

DataCite – A global approach for better data sharing

Jan Brase
DataCite

Towards linked science - Open Data and DataCite Estonia
University of Tartu Library
October 23rd 2014



Science Paradigms

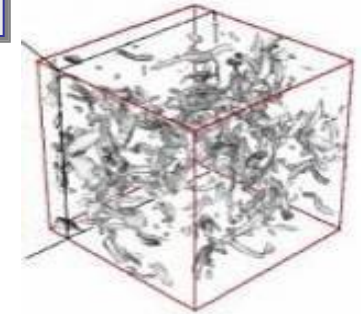
Thousand years ago:
science was **empirical**
describing natural phenomena



Last few hundred years:
theoretical branch
using models, generalizations

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K \frac{c^2}{a^2}$$

Last few decades:
a **computational** branch
simulating complex phenomena



Today:
data exploration (eScience)
unify theory, experiment, and simulation



Consequences for Libraries

Scientific Information is more than a journal article or a book

Libraries should open their catalogues to any kind of information

The catalogue of the future is NOT ONLY a window to the library's holding, but

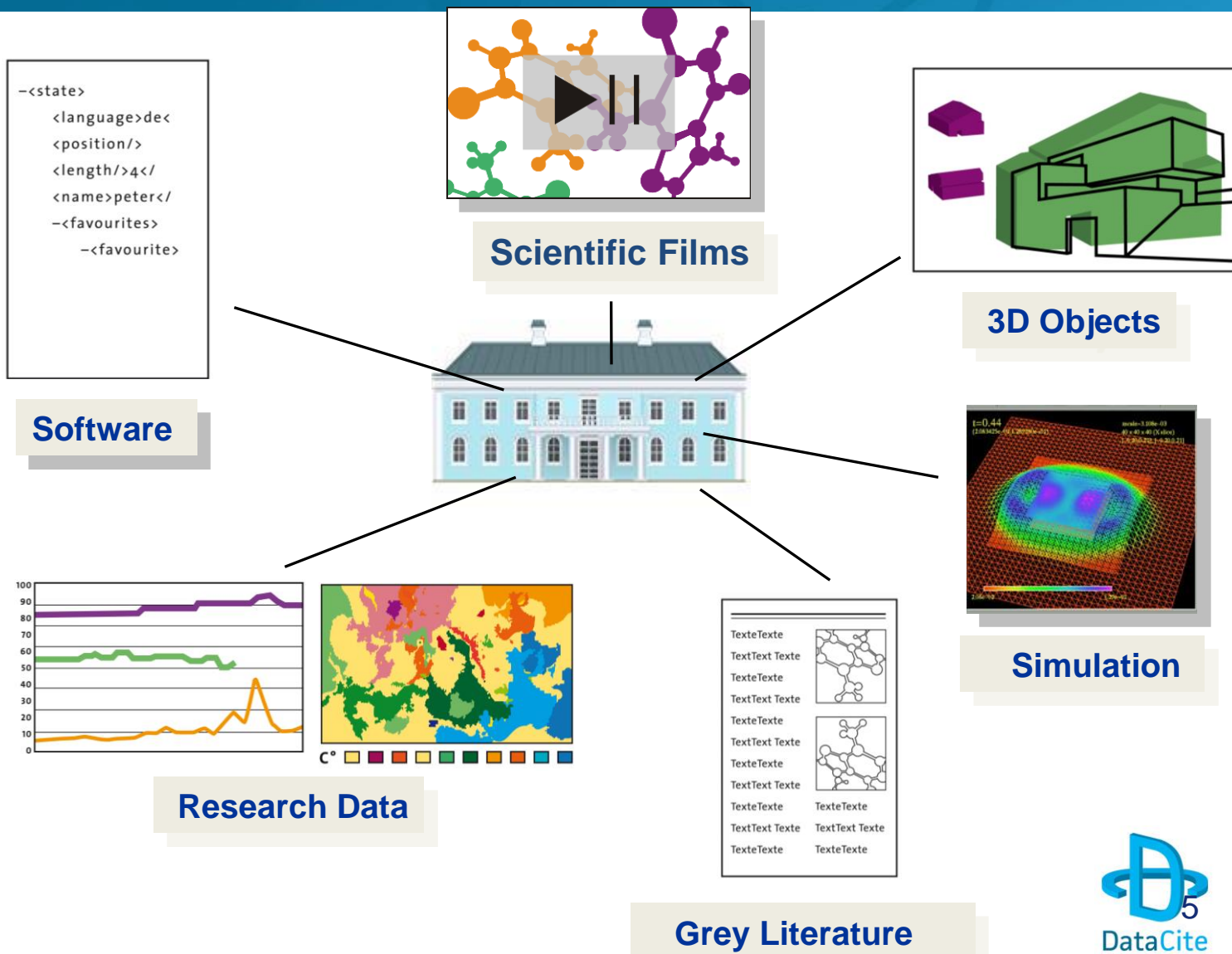
A portal in a net of trusted providers of scientific content

We do not have it
BUT

We know where you can find

And here is the link to it!

Including non-classical publications



Why is this a role for libraries?

- Libraries have a history in bringing scientific information to the public
- Libraries have a tendency to be persistent
 - A project will be forgotten in 40 years, the library will very likely still exist then
- Libraries are very trustworthy organisations

DataCite



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

How to achieve this?

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

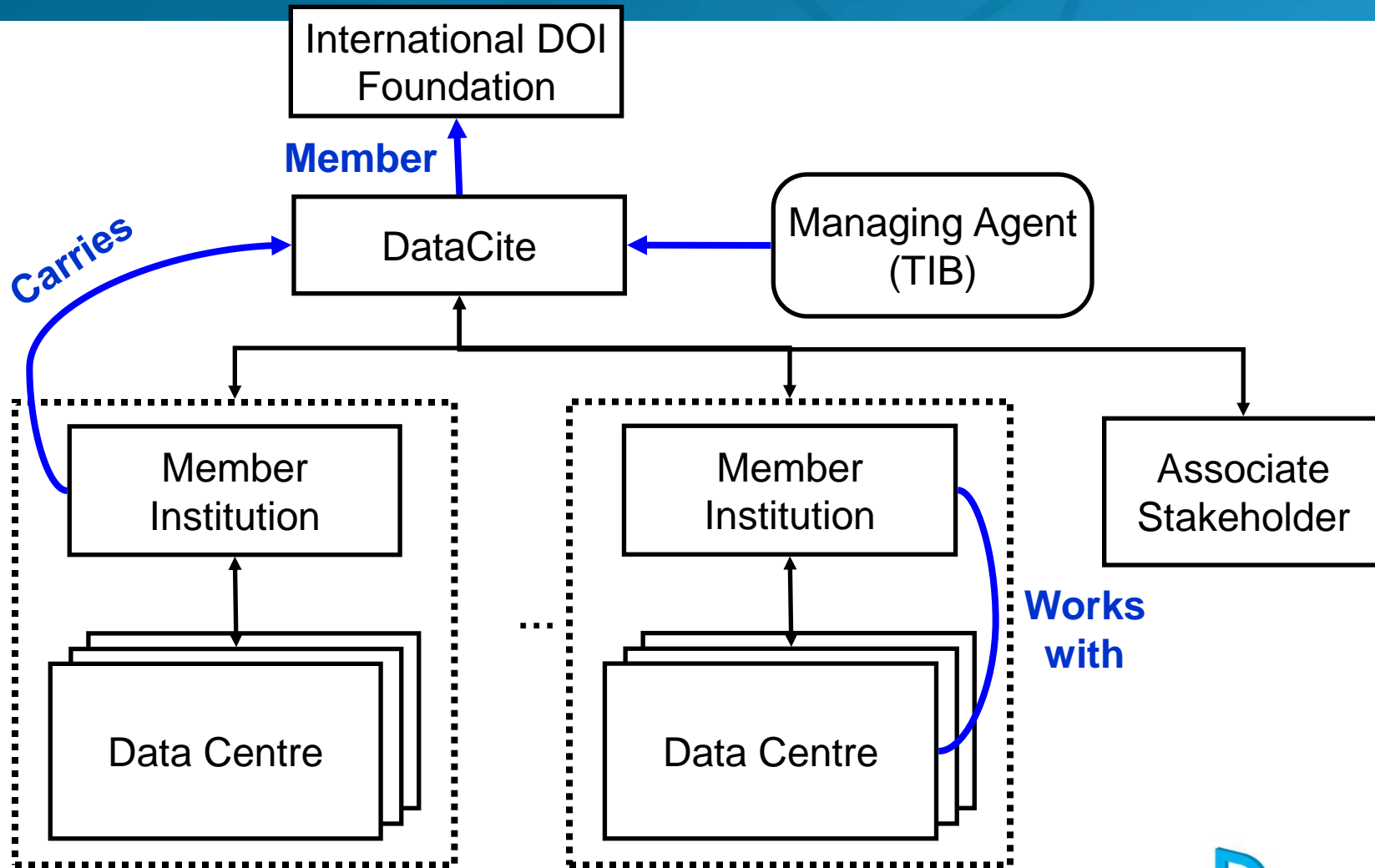
focused on working with data centres and
organisations that hold content

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/WDC/DPhase_mpeps](https://doi.org/10.1594/WDC/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)

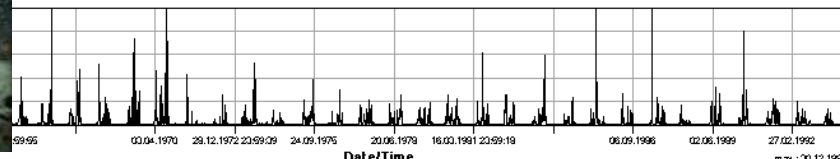
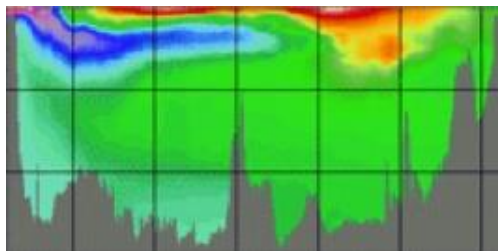
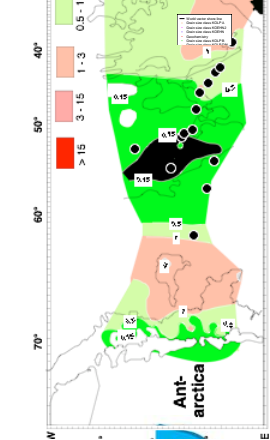
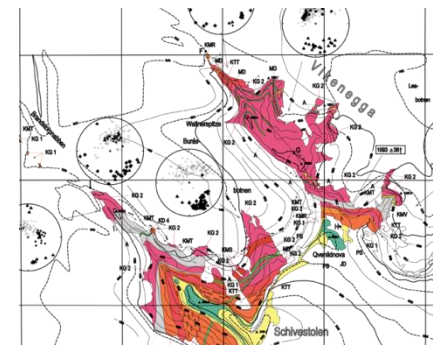
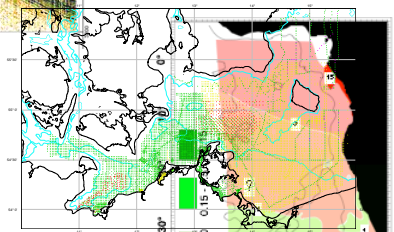
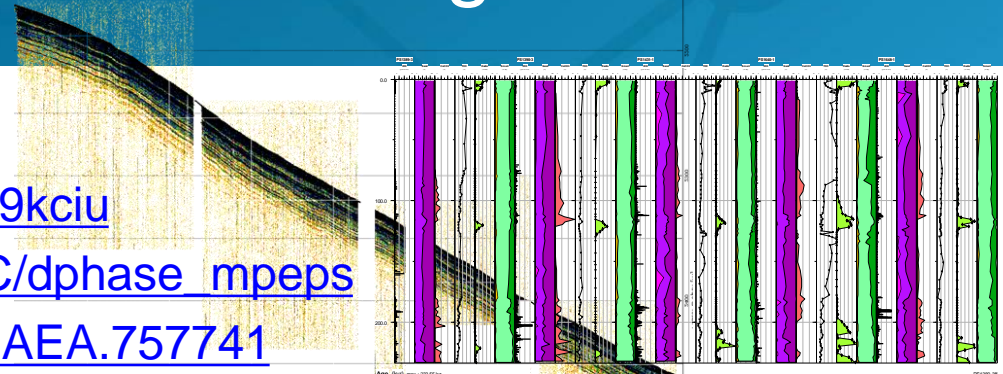
Medical case studies => [doi:10.1594/eaacinet2007/CR/5-270407](https://doi.org/10.1594/eaacinet2007/CR/5-270407)

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>

DataCite search

Searchterm: *

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

Searchterm: relatedIdentifier:*

Searchterm:
relatedIdentifier:issupplementto\:10.1029*

Searchterm:relatedIdentifier:*\:10.1055*

OAI and Statistics

OAI Harvester

<http://oai.datacite.org>

DataCite statistics (resolution and registration)

<http://stats.datacite.org>

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>

Identifier combination for fragments

Media fragment identifier (MFID), W3C

- <http://doi.org/10.5446/12780> (video)
- <http://doi.org/10.5446/12780#t=00:20,00:27>

DOI resolver agnostic solution for media fragments

Thank you!

