

# DOI uses cases for data

Jan Brase  
DOI outreach meeting

November 21<sup>st</sup>  
Milano

# What if any kind of scientific content would be citable?

High visibility of the content

Easy re-use and verification.

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

Encouraging the *Brussels declaration on STM publishing*

Avoiding duplications

Motivation for new research

# Prologue

In 1998 a paper was published from GFZ Potsdam analysing the idea of using DOI names for data citation  
At the 17th International CODATA Conference (2000) in Italy the German CODATA group discussed this further



Leading to a German working group 2000 – 2002 founded by the German Research foundation (DFG)

# Project

2003 – 2007 The DFG founded a project „Citability of primary data“

Members:

- WDC MARE Bremen (PANGAEA)
- WDC CLIMATE Hamburg
- GFZ Potsdam
- TIB Hannover

# 2004: First DOI names registered for data

First DOIs from project partners:

18.03.2004: doi:10.1594/WDCC/EH4\_OPYC\_SRES\_A2 (DOI #1)

22.07.2004: doi:10.1594/GFZ/ICDP/KTB/KTB-GEOCH-GASCHR-P

14.12.2004: doi:10.1594/PANGAEA.119754

End of 2004: about 30 DOIs registered and in library catalogue TIBORDER

**DKRZ DEUTSCHES KLIMARECHENZENTRUM** World Data Center for Climate Hamburg

**DOI #1**

**TIBORDER - Dokumentlieferdienst der TIB Hannover - results/titledata - Microsoft Internet Explorer zur Verfügung gestellt von**

**DOI #1**

**DOI for Scientific and Technical Data**  
10.1594/WDCC/EH4\_OPYC\_SRES\_A2

**Citation elements**

**Creator**  
(person(s) or institute(s) responsible for this assemblage of data: e.g. author, data collector, editor...)  
Stendel, Martin; Schmith, Torben; Roeckner, Erich; Cubasch, Ulrich

**Publication Year**  
2004

**Title**  
ECHAM4\_OPYC\_SRES\_A2: 110 YEARS COUPLED A2 RUN 6H VALUES

**DOI Publisher**  
WDCC at DKRZ

**Identifier**  
DOI:10.1594/WDCC/EH4\_OPYC\_SRES\_A2

The DataCite consortium proposes a citation format (Creator (PublicationYear): Title. DOI Publisher. Identifier) citation export in different formats.

**Detailed Metadata**  
[http://cera-www.dkrz.de/WDCC/ui/Entry.jsp?acronym=EH4\\_OPYC\\_SRES\\_A2](http://cera-www.dkrz.de/WDCC/ui/Entry.jsp?acronym=EH4_OPYC_SRES_A2)

**Data Access**  
[http://cera-www.dkrz.de/WDCC/ui/EntryList.jsp?acronym=EH4\\_OPYC\\_SRES\\_A2](http://cera-www.dkrz.de/WDCC/ui/EntryList.jsp?acronym=EH4_OPYC_SRES_A2)

**Summary**  
The SRES data sets were published by the IPCC in 2000 and classified into four different scenario families (A1, A2, B1, B2). SRES-A2 storyline describes a very heterogeneous world with the underlying theme of self-reliance and preservation of local identities. It results in this scenario continuous increasing population together with a slower economic growth and technological change. The model consists of the atmospheric component which based on the weather forecast model of ECMWF. The atmospheric component is the standard model version of a 19-level hybrid sigma-pressure coordinate system.

**TIBORDER Katalog**

suchen [und] | Person, Autor | sortiert nach | Erscheinungsjahr

stendel, martin | suchen

Kundennummer: | Datenbanken

Suchgeschichte | Kurzliste | Titledaten

Ihre Aktion suchen [und] (Person, Autor) stendel, martin 1 von 8

**Titel:** ECHAM4\_OPYC\_SRES\_A2: 110 years coupled A2 run 6h values / World Data Center for Climate (WDCC), Hamburg. **Martin Stendel**; Torben Schmith; Erich Roeckner ...

**Beteiligt:** Körperschaft: **Martin Stendel** ; Torben Schmith ; Erich Roeckner

**Erschienen:** World Data Center for Climate (WDCC) 2003-01-26

**Umfang:** 961261722720 Bytes.

**Anmerkung:** mode: abstract  
resourceType: PrimaryData  
issueNumber: 1

**Inhalt:** This is a transient scenario experiment where changes in greenhouse gas concentrations and sulfur emissions from 1990 until 2100 are prescribed according the SRES A2 scenario (IPCC, Third Assessment Report). The model experiment is initialized in 1990 using data from the 'GSDIO' experiment described in Roeckner et al. (1999) J. Climate 12, 3004-3032. The model consists of the atmospheric component ECHAM4 (MPI report No. 218) and the isopycnal ocean model OPYC (DKRZ technical report No. 7). The coupling procedure and results are documented in Roeckner et al. (1996), Climate Dynamics 12, 737-754 and in Bacher et al. (1998) Climate Dynamics 14, 431-450.

**Technische Angaben:format:** GRIB

**Bestandsinfo:** <B9> Hannover, Universitätsbibliothek Technische Informationsbibliothek  
Volltext/Image: [Anzeigen](#) [lizenzpflichtig!](#)

1 von 8

ice 2014 / Lautenschlager (WDCC/DKRZ)

**DataCite**

# Think global – act local

## **Science is global**

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

## **Science is carried out locally**

- By local scientist
- Being part of local infrastructures
- Having local funders

# DataCite

Global consortium carried by local institutions  
focused on improving the scholarly infrastructure  
around datasets and other non-textual  
information

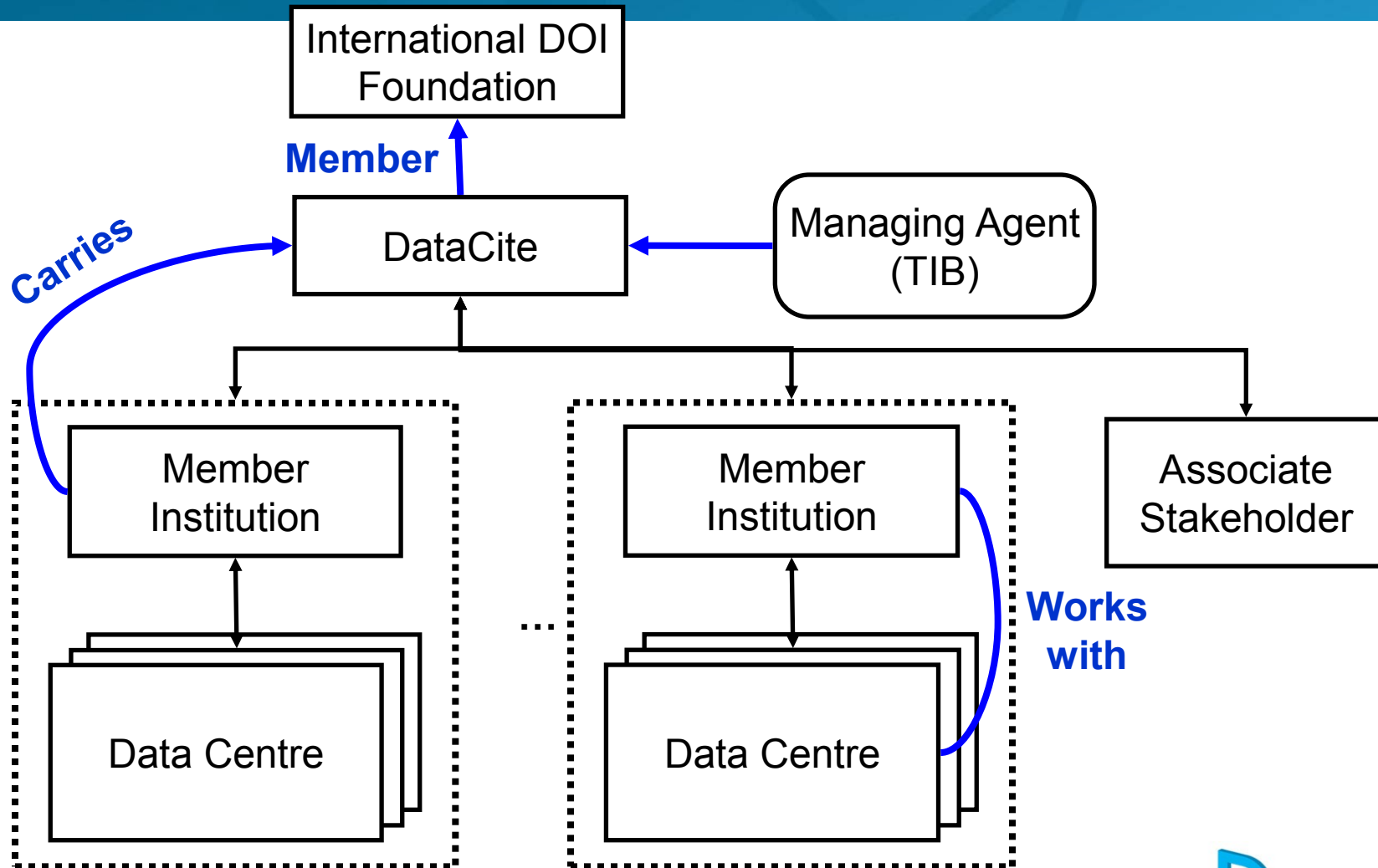
focused on working with data centres and  
organisations that hold content

Providing standards, workflows and best-practice

Initially, but not exclusively based on the DOI system

Founded December 1st 2009 in London

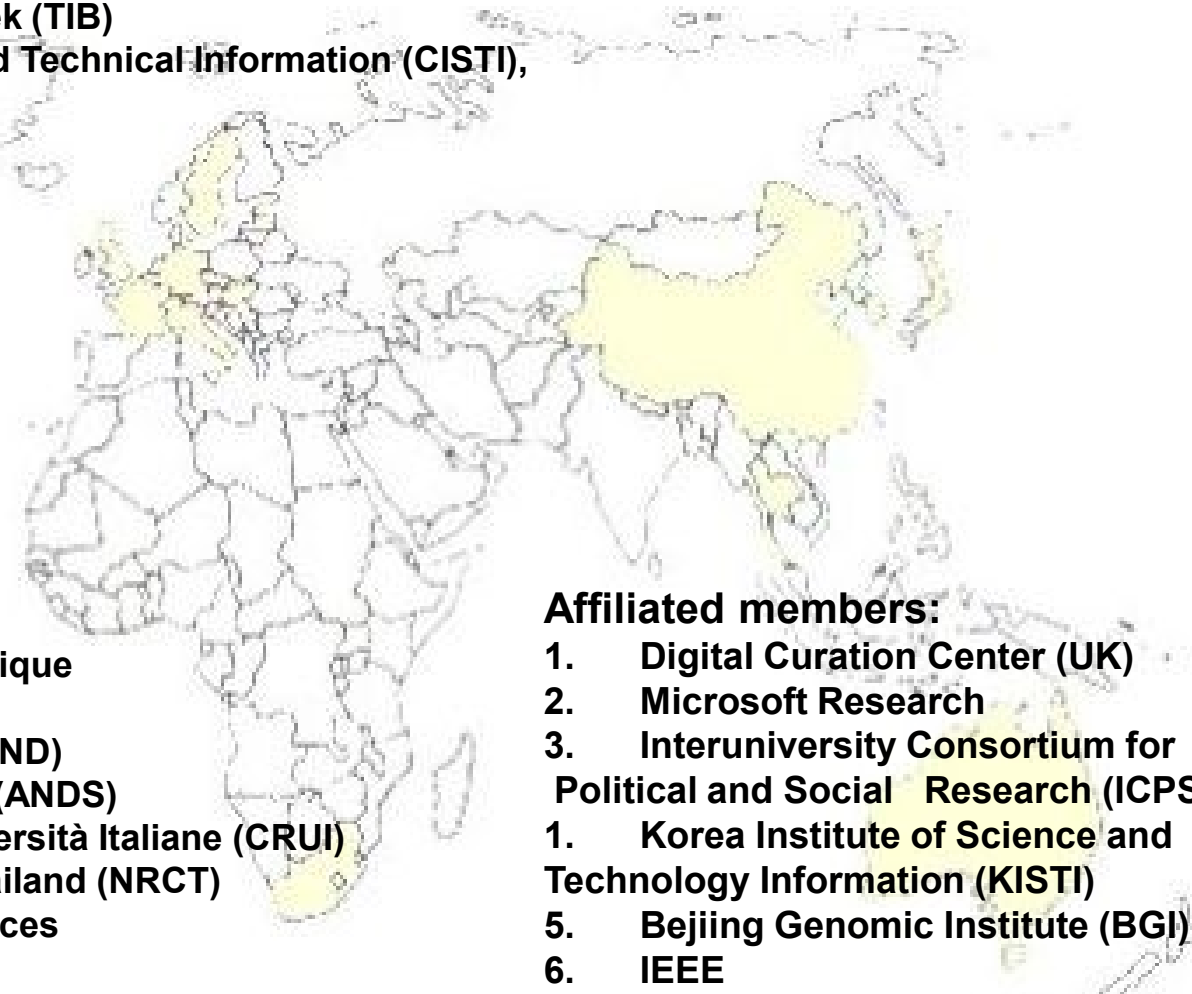
# DataCite structure





# DataCite members

1. Technische Informationsbibliothek (TIB)
2. Canada Institute for Scientific and Technical Information (CISTI),
3. California Digital Library, USA
4. Purdue University, USA
5. Office of Scientific and Technical Information (OSTI), USA
6. Library of TU Delft, The Netherlands
7. Technical Information Center of Denmark
8. The British Library
9. ZB Med, Germany
10. ZBW, Germany
11. Gesis, Germany
12. Library of ETH Zürich
13. L'Institut de l'Information Scientifique et Technique (INIST), France
14. Swedish National Data Service (SND)
15. Australian National Data Service (ANDS)
16. Conferenza dei Rettori delle Università Italiane (CRUI)
17. National Research Council of Thailand (NRCT)
18. The Hungarian Academy of Sciences
19. University of Tartu, Estonia
20. Japan Link Center (JaLC)
21. South African Environmental Observation Network (SAEON)
22. European Organisation for Nuclear Research (CERN)



## Affiliated members:

1. Digital Curation Center (UK)
2. Microsoft Research
3. Interuniversity Consortium for Political and Social Research (ICPS)
1. Korea Institute of Science and Technology Information (KISTI)
5. Beijing Genomic Institute (BGI)
6. IEEE
7. Harvard University Library
8. World Data System (WDS)
9. GWDG

# What type of data are we talking about?

Earth quake events =>

[doi:10.1594/GFZ.GEOFON.gfz2009kciu](https://doi.org/10.1594/GFZ.GEOFON.gfz2009kciu)

Climate models => [doi:10.1594/WDCC/dphase\\_mpeps](https://doi.org/10.1594/WDCC/dphase_mpeps)

Sea bed photos => [doi:10.1594/PANGAEA.757741](https://doi.org/10.1594/PANGAEA.757741)

Distributes samples => [doi:10.1594/PANGAEA.51749](https://doi.org/10.1594/PANGAEA.51749)

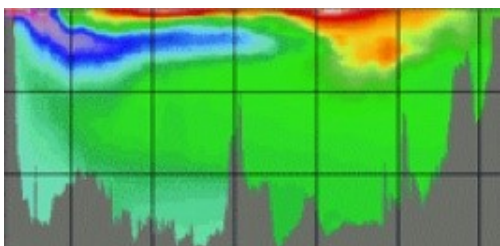
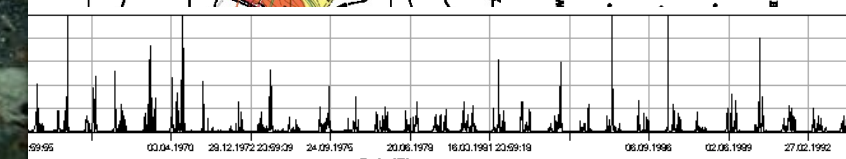
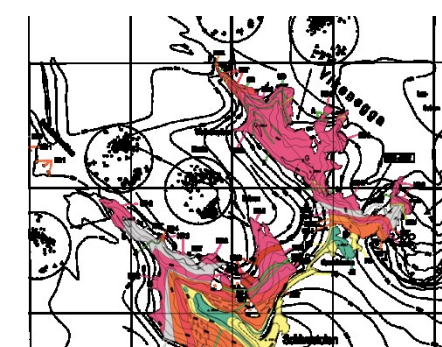
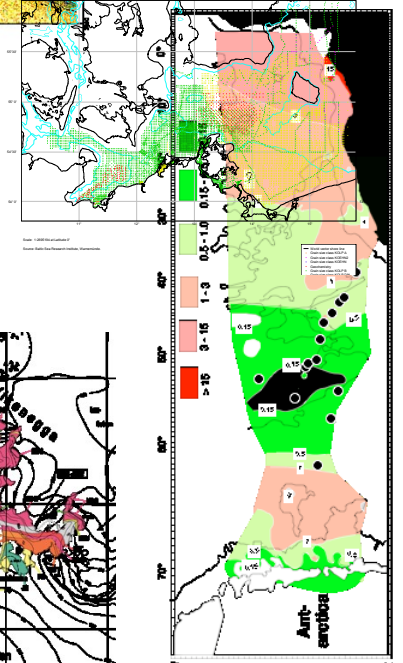
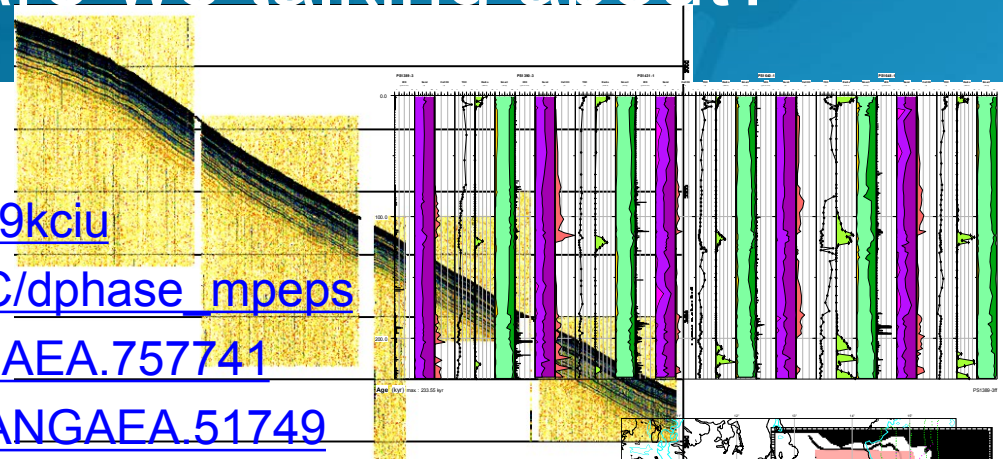
Medical case studies => [doi:10.1594/eaacinet2007/CR/5-270407](https://doi.org/10.1594/eaacinet2007/CR/5-270407)

Computational model => [doi:10.4225/02/4E9F69C011BC8](https://doi.org/10.4225/02/4E9F69C011BC8)

Audio record => [doi:10.1594/PANGAEA.339110](https://doi.org/10.1594/PANGAEA.339110)

Grey Literature => [doi:10.2314/GBV:489185967](https://doi.org/10.2314/GBV:489185967)

Videos => [doi:10.3207/2959859860](https://doi.org/10.3207/2959859860)



**Anything that is the foundation  
of further reserach  
is research data**

**Data is evidence**

# DataCite in 2014

Over 3,700,000 DOI names registered so far.

350 data centers.

14,000,000 resolutions so far in 2014.

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

DataCite MetadataStore

<http://search.datacite.org>

# OAI and Statistics

OAI Harvester

<http://oai.datacite.org>

DataCite statistics (resolution and registration)

<http://stats.datacite.org>

Content negotiation (with CrossRef)

<http://www.crosscite.org/cn/>

# 2012: STM, CrossRef and DataCite Joint Statement

1. To improve the availability and findability of research data, the signers encourage authors of research papers to **deposit researcher validated data in trustworthy and reliable Data Archives**.
2. The Signers encourage Data Archives to **enable bi-directional linking between datasets and publications** by using established and community endorsed unique persistent identifiers such as database accession codes and DOI's.
3. The Signers encourage publishers and data archives to make visible or increase **visibility of these links** from publications to datasets and vice versa

# Example

## **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>

# More Data example

- Higgs particle

ATLAS Collaboration ( 2013 ) HepData,

<http://doi.org/10.7484/INSPIREHEP.DATA.A78C.HK44>

- ECOLI outbreak

Li, D et al (2011):

*Genomic data from Escherichia coli O104:H4 isolate TY-2482.*

BGI Shenzhen.

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



# Latest developments

ODIN project with ORCID.

<http://datacite.labs.orcid-eu.org/>

MoU with Thomson reuters to cooperate on data citation index

DataCite plugin for next D-Space release

Agreement with Re3Data and DataBib to include their service in 2016

MoU with RDA to become organisational affiliate

Joint Declaration of Data Citation Principles

<https://www.force11.org/datacitation>

# Thank you!

