DataCite – International consortium for data citation

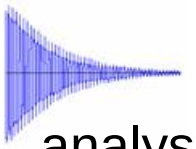
Jan Brase, DataCite - TIB

April 25th
Bielefeld conference

TIB | German National Library of Science and Technology

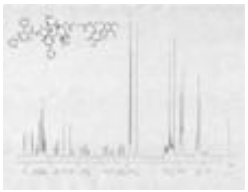
DataCite

Problem with data: The research trajectory

Data  … is lost!

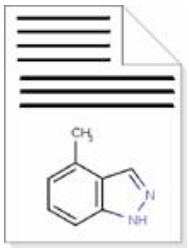
are

analysed
synthesised
interpreted

become **Information**  … is traceable

is

published

becomes **Knowledge** 

… is accessible

Publication

TIB | German National Library of Science and Technology

DataCite

What if data would be citable?

High visability of the data

Easy re-use and verification of the data sets.

Scientific reputation for the collection and documentation of data (Citation Index)

Encouraging the *Brussels declaration on STM publishing*

Avoiding duplications

Motivation for new research

How to achive this?

Science is global

- it needs global standards
- Global workflows
- Cooperation of global players

Science is carried out locally

- By local scientist
- Beeing part of local infrastrucures
- Having local funders

DataCite

Global consortium carried by local institutions

focused on improving the scholarly infrastructure around datasets and other non-textual information

focused on working with data centres and organisations that hold data

Providing standards, workflows and best-practice

Initially, but not exclusivly based on the DOI system

Founded December 1st 2009 in London

History

03

- DFG funded project with German WDCs

05

- TIB begins to issue DOI names for datasets

03. 09

- Paris Memo-randum

12. 09

- DataCite Asso-ciation founded in London
- 7 members

12. 10

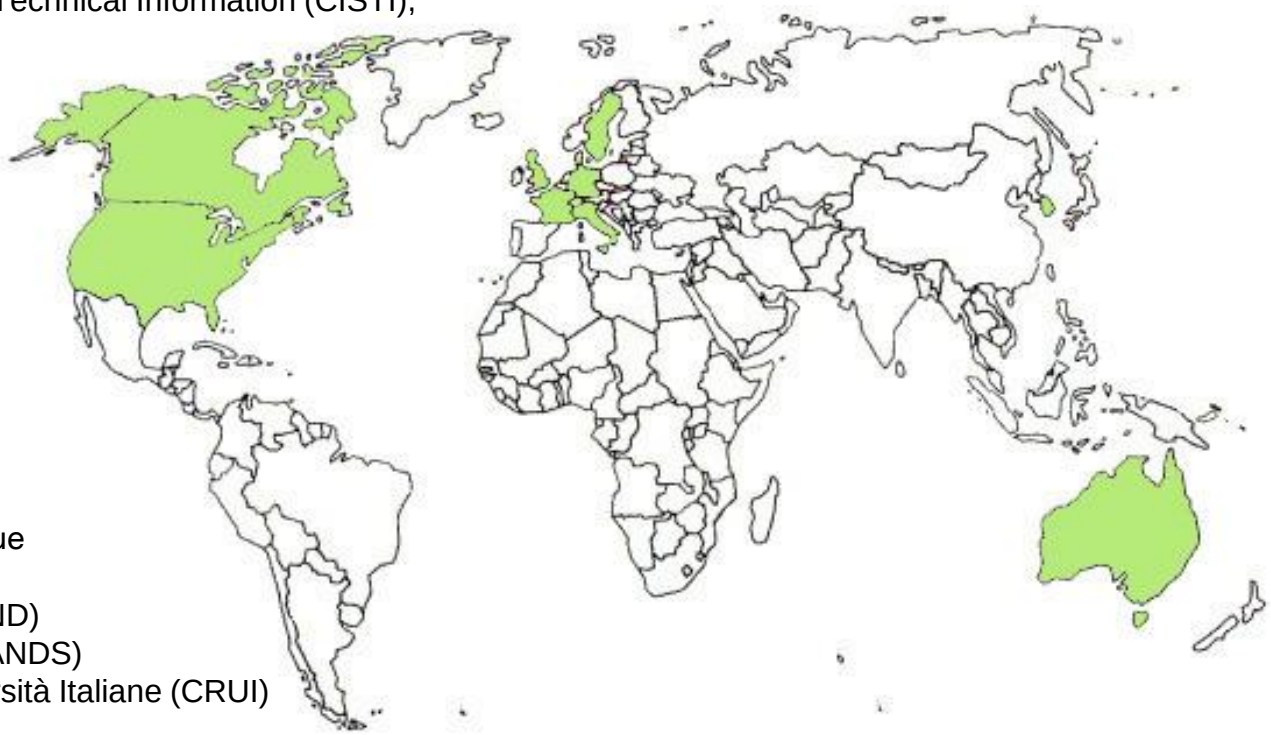
- 12 members
- All members assigned DOIs
- Over 800,000 items registered
- Pilot projects with Data Centres

12. 11

- 16 members
- Over 1,2 million DOI names
- Metadata store

DataCite members

Technische Informationsbibliothek (TIB)
Canada Institute for Scientific and Technical Information (CISTI),
California Digital Library, USA
Purdue University, USA
Office of Scientific and Technical
Information (OSTI), USA
Library of TU Delft,
The Netherlands
Technical Information
Center of Denmark
The British Library
ZB Med, Germany
ZBW, Germany
Gesis, Germany
Library of ETH Zürich
L'Institut de l'Information Scientifique
et Technique (INIST), France
Swedish National Data Service (SND)
Australian National Data Service (ANDS)
Conferenza dei Rettori delle Università Italiane (CRUI)

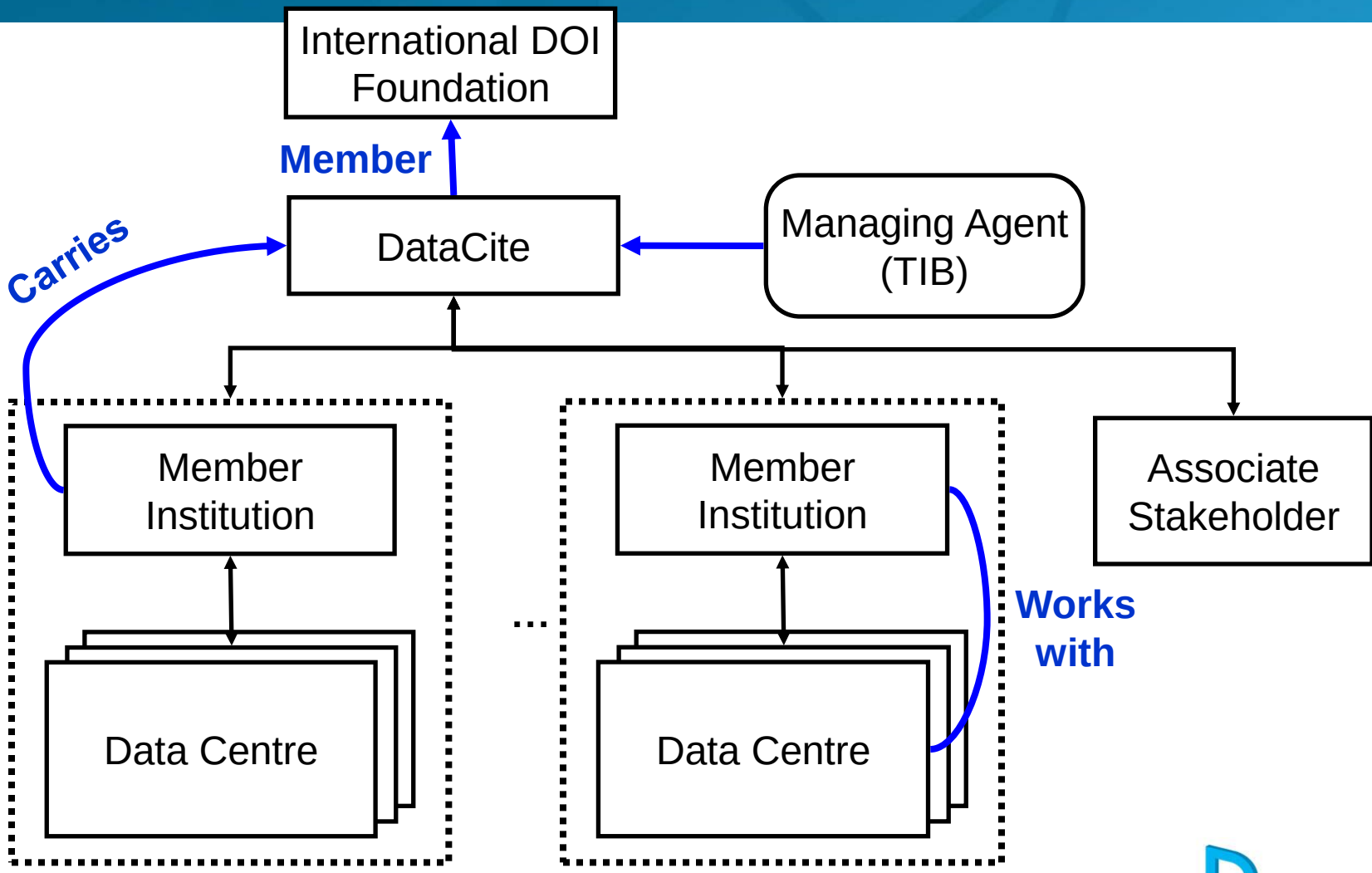


Affiliated members:
Digital Curation Center (UK)
Microsoft Research
Interuniversity Consortium for Political and Social Research (ICPSR)
Korea Institute of Science and Technology Information (KISTI)

TIB | German National Library of Science and Technology

DataCite

DataCite structure



DataCite‘s main goals

Act as DOI registration agency

Actively involved in developing standards and workflows

CODATA-TG, STM, ICSTI, Data citation index

Central portal allowing access to the metadata from all registered objects. (OAI)

Community for exchange of all relevant stakeholders in the area access to and linking of data (data centers, publishers, libraries, research organisation, science unions, funders)

DataCite in 2012

Over 1,300,000 DOI names registered so far

DataCite Metadata schema published (in cooperation with all members) http://schema.datacite.org

DataCite MetadataStore

http://search.datacite.org

OAI Harvester

http://oai.datacite.org

DataCite search

Searchterm: *

Searchterm: uploaded:[NOW-7DAY TO NOW]

Searchterm: relatedIdentifier:*

Searchterm: relatedIdentifier:issupplementto\:10.1029*

Searchterm:relatedIdentifier:*\:10.1055*

DataCite Content Service

Service for displaying DataCite metadata

Different formats (BibTeX, RIS, RDF, etc.)

Content Negotation (through MIME-Typ)

- Access through DOI proxy (http://dx.doi.org)
- First implemented by CNRI and CrossRef:

Alpha available:

http://data.datacite.org

Examples

curl -L -H "Accept: application/x-datacite+text"
"http://dx.doi.org/10.5524/100005"

⇒ *Li, j; Zhang, G; Lambert, D; Wang, J (2011): Genomic data from Emperor penguin. GigaScience. http://dx.doi.org/10.5524/100005*

curl -L -H "Accept: application/rdf+xml"
http://dx.doi.org/10.5524/100005

⇒ *RDF-file*

curl -L -H "Accept: application/raw" http://dx.doi.org/10.5524/100005

=> *?*

Citation

The dataset:

Storz, D et al. (2009):

Planktic foraminiferal flux and faunal composition of sediment trap L1_K276 in the northeastern Atlantic.

http://dx.doi.org/10.1594/PANGAEA.724325

Is supplement to the article:

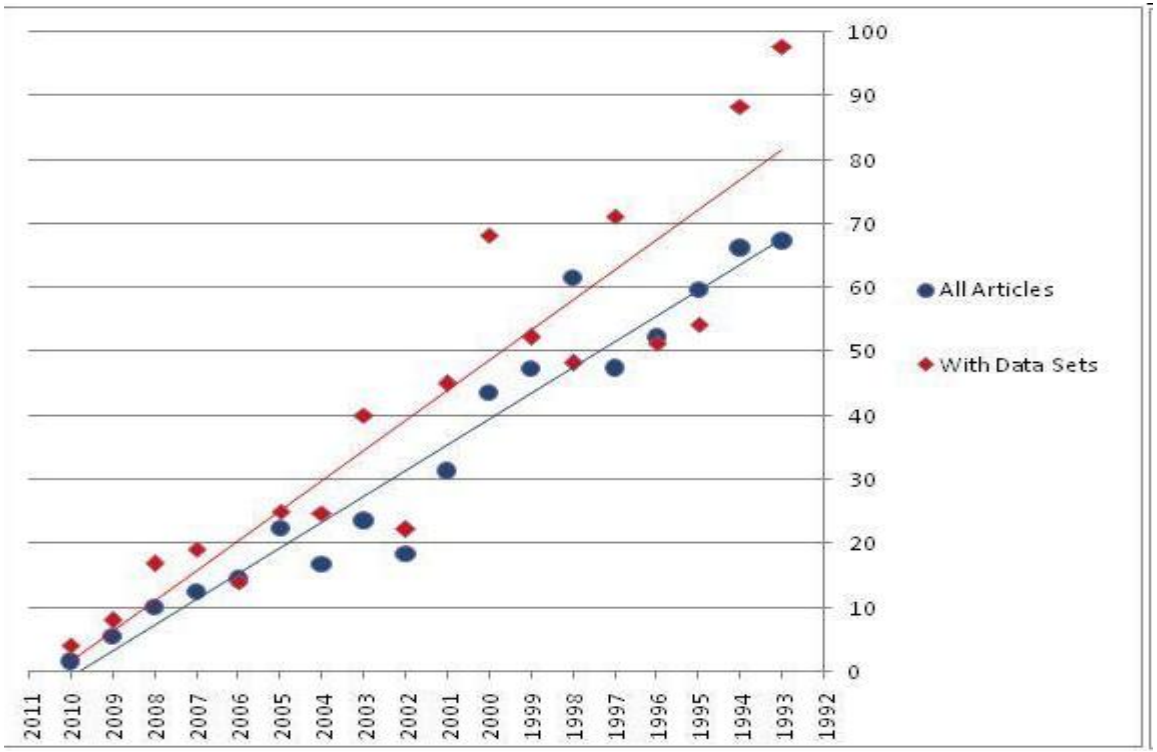
Storz, David; Schulz, Hartmut; Waniek, Joanna J; Schulz-Bull, Detlef; Kucera, Michal (2009): *Seasonal and interannual variability of the planktic foraminiferal flux in the vicinity of the Azores Current.*

Deep-Sea Research Part I-Oceanographic Research Papers, **56(1),** 107-124,

http://dx.doi.org/10.1016/j.dsr.2008.08.009

Availabilty of data increases citation by 35%

Jon Sears (AGU) - Abstract for the AGU 2011



Result

EHEC outbreak 2011:

Li, D; Xi, F; Zhao, M; Chen, W; Cao, S; Xu, R; Wang, G; Wang, J; Zhang, Z; Li, Y; Cui, C; Chang, C; Cui, C; Luo, Y; Qin, J; Li, S ; Li, J; Peng, Y; Pu, F; Sun, Y; Chen, Y; Zong, Y; Ma, X; Yang, X; Cen, Z; Song, Y; Zhao, X; Chen, F; Yin, X; Rohde, H; Liang, Y; Li, Y and the Escherichia coli O104:H4 TY-2482 isolate genome sequencing consortium (2011):

Genomic data from Escherichia coli O104:H4 isolate TY-2482.
BGI Shenzhen. doi:10.5524/100001

http://dx.doi.org/10.5524/100001

Meet us and discuss with us

- DataCite summer meeting, June 14th, Copenhagen (in conjunction with Nordbib conference „Structural frameworks for open, digital research”, June 11.-13.)

- http://www.datacite.org
- contact@datacite.org

TIB | German National Library of Science and Technology

DataCite