured
2. **Deploy central online hub** for acquiring, managing and presenting DLMs
3. **Integrate new sources and clients of aggregated metrics**

Data Usage Stats	Data Citations
DataONE Federation	PubMed Central
DOI resolver logs via DataCite	Crossref
Institutional repository (Re3data)	Europe PMC

4. Encourage growth and uptake of DLMs through an engaged stakeholder community

Project Details

Project: **Making Data Count: Promoting a New Normal**

ALFRED P. SLOAN FOUNDATION

Funding: **750 K for 2 years** (June 2017 – June 2019)

Partners: **CDL, DataCite and DataONE**

- Collaborate with other data metrics initiatives: Crossref Event Data, JISC IRUS UK, NISO Altmetrics working group, RDA/WDS Scholix, etc.
- Start: RDA BoF with relevant stakeholders

Software Repositories

Matdcal

Kirk Bevan

Simulation Tool published 2015 via nanoHUB

Non-equilibrium Green's Function Density Functional Theory Simulator

https://doi.org/10.4231/D3JH3D36M Cite

By Kirk Bevan

McGill University

Non-equilibrium Green's Function Density Functional Theory Simulator

Launch Tool

Version 3.0 - published on 09 Jan 2015

doi:10.4231/D3JH3D36M cite this

Advanced-Expert

380 users, detailed usage

18 users in 2 classes

1 Citation(s)

1 question (Ask a question)

0 review(s) (Review this)

0 wish(es) (New Wish)

Share: ...

Citations Non-affiliated (1) | Affiliated (0)

Non-affiliated authors

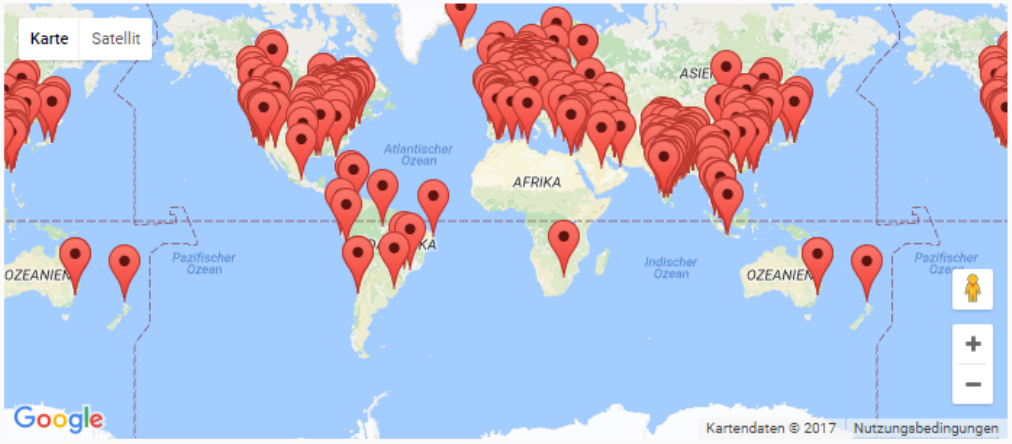
Yap Siong (2011), "Molecular Electronics As A Future Electronic Device": pg. -.

BibTex EndNote

User Details

World usage

Location of all "Matdcal" Users Since Its Posting

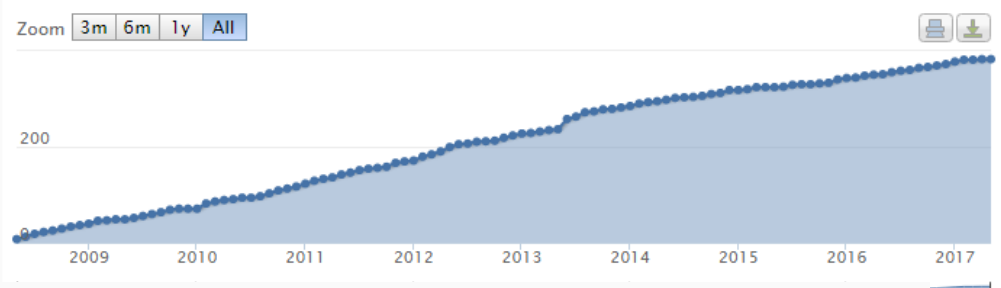


Monthly | Yearly | Cumulative

Simulation Users

380

Apr 2017



Users By Organization Type

Type	Users
Educational - University	270 (71.05%)
Unidentified	81 (21.32%)
Industry	13 (3.42%)

Users by Country of Residence

Country	Users
UNITED STATES	68 (34%)
INDIA	50 (25%)
CHINA	22 (11%)
GERMANY	12 (6%)

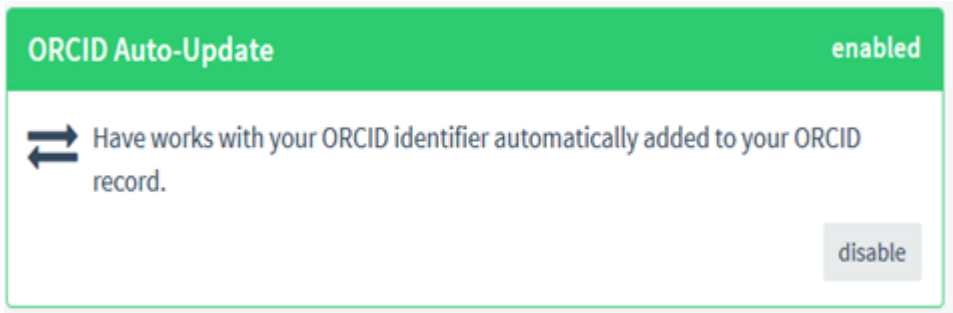
ORCID in a Nutshell

- ORCID provides a persistent digital identifier
- Distinguishes you from every other researcher
- Integrates in key research workflows such as manuscript and grant submission,
- Supports automated linkages between you and your professional activities ensuring that your work is recognized.

ORCID
Connecting Research and Researchers

ORCID Profile Service

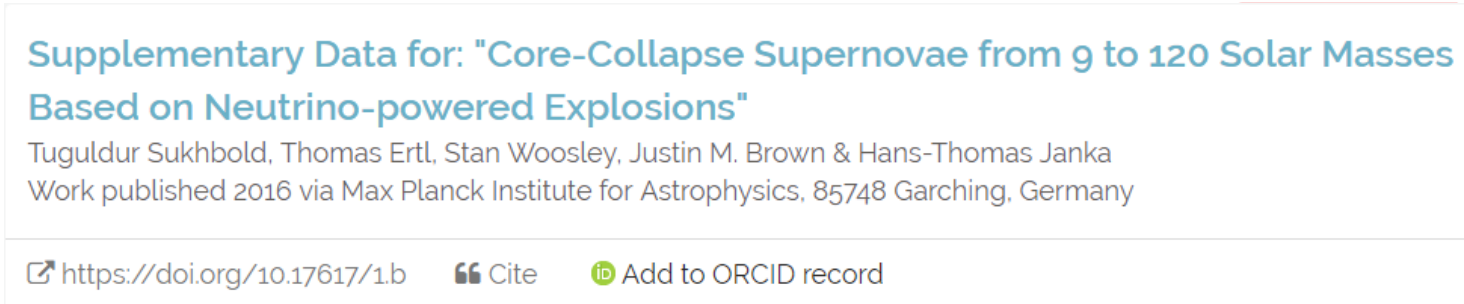
- DataCite – CrossRef – ORCID Collaboration
- Automatic ORCID Profil Update if ORCID is submitted with DOI metadata (ORCID push)



If you authorize Crossref and DataCite to update your ORCID record

Crossref DataCite ORCID

and you add your ORCID to your paper or dataset submission

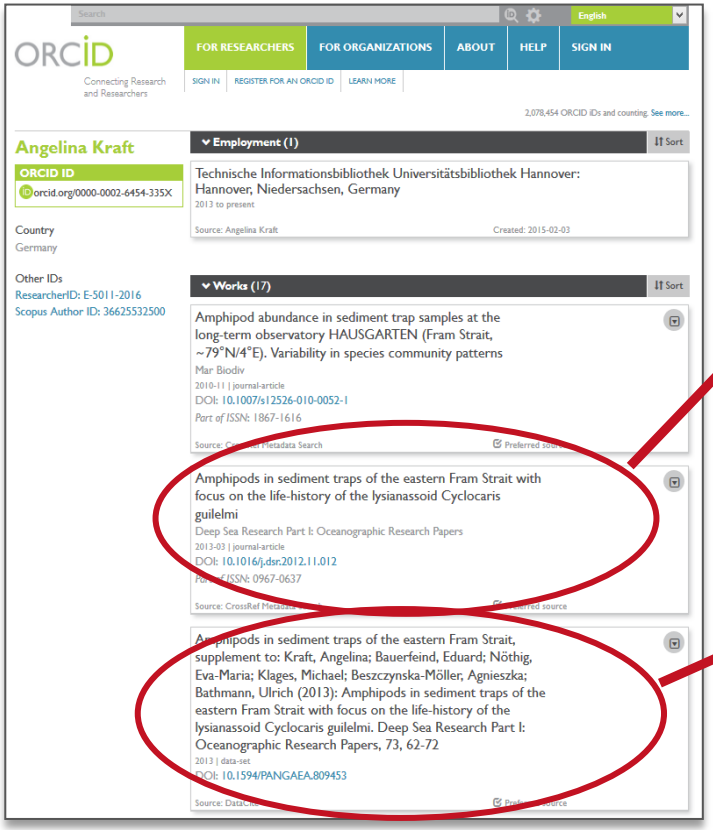


AUTOMATICALLY!

https://profiles.datacite.org/ https://search.datacite.org/

Research Data Publications with ORCID

Digital CV
(e.g. institutional CRIS, ORCID …)

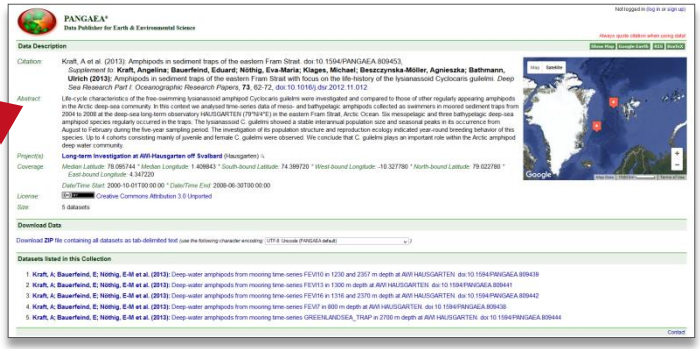


Paper / Journal

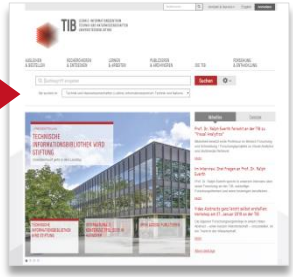
Institut

Data Repository

DOI



Portal



To Dos

1. Register **ORCID iD**
2. Ask your library to provide you with **DOIs** for your research output
3. **Find a repository** with DOI registration services
4. Write a Data Management Plan for your next research project
5. Activate the automatic push of your DOI publications in DataCite Search

DataCite Services

ASSIGN DOIS	https://mds.datacite.org https://api.labs.datacite.org
METADATA SEARCH	https://search.datacite.org/
EVENT DATA	https://dlm.datacite.org https://ls.datacite.org
DATA METRICS	https://makedatacount.org/
PROFILES	https://profiles.datacite.org
RE3DATA	http://re3data.org
CITATION FORMATTER	http://crosscite.org/citeproc/
STATISTICS	http://stats.datacite.org
SERVICE STATUS	http://stats.datacite.org http://twitter.com/datacitetech
OAI-PMH	http://oai.datacite.org
Content Resolver	http://data.datacite.org/
API	https://api.datacite.org/

DataCite
FIND, ACCESS, AND REUSE DATA

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

TIB

THANK YOU!

Further information:

www.tib.eu

www.datacite.org

Contact:
Britta Dreyer
Phone + 49 (0)511 762-17642, britta.dreyer@tib.eu