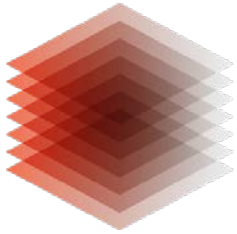
LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

Digital Object Identifier (DOI): Enabling Scientific Publication and Citation – Role of Libraries

Britta Dreyer

Санкт-Петербург, 20. – 21. June 2016
Arbicon 2016

Agenda

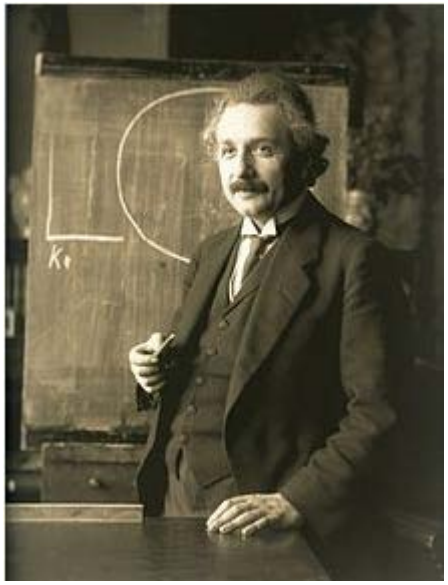
1. The Value of Open Data
2. DOI and DataCite
3. DOI Registration in practice
4. Research Data Services at the German National Library of Science and Technology (TIB)

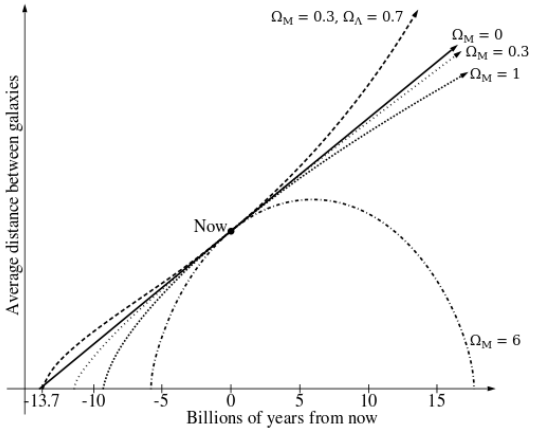
Russian Scientists

Alexander Alexandrowitsch Friedmann (1880 – 1925)

- Born in St. Petersburg, developed the Friedmann-Equations, inspired Albert Einstein

$E = mc^2$



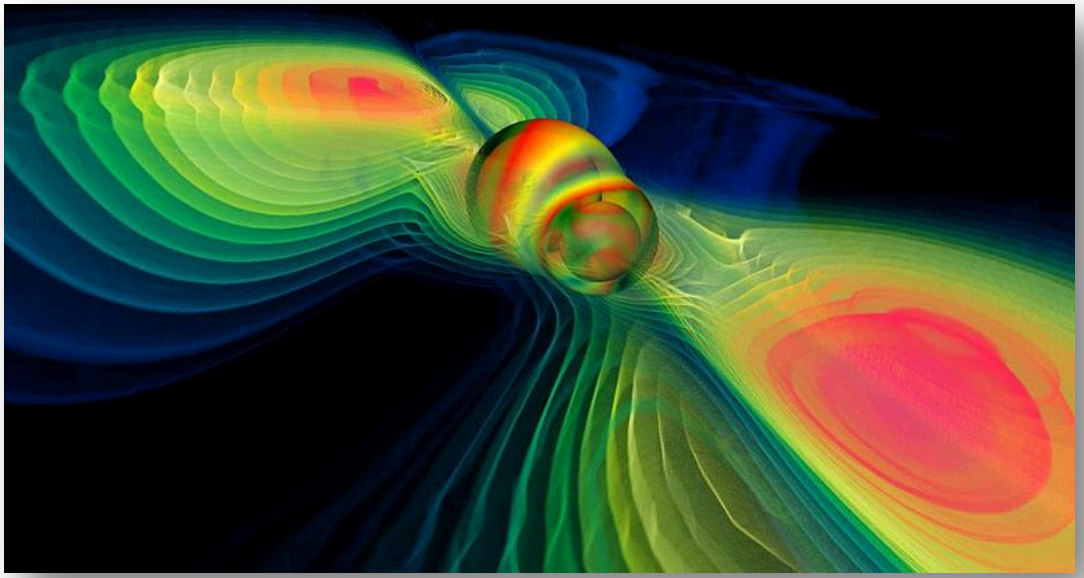


Published Research Data with DataCite DOIs

- On 14 September 2015 first observation of gravitational waves was made - 100 years after publication of the theory of general relativity by Albert Einstein
- Made available by the LIGO and Virgo Collaborations on 11 February 2016 through our member the California Digital Library with DataCite DOIs

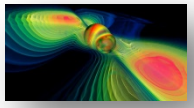
Albert Einstein himself doubted that they could ever be detected...

DataCite
FIND, ACCESS, AND REUSE DATA



Source:
Benger, W: When black holes collide
https://commons.wikimedia.org/wiki/File:When_Black_Holes_Collide.jpg
CC-BY 2.0

State of the Art Data Publication!



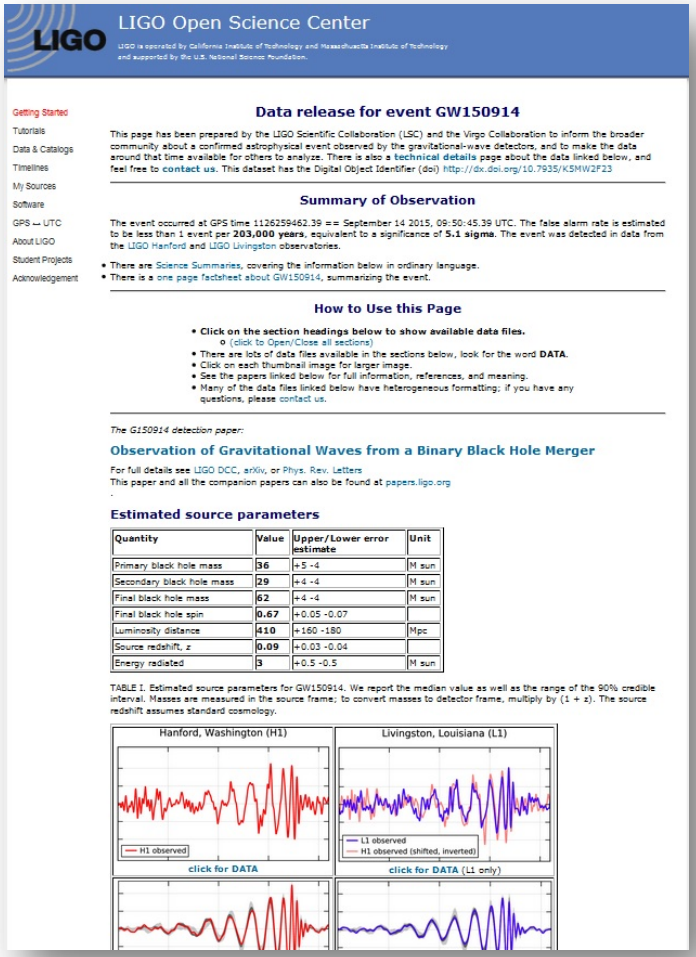
Explore: **GW150914**

View: http://**doi.org/10.7935/K5MW2F23**

The data behind the collision of two black holes

- collected by LIGO's twin detectors,
- citable via a DataCite DOI, and
- **open for everyone!**

Made available by LIGO Open Science Center at the California Institute of Technology and Massachusetts Institute of Technology –
including technical reports, graphs, calibration data & even audio files!

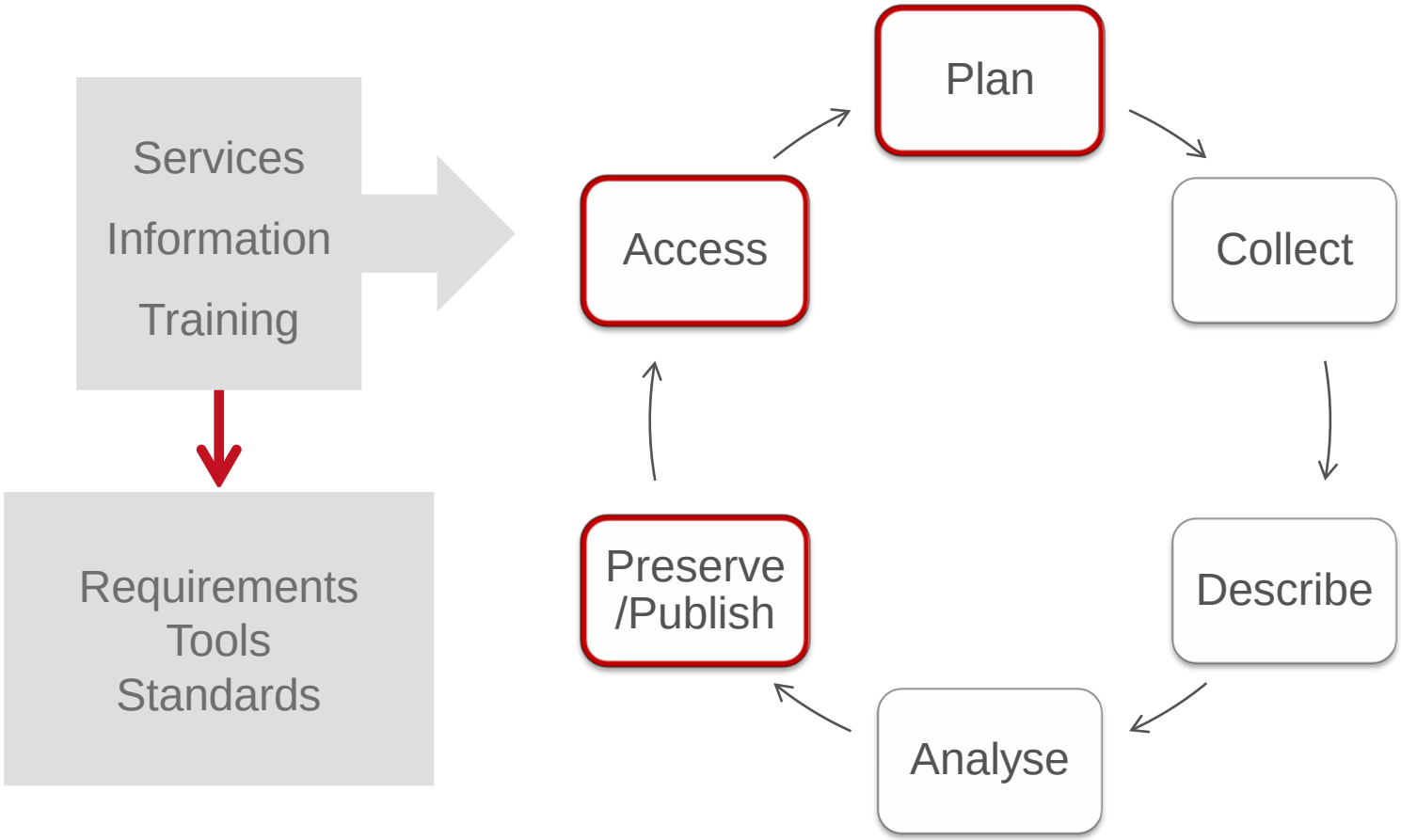


Source: Screenshot GW150914 Landing Page: http://doi.org/10.7935/K5MW2F23

Requirements for Research Data Preservation and Sharing

1. Trustworthy research data repositories
2. Data policies
3. Standards for data citation, metadata, licensing
4. Intellectual property rights and proprietary data
5. Methods and tools adapted to the scientific workflows
6. Cost recovery strategies
7. **Motivation for change**

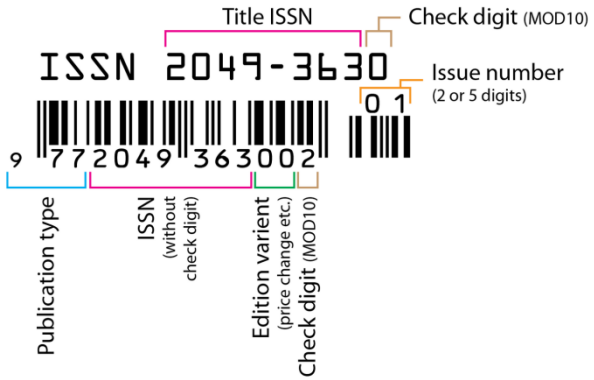
Research Data Management – Key Roles for Libraries

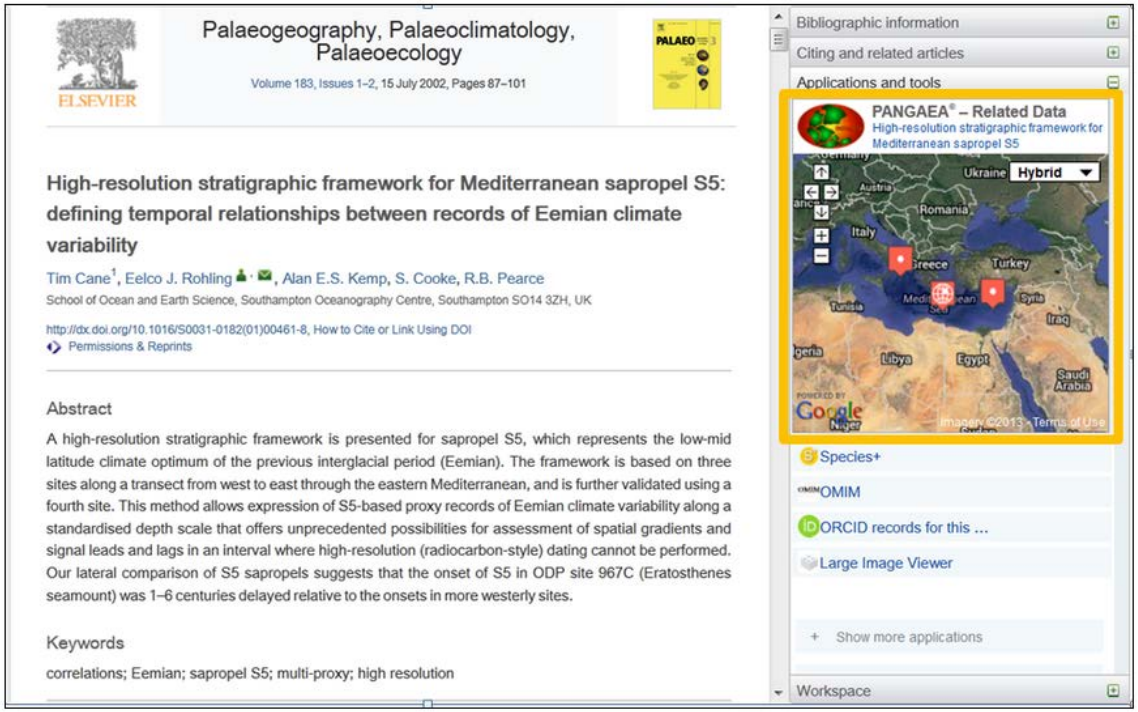


Quelle: http://www.dcc.ac.uk/resources/curation-lifecycle-model

Digital Persistent Identifier

ISBN and ISSN are here since 1970s → **Hyperlinking** not possible→ Granularity options





https://www.elsevier.com/connect/bringing-data-to-life-with-data-linking

For a landing page of this dataset please follow http://dx.doi.org/10.1594/PANGAEA.849289

Quelle: images labeled for reuse http://images.google.de/

Digital Object Identifier (DOI)

- Persistent Identifier and hyperlink
- Typically expressed as a URL
- Specifies the resource – not the location
- Has metadata attached
- ANSI/NISO Z39.84-2000 (reconfirmed 2005, 2010)
- Standard DIN ISO 26324 (2012)
- Unique alphanumeric string assigned by a registration agency

DOI Name				
Directory indicator	.	Registrant code	/	Suffix
Example: 10	.	1234	/	data567

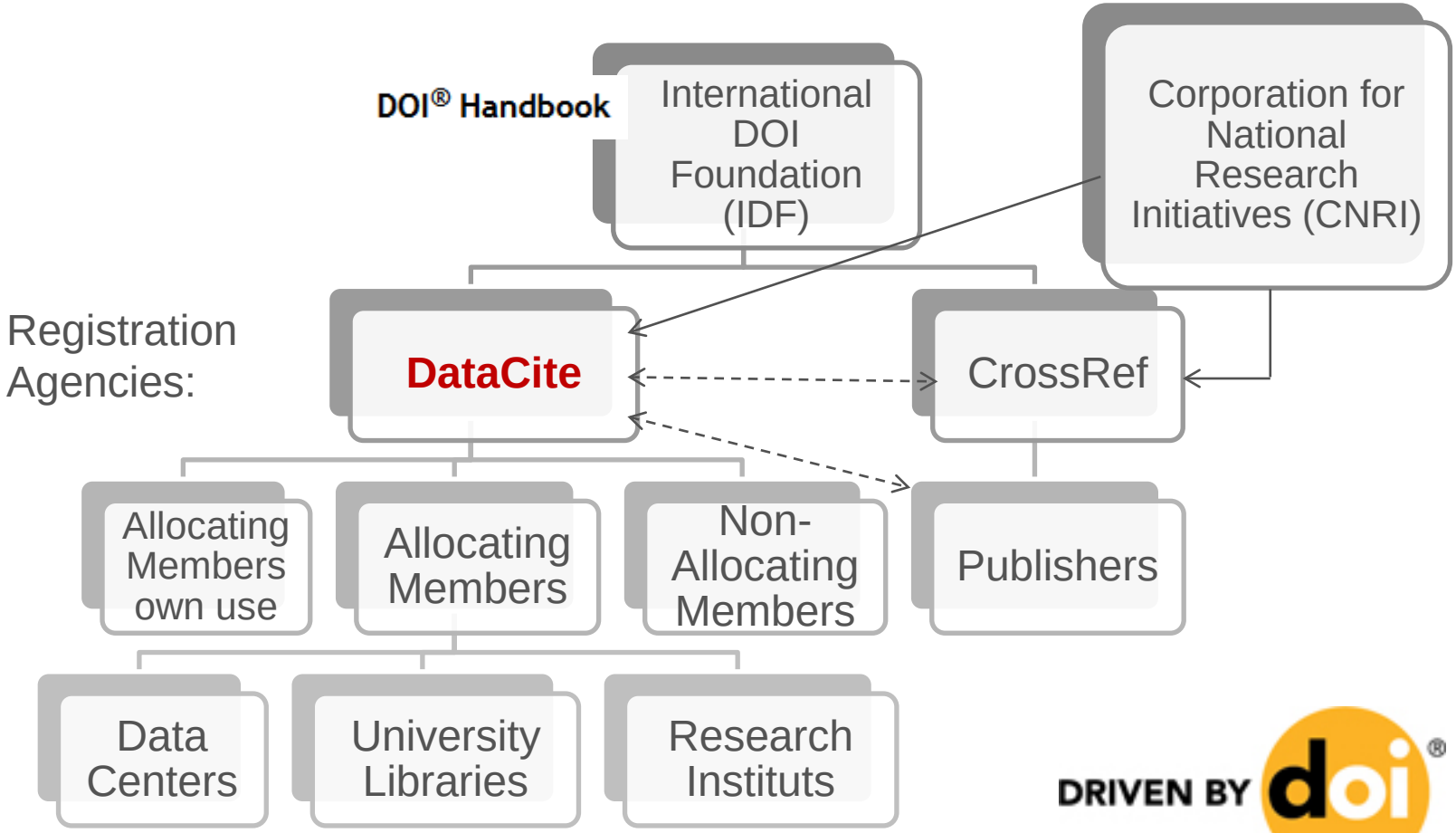
DataCite
FIND, ACCESS, AND REUSE DATA

FORCE11
The Future of Research Communications and e-Scholarship

NISO
How the information world CONNECTS

RDA
RESEARCH DATA ALLIANCE

DataCite Structure



DRIVEN BY doi®

DataCite
FIND, ACCESS, AND REUSE DATA

DOI Registration Agencies

Chinese library for digital resources	Airiti, Inc.	Crossref
Chinese knowledge database	China National Knowledge Infrastructure (CNKI)	DataCite
Universal unique identifier system for movie and television assets	EIDR (Entertainment Identifier Registry)	ISTIC (The Institute of Scientific and Technical Information of China)
DOI names for Japanese journal articles	JaLC (Japan Link Center)	Korea Institute of Science and Technology Information (KISTI)
mEDRA is the European DOI Registration Agency for some European countries.	mEDRA (Multilingual European DOI Registration Agency)	OP (Publications Office of the European Union)

DOI names for Chinese journals, data sets and dissertations

KISTI is a government-funded research institute

Publishing house of the EU institutions, responsible for publishing and distributing EU publications

DataCite in a Nutshell

- Found in December 2009
- 27 allocating members and 8 non-allocating members in 21 countries
- ~ 700 data centers around the world
- DataCite **DOIs** have been assigned to 7.7 Mio DOIs of research datasets
  → making them **public, citable, traceable**.
- German non-profit association - Business Office at German National Library of Science (TIB)
- Member of Research Data Alliance (RDA) and ORCID

DataCite
FIND, ACCESS, AND REUSE DATA

Global DOI Providers for General Resources

DataCite
FIND, ACCESS, AND REUSE DATA

Strong Collaboration

Crossref

Focus:
data community
Mission:
further data sharing

- Linking scholarly output
- Citation formatter
- Content negotiation
- Event Data

Focus:
publisher Community
Mission:
sustainable infrastructure for scholarly communication

The metadata search services do not operate accross registration agency infrastructures.

DataCite Members

DataCite
FIND, ACCESS, AND REUSE DATA

DataCite Membership

Open to all types of organizations

- Allocating Members
- Non-Allocating Members

Members....

- demonstrate an elevated level of **commitment to open-data and open research.**
- become **part of a global data-sharing community**, learning, collaborating, and advocating with a leading-edge network of data research experts.
- **create and manage persistent identifiers (DOIs)** for research outputs (Allocating Members only).
- **play a critical role in advancing DataCite's data-sharing** mission through their financial support.
- **shape the future DataCite** – and data research in general – by serving on our Executive Board.

DataCite – Staff on Board since August 2015

Executive Director	Trisha Cruse
Business Manager	Britta Dreyer
Communication Director	Laura Rueda
Technical Director	Martin Fenner
Starting August 1st: Application Developer	Kristian Garza

DataCite
FIND, ACCESS, AND REUSE DATA

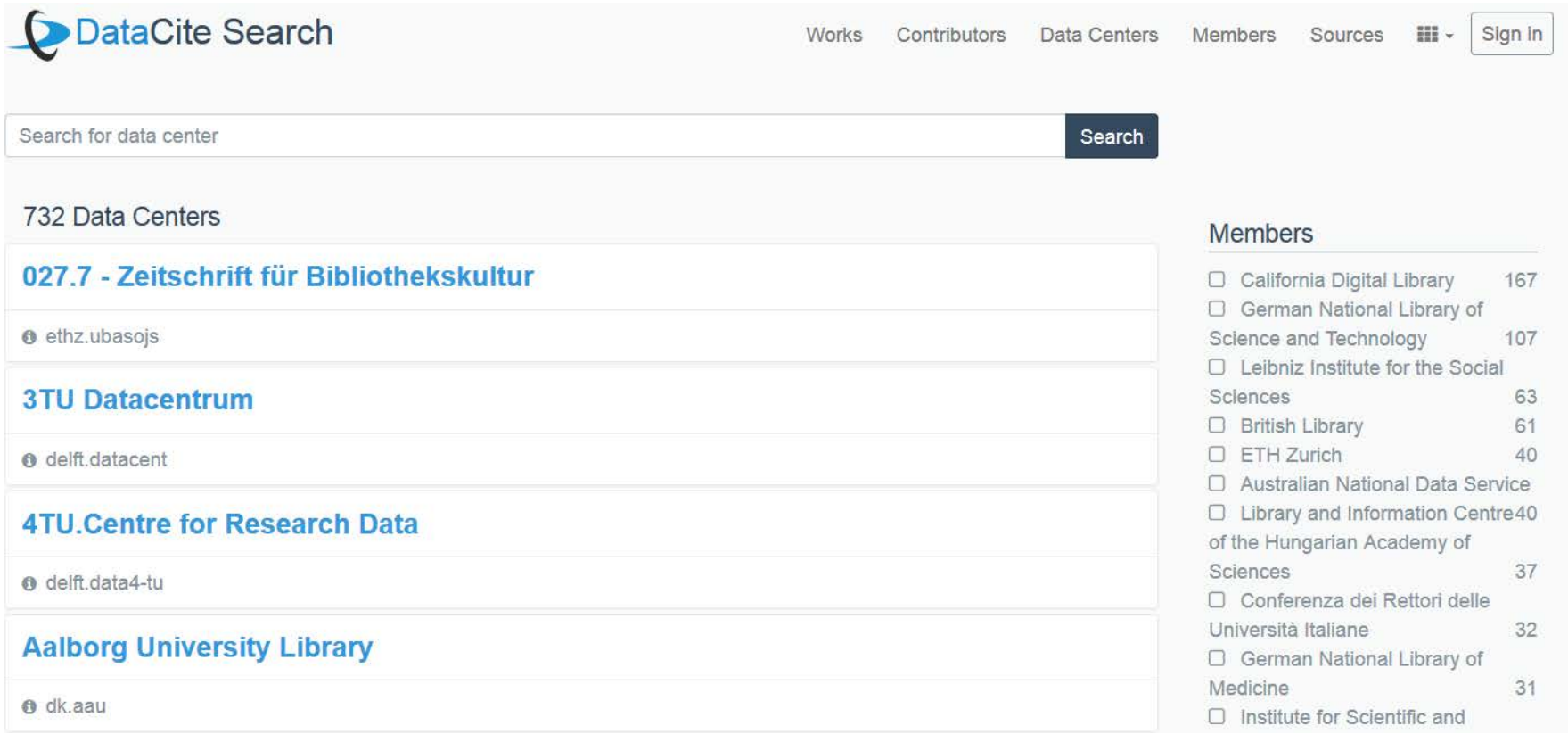
THOR

Services

ASSIGN DOIS	https://mds.datacite.org https://api.labs.datacite.org
METADATA SEARCH	https://search.datacite.org/ https://labs.search.datacite.org
EVENT DATA	https://dlm.datacite.org https://ls.datacite.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://mds.datacite.org/static/apidoc https://api.labs.datacite.org/ (soon with Documentation and new JSON API

DataCite
FIND, ACCESS, AND REUSE DATA

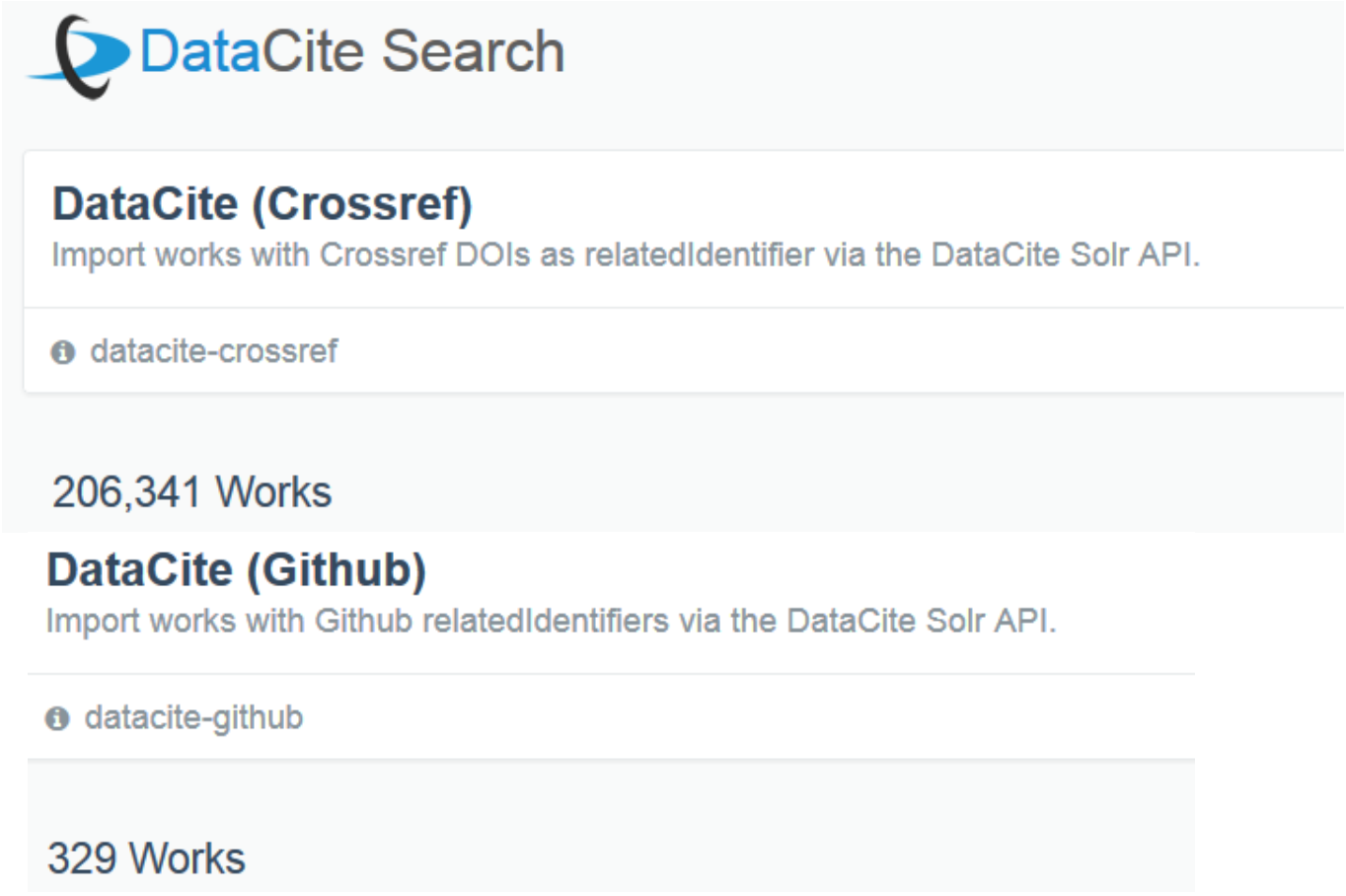
Metadata Search



DataCite
FIND, ACCESS, AND REUSE DATA

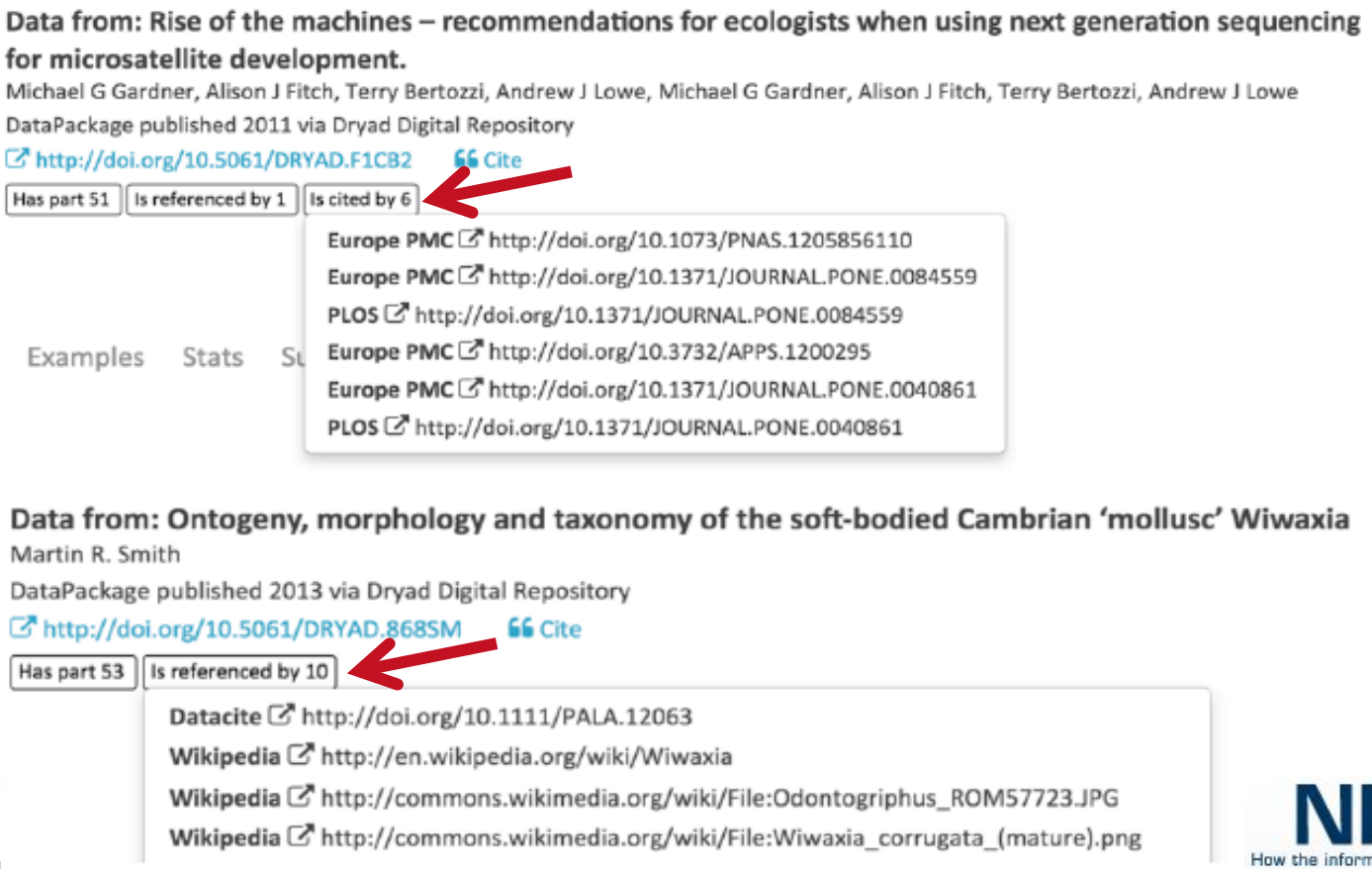
Event Data

- Collect events around scholarly content (DOI Event Tracker)



Data Level Metrics

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



What is an ORCID ID?

- ORCID ID are permanent identifier for researchers
- Protects the researchers unique scholarly identity
- Creation takes 30 seconds – full privacy control
- Lasts longer than your email address – keep contact information up-to-date
- Supports 37 types of „works“
- Push information automatically from Web of Science, Impactstory etc.
- Over 1000 journals are using ORCID to ease their manuscript submission system

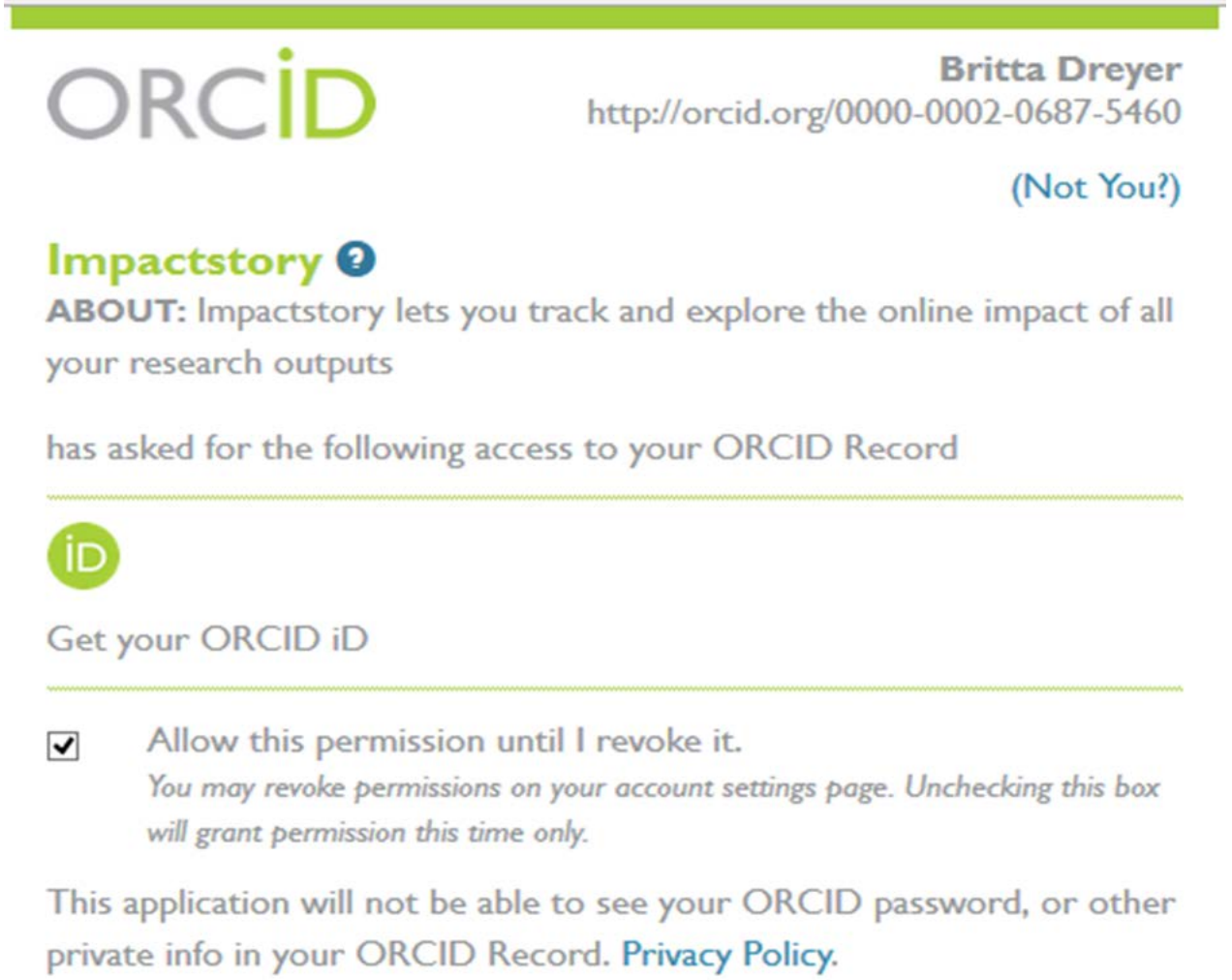
ORCID
Connecting Research and Researchers

Impactstory is a non-profit that helps scientists learn where their research is being cited, shared, saved and more. Discover your impact today.

http://blog.impactstory.org/ten-things-you-need-to-know-about-orcid-right-now/

ORCID Profile

- ✓ Connect funding dollars with tangible outputs
- ✓ Track citations beyond journal articles

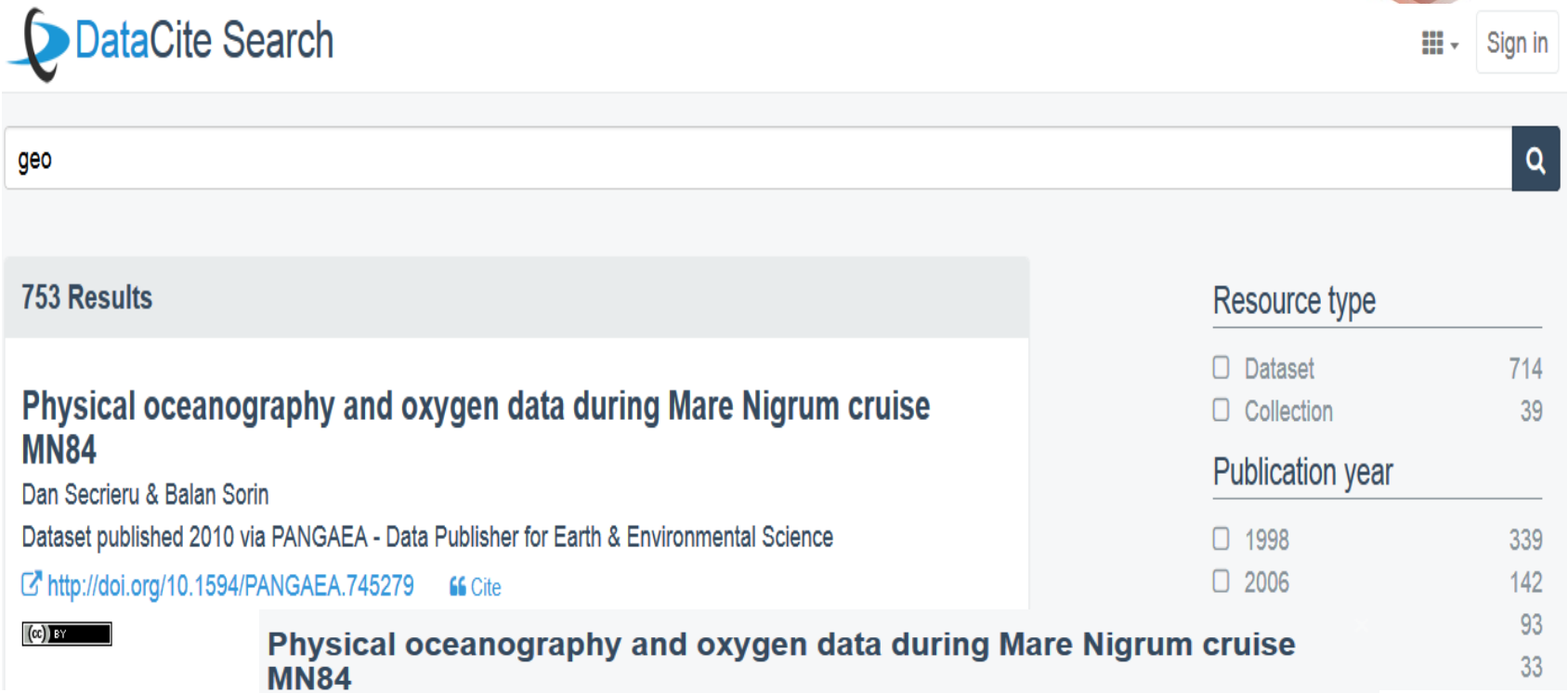


ORCID Profile

- Automatic ORCID profile update when DOI is minted (ORCID push)

- DataCite – CrossRef – ORCID collaboration

Citation Formatter



Citation Recommendation

First Five Metadata schemes:

Creator (PublicationYear): Title. Publisher. Identifier

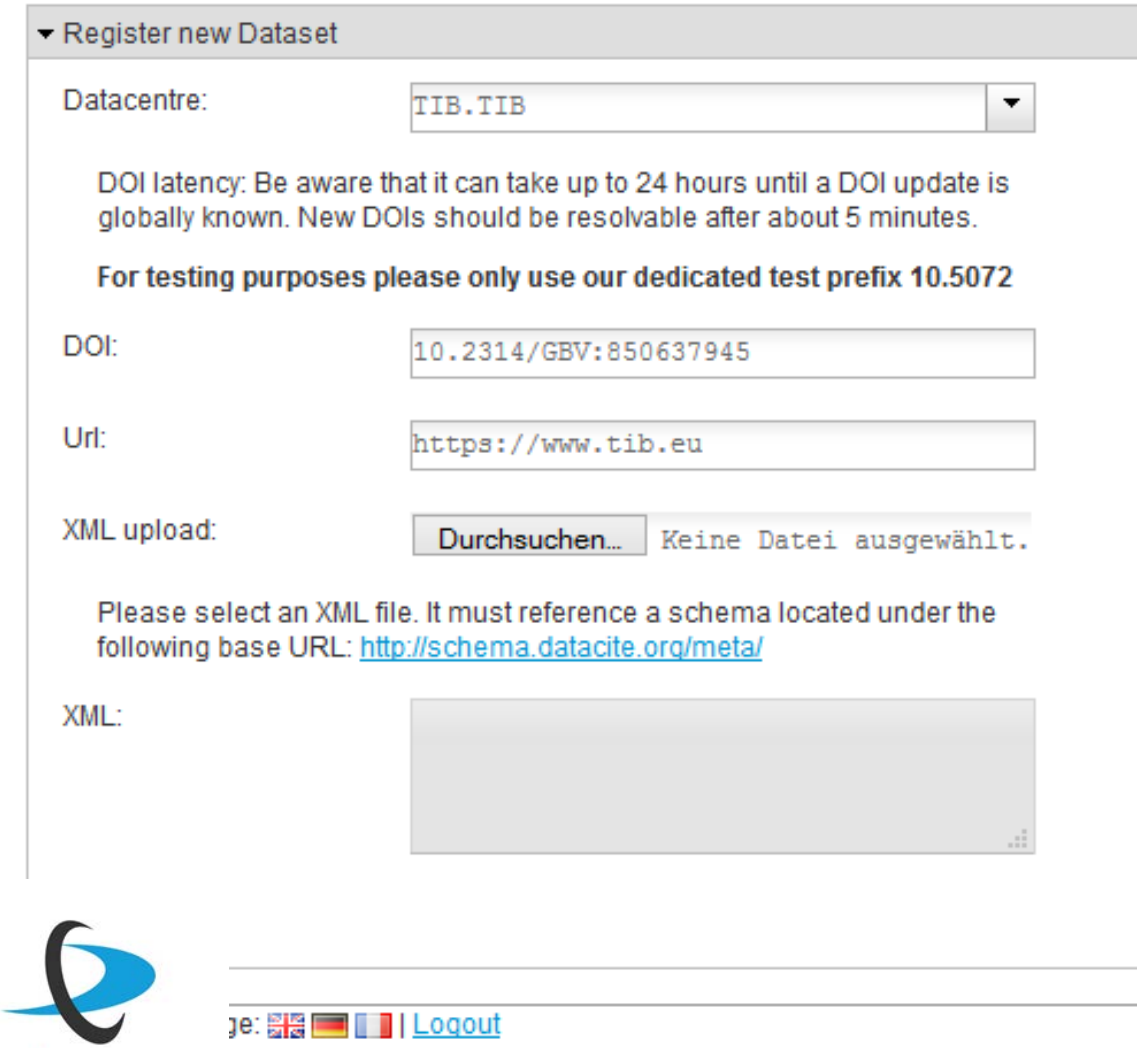
Metadata Version 4 citation recommendation:

Creator (PublicationYear): Title. Version. Publisher. ResourceType. Identifier

Example:
Denhard, Michael (2009): dphase_mpeps: MicroPEPS LAF-Ensemble run by DWD for the MAP D-PHASE project. World Data Center for Climate. Dataset. http://doi.org/10.1594/WDCC/dphase_mpeps

DataCite
FIND, ACCESS, AND REUSE DATA

DOI Registration in Practice



DataCite Metadata Store

Registration via DataCite User Interface or REST API

https://www.tib.eu/de/suchen/id/TIBKAT%3A850637945/

Creation process of a prefix?

Handle Registry

Namespace reserved for **Prefix** 10.1234

CNRI

Prefix 10.1234

DataCite Server

Create new Prefix 10.1234

DataCite **Metadata Store**

Assign new Prefix 10.1234

Allocator Account

Assign new Prefix 10.1234

Data Center Account

DataCite
FIND, ACCESS, AND REUSE DATA

Prefixes

Please keep in mind that DOIs within one prefix name space are inseparable

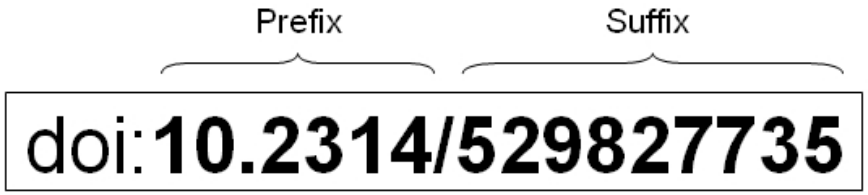
Recommendations:

Use additional prefix if:

- Objects are on separate plattforms/servers (apply internal/logical separation)
- Different and separate customers
- Different systems or products

Suffix

- Suffix defined by registrant
- Recommendation: Opaque string (names can change)
- Characters for a DOI name:



A – Z	: (colon)
a – z	. (full-stop)
0 – 9	- (hyphen)
/ (slash)	_ (underscore)

Media Fragment Identifier

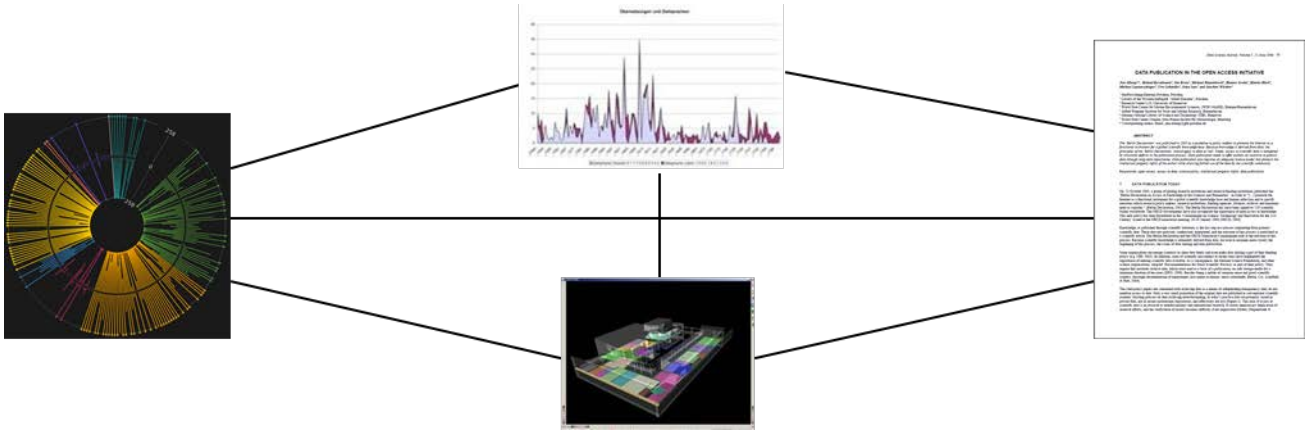
AV Media: http://dx.doi.org/10.5446/10133#t=00:45,00:48

Book: http://dx.doi.org/10.12345/9876#5

http://www.tib-hannover.de/fileadmin/informationsmaterial/flyer/TIB_DOI_Service_DOKU_ENG.pdf

DataCite Metadata

- IDF Requirement
- Based on Dublin Core
- Applicable accross research disciplines
- Linkage between different types of resources
- Connecting resources, creators, institutions and funders

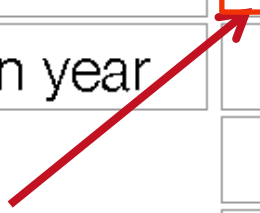


DataCite
FIND, ACCESS, AND REUSE DATA

DataCite Metadata Properties

Current version Metadata Schema 3.1 https://schema.datacite.org/

Mandatory	Recommended	Optional
Identifier	Subject	Language
Creator	Contributor	Alternate ID
Title	Date	Size
Publisher	Resource Type	Format
Publication year	Related identifier	Version
	Description	Rights
	GeoLocation	



Will be mandatory with Metadata version 4.0

Key for search queries and harvesting

Metadata Schema V 4.0

- Roll-out in September 2016
- Allowing more than one nameIdentifier per creator or contributor
- Addition of new optional subproperties for creatorName and contributorName
  - familyName
  - givenName
- Addition of a new property: FundingReference, with subproperties
  - funderName
  - funderIdentifier
    - funderIdentifierType
  - awardNumber
  - awardURI
  - awardTitle

https://projects.tib.eu/confluence/display/datacite/Documentation+repo

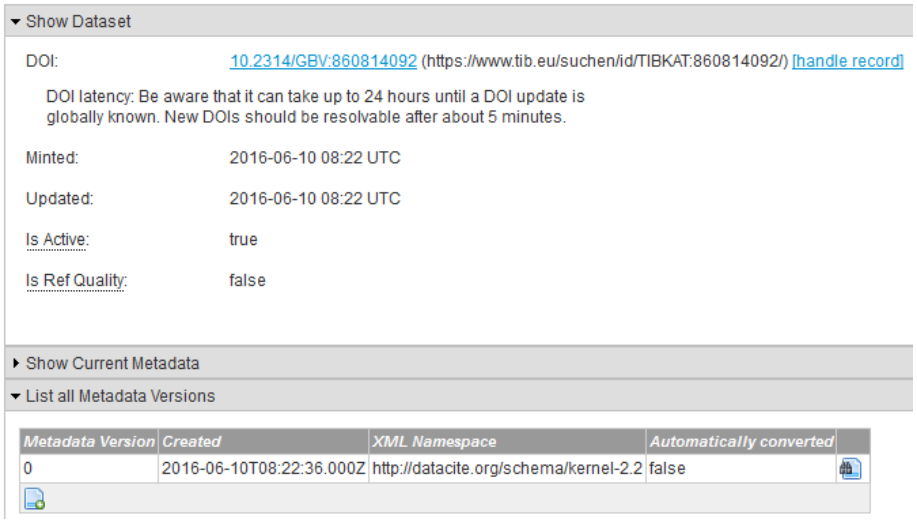
DOI - New Versions, Changes, Removal

Case: **New version of registered resource**
Solution: New DOI for each version – use metadata attribute „relatedIdentifier“ to link them conveniently

Case: **Change to metadata** i.e. typo fix, transliteration of a name
Solution: Upload the updated xml file via UI or API

Case: **Resource is removed**
Solution: DOI, metadata and landing page remain – worst case DOI inactive

Tombstone page:



Open Access and Licensing

Open Access is the main goal!

- But it is possible to restrict access to data if necessary → Landing page should inform of access rights.

- Metadata and landing page should always be available.

Russian versions of CC Licenses:
Wikimedia Wikisource.
Professional 2011 3.0 translations by IIS:
Russian translation of CC BY 3.0 (Unported)
Russian translation of CC BY-SA 3.0 (Unported)
Russian translation of CC BY-ND 3.0 (Unported)
Russian translation of CC BY-NC-SA 3.0 (Unported)
Russian translation of CC BY-NC 3.0 (Unported)
Russian translation of CC BY-NC-ND 3.0 (Unported)

German National Library of Science and Technology (TIB)

Research library for science and technology, architecture, chemistry, computer science, mathematics and physics

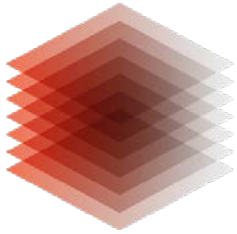
Member of **Leibniz Association**
500 members of staff

Global supplier for scientific and technical information

Founding member of DataCite

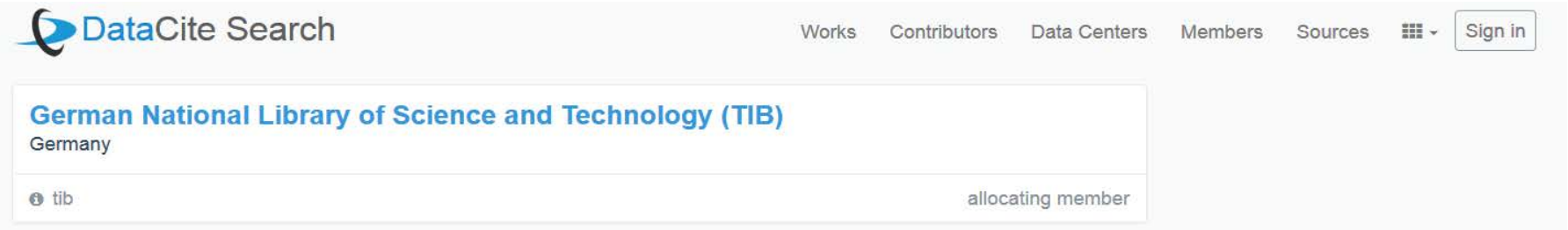
- 55,345 journal subscriptions (15,967 print; 39,378 digital)
- 9.1 m items, 17.3 m patents & standards

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

TIB Resource Types



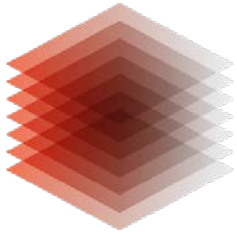
Total Metadata Records: 1,030,112 Works

Resource Types

Dataset	361,983
Text	247,634
Image	124,542
Collection	26,122
Audiovisual	628
Other	174
Film	54
Software	22
Event	21
Sound	2
Interactive resource	1

https://search.labs.datacite.org/

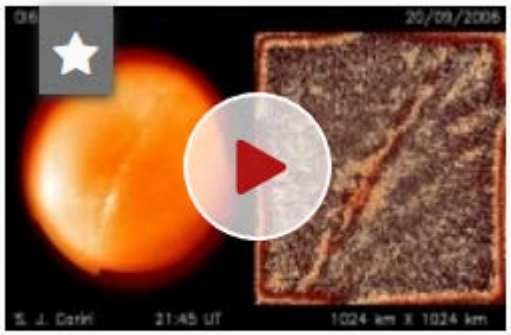
LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

Internal Customers

- Retro-digitisation
- Audio visual media
- Research reports
- Handbuch.io
- University document server



Periodic gravity waves in the lower thermosphere

TIB AV-PORTAL

TIB Portal: Search and Access to Digital Content

Page 1 from 21,695 hits

Sort by: Best results | Currentness | Title

Drilldown

Year of publication

Ok

Type of media

‹ All types of media

- [ ] Article (Journal) (3,362,203)
- [ ] Report (789,332)
- [ ] Conference paper (665,529)
- [ ] Book (58,958)
- [x] Research Data (21,695)
- [ ] Article/Chapter (Book) (17,495)
- [ ] Conference Proceedings (9,428)
- [ ] Theses (9,058)
- [ ] No value (5,094)
- [ ] Paper (4,185)

« ‹ 1 2 3 4 5 6 7 › »

Non-tidal Ocean Geopotential Coefficients averaged over certain Time Period (3.9393604)

GRACE Science **Data** System | DataCite | 2002
Supplier: GRACE Science **Data** System
Keywords: ECMWF atmospheric **data**

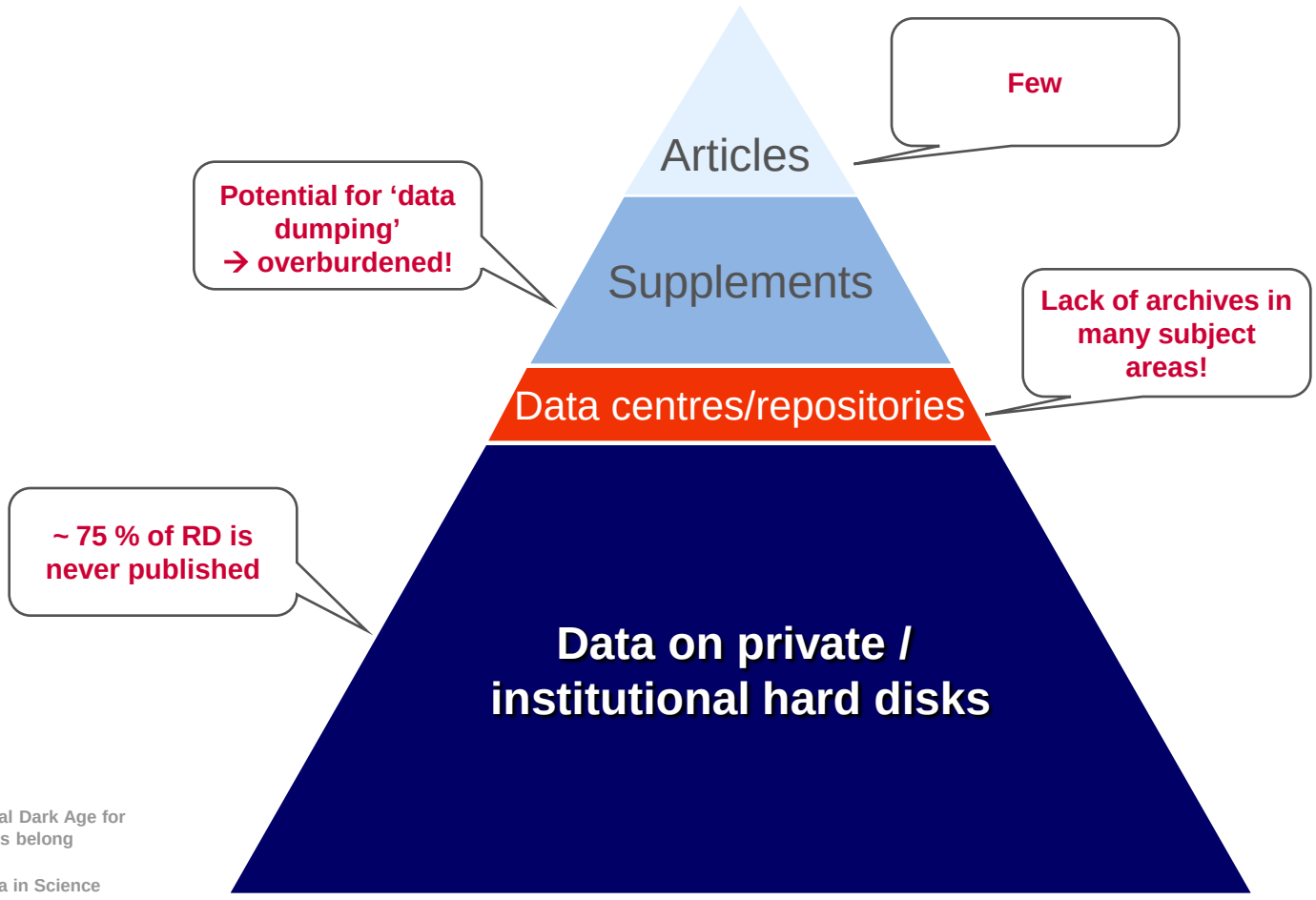
Static Field Geopotential Coefficients estimated from Satellite Data only (3.9393604)

GRACE Science **Data** System | DataCite | 2002
Static Field Geopotential Coefficients estimated from Satellite **Data** only ...
Supplier: GRACE Science **Data** System

Non-tidal Atmosphere Geopotential Coefficients averaged over certain Time Period (3.9393604)

GRACE Science **Data** System | DataCite | 2002
Supplier: GRACE Science **Data** System
Keywords: ECMWF atmospheric **data**

1. Reality of data publishing



Modified based on
STM / Smit, E: Avoiding a Digital Dark Age for Data: why data and publications belong together
ICSTI workshop Delivering Data in Science
PARIS, 5 March 2012

RADAR: Research Data Repository

- **Digital data repository** (RADAR) for archiving & publishing research data

Goal:

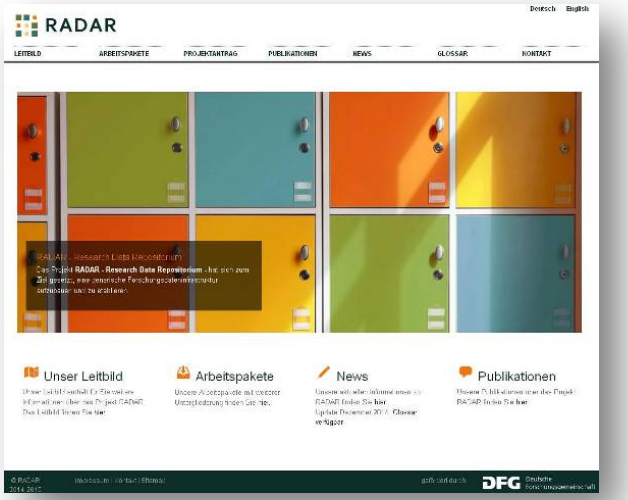
Preservation & reuse of research data

Focus:

Repository for heterogeneous research data (‚**Long Tail**‘), addition to big data archives → for customers without own computing centers/capacities, dedicated to open research data standards

Further information: www.radar-projekt.org

Service (July 2016): www.radar-service.eu



RADAR: Service & Business Model

Services:

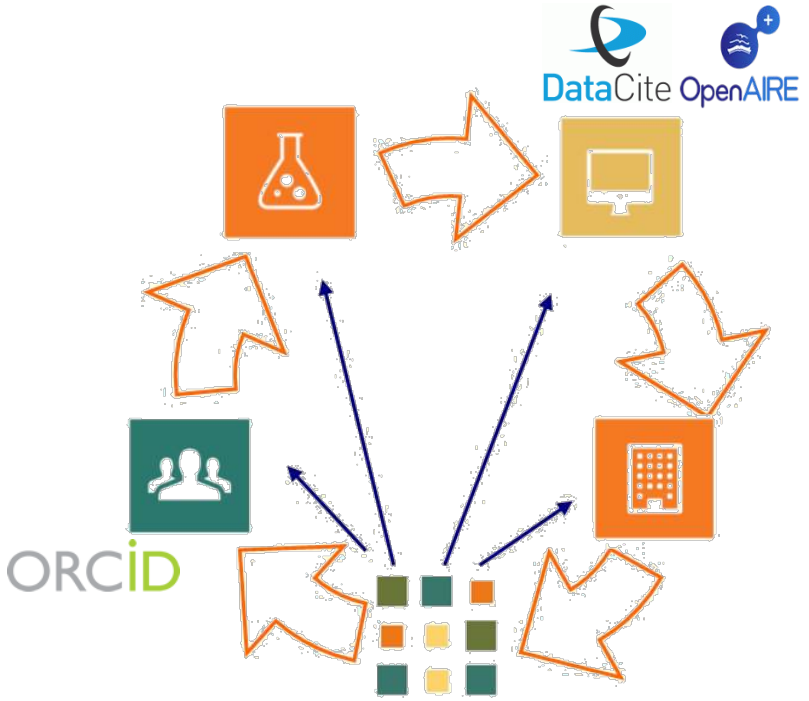
Basic service: **Archival Storage** (with EPIC-Handle)
Extended service: **Data Publication** (with **DOI**)

Features:

- Data Life Cycle support
- Sign Up using ORCID
- REST **API** for clients (**customizable**)
- Interoperability & cross-linking of published datasets via API: DataCite, ORCID & others
- **Optional Peer-Review** Support
- **Download Statistics**

Prices:

Annual fee + data volume
(price information: www.radar-projekt.org/display/RE/Home)



Generic end-point repository with services for scientists/institutions

Resources

Citations recommendations
http://jats4r.org/citations-recommendations/

How to develop RDM services
http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services

DOI Handbook
https://www.doi.org/hb.html

Workshops for researcher
http://anulib.anu.edu.au/research-learn/research-data-management-0
http://www.data-archive.ac.uk/media/2894/managingsharing.pdf
http://datalib.edina.ac.uk/mantra/
http://opus.bath.ac.uk/412/

DataCite
FIND, ACCESS, AND REUSE DATA

DataCite Plugins

Dataverse software will be able to mint DOIs directly through the DataCite Metadata Store API.
Https://github.com/IQSS/dataverse/releases/tag/v4.3

OJS DataCite Plugin
https://github.com/pkp/ojs/tree/master/plugins/importexport/datacite

Dspace
https://wiki.duraspace.org/display/DSDOC/All+Documentation

Eprints
https://github.com/eprints/datacite

DataCite
FIND, ACCESS, AND REUSE DATA

DataCite - Upcoming webinar and event

Webinar:

DataCite Metadata Schema Version 4.0 in June/July

Event:

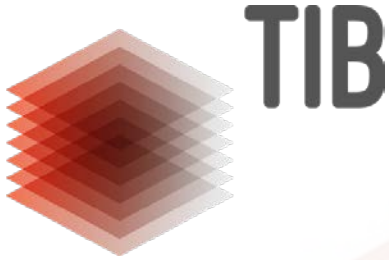
PIDapalooza Reykjavik, Nov 2016

November 9 -1, 2016 in Reykjavik, Iceland

http://pidapalooza.org/

DataCite
FIND, ACCESS, AND REUSE DATA

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

Спасибо and thank you!

Contact:
Britta Dreyer
T +49 (0)511 762-17642,
TIB DOI-Service: britta.dreyer@tib.eu
DataCite Business Office: britta.dreyer@datacite.org