

# Embracing research data

Ingrid van de Stadt, Elsevier

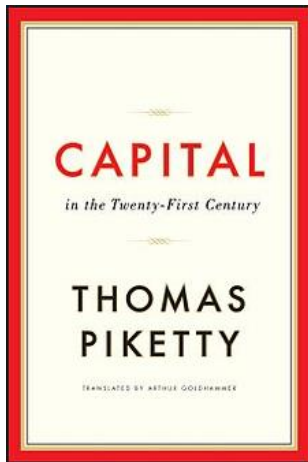
• 27 November 2016

# Embracing research data

- **Some backgrounds on Research Data**
- **Tools and programmes supporting research data**
  - Data repository (Mendeley Data)
  - Data journals
  - Linking-data programme
  - Data search
  - Research protocols (HiveBench)
  - Industry standards
- **Elsevier's Research Data Policy**

# Impact of Data Sharing

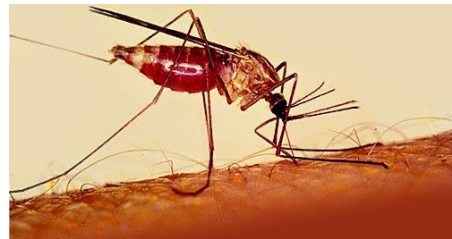
## Social Science



Raw & Normalized Data, Analyses, and Methods have all been made publicly available on a dedicated website

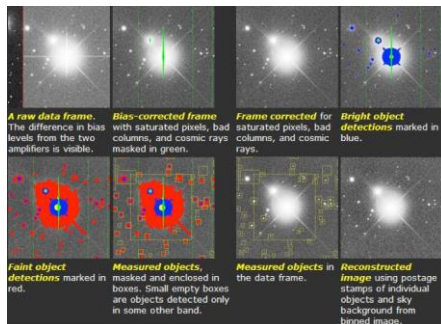
Generated great discussions and ideas about the topic

## Medical Science



WWARN, the first malaria data sharing network, has used pooled analysis of shared data to provide evidence to **help improve dosing regimens** of malaria treatments

## Astronomy



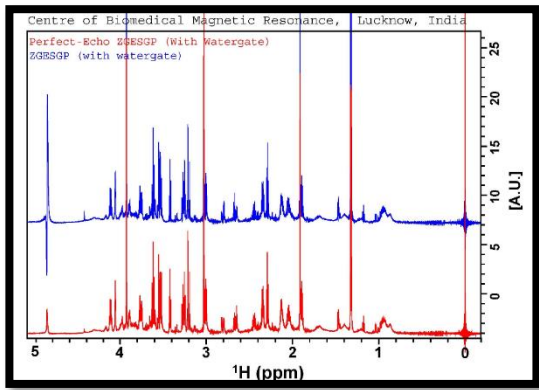
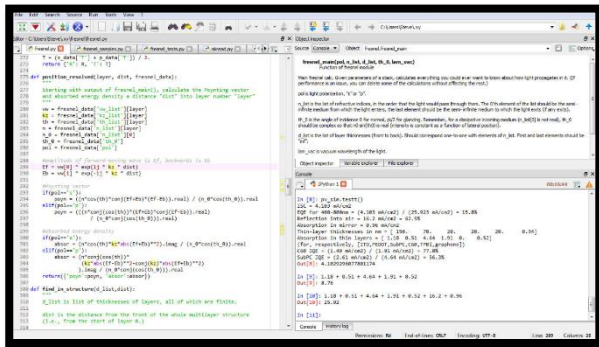
Sloan Digital Sky Survey (SDSS), University of Washington.  
Early data releases of a Telescope Observatory

# When we talk about data, we really talk about the following:

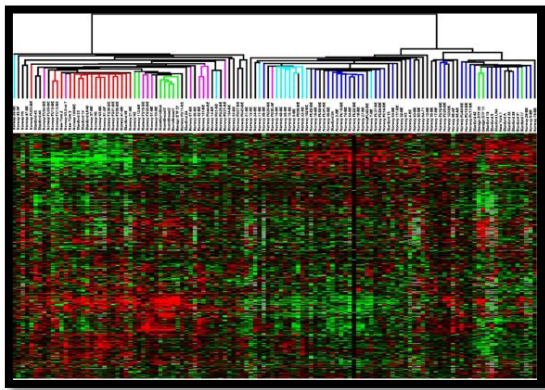
Machine & environment settings



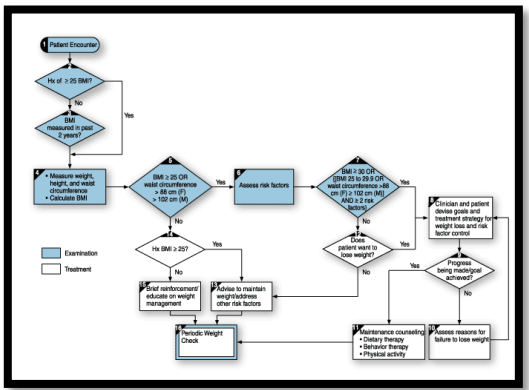
Scripts & analyses



Raw data

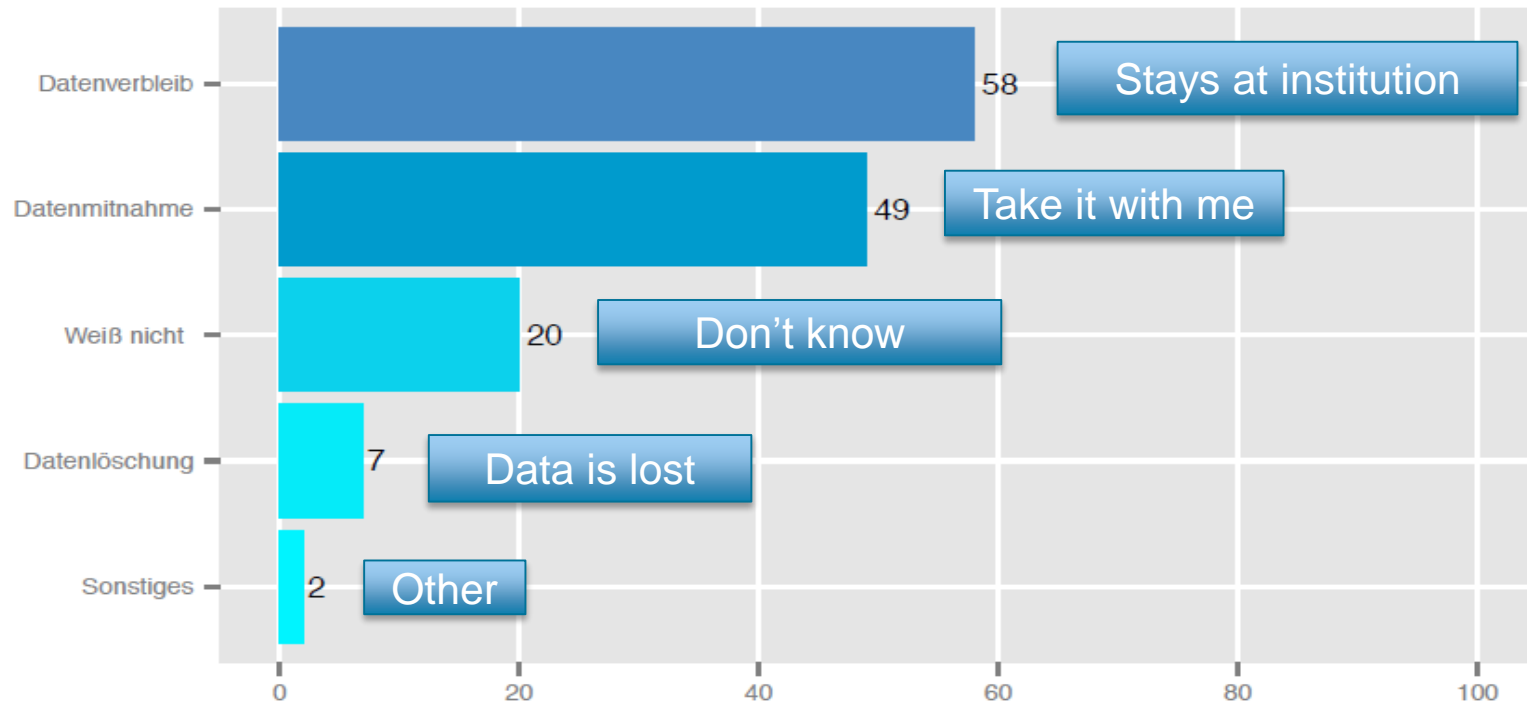


Processed data

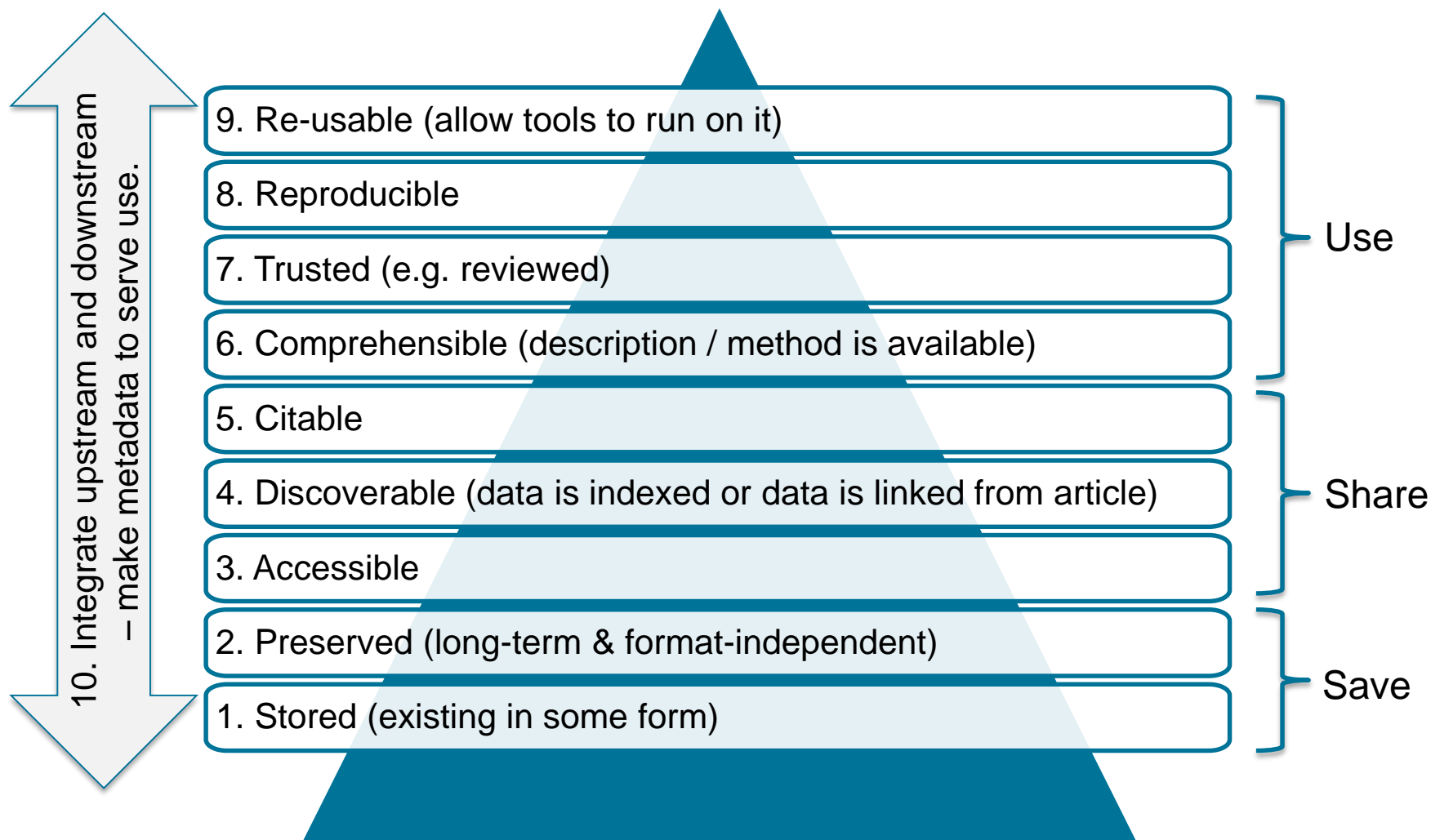


Protocols, methods, algorithms

# When You Leave Your Institution, What Happens To Your Data?



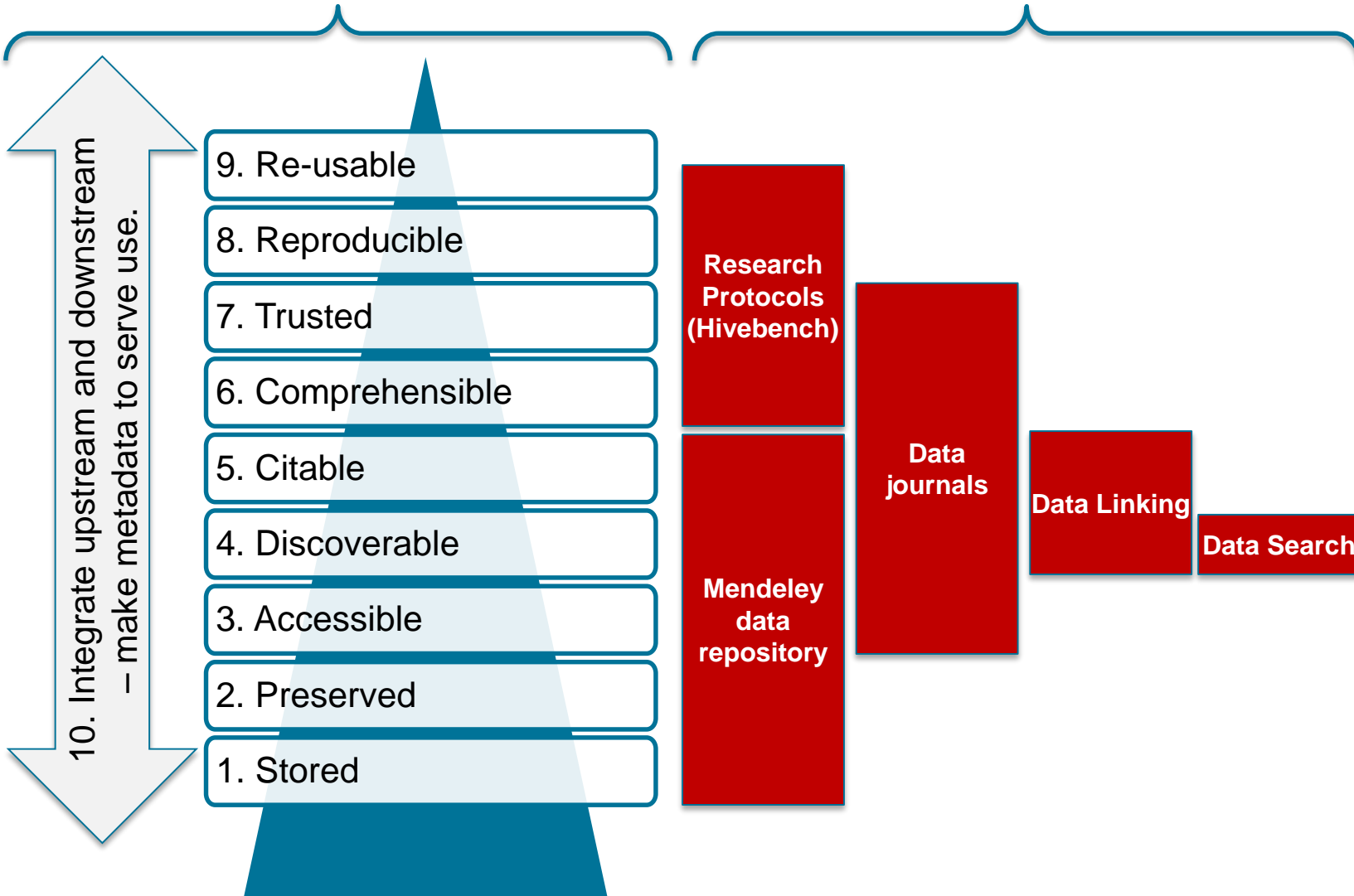
## The 10 components for effective research data



# Tools and programmes supporting research data

# The 10 components for effective research data

# Elsevier initiatives





# Manage, store: Mendeley Data

<http://data.mendeley.com/>

An open repository for posting & reusing research data

MENDELEY DATA <sup>Beta</sup> Browse My datasets New dataset Mike Jones

## Put your research data online today

so it can be cited, shared and secure

Start uploading

## ❖ Discoverable and citable through DOI's

### For datasets big and small

Store your research data online

Quickly and easily upload files of any type and we will host your research data for you. Your experimental research data will have a permanent home on the web that you can refer to.



# Manage, Store: Mendeley Data

MENDELEY DATA <sup>Beta</sup> Browse My datasets New dataset Log in Create account

## Reproducible experiments on dynamic resource allocation in cloud data centers

DOI: 10.17632/xz6gv65m6d.6  
Contributor(s): Andreas Wolke

**Description of this data**  
In Wolke et al. we compare the efficiency of different resource allocation strategies experimentally. We focused on dynamic environments where virtual machines need to be allocated and deallocated to servers over time. In this companion paper, we describe the simulation framework and how to run simulations to replicate experiments or run new experiments within the framework.

**Experiment data files** Download all files (6)

- Results.zip 63 KB
- CSV files with simulation and experimentation results.
- github.paper.IS2015-master.zip 8 MB  
<https://github.com/jacksonson/paper.IS2015/tree/7165452f4e9c540f98e1e57058de06f9fb192eBf>
- github.workload-master.zip 222 MB  
<https://github.com/jacksonson/workload/tree/713dc5382b82e4ec1e1b6a998c80af3f7c08219f>
- Dockerfile 1 KB  
Used to create the Docker container provided in IS2015.tar.gz
- IS2015.tar.gz 1.3 GB  
Docker container file with installed simulation framework. Run simulations: (cd /root/work/paper.IS2015/control/Control && ./startsim\_reprozip) Run analysis: (cd /root/work/paper.IS2015/analysis && ./startanalysis-sim)
- reprozip.rpz 160 MB  
ReproZip package of the simulation executed in the Docker container.

**Version 6** | Published: 13 Dec 2015

This data is associated with the following peer reviewed publication:

Reproducible experiments on dynamic resource allocation in cloud data centers

Cite this article

Published in:  
Information Systems

**latest version**

**Version 6** 2015-12-13  
Published: 2015-12-13  
DOI: 10.17632/xz6gv65m6d.6

Cite this dataset

Wolke, Andreas (2015), "Reproducible experiments on dynamic resource allocation in cloud data centers", Mendeley Data, v6  
<http://dx.doi.org/10.17632/xz6gv65m6d.6>

**Previous versions**

Version 5	2015-11-21
Version 4	2015-11-16
Version 3	2015-11-14
Version 2	2015-11-14
Version 1	2015-10-12

Version comparison

Version 5

Linked to published papers – or not

Linked to Github – or not

Citable data and Versioning

# Data journals: SoftwareX

Home > Books & Journals > SoftwareX

## SoftwareX

Editors-in-Chief: Dr. Kate Keahey, Dr. Frank Seinstra, Dr. David Wallom  
View full editorial board

Open Access



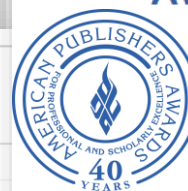
ISSN: 2352-7110



### Code metadata

Current code version	v0.6
Permanent link to code/repository used of this code version	<a href="https://github.com/ElsevierSoftwareX/SOFTX-D-15-00005">https://github.com/ElsevierSoftwareX/SOFTX-D-15-00005</a>
Legal Code License	<i>NCSA open source license</i>
Code versioning system used	<i>git</i>
Software code languages, tools, and services used	<i>C, C++, Python, Bash; MPI, OpenMP, CUDA</i>
Compilation requirements, operating environments & dependencies	<i>Compilers: GNU/Intel/Cray; OS: Linux (RedHat, Debian, Ubuntu, CentOS, SUSE); Dependencies: GDAL, GEOS, PROJ4, SPRNG, PySAL, OpenGeoDa, etc.</i>
If available Link to developer documentation/manual	<a href="https://github.com/cybergis/cybergis-toolkit">https://github.com/cybergis/cybergis-toolkit</a>  <a href="http://cybergis.cigi.uiuc.edu/cyberGISwiki/doku.php/ct">http://cybergis.cigi.uiuc.edu/cyberGISwiki/doku.php/ct</a>
Support email for questions	CyberGIS Helpdesk ( <a href="mailto:help@cybergis.org">help@cybergis.org</a> )

Table options ▼

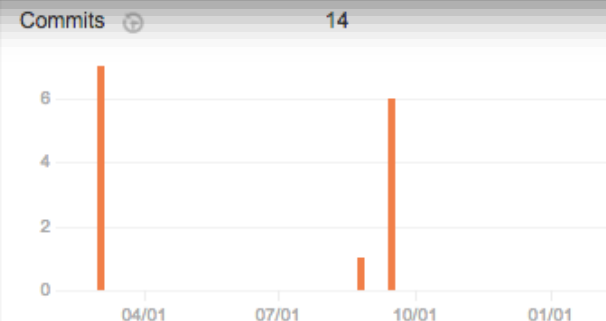


## AWARD FOR INNOVATION IN JOURNAL PUBLISHING



Elsevier  
SoftwareX

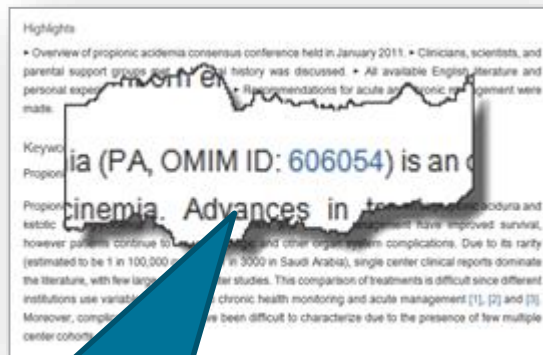
Editors Dr. Kate Keahey, Dr. Frank Seinstra and Professor David Wallom



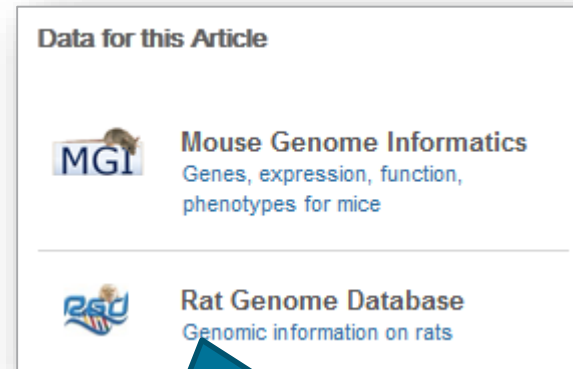
Powered by **GitHub** and **Scopus**

## Data-linking program

- Elsevier has an extensive programme with 60+ leading domain-specific data repositories to interlink articles and data
- Makes it easier to find relevant data and place data into the right context
- Linking through in-article accession numbers, data DOI's, or data banners



Linking through in-article data accession numbers



Database banners shown next to the article on ScienceDirect

# Data-linking program – example Pangaea



## Calcium carbonate corrosiveness in the South Atlantic during the Last Glacial Maximum as inferred from changes in the preservation of *Globigerina bulloides*: A proxy to determine deep-water circulation patterns?

A.N.A Volbers, R Henrich

University of Bremen, Faculty of Geosciences, Department of Paleoceanography and Sedimentology, P.O.Box 330440, D-28334 Bremen, Germany

### Abstract

The modern Atlantic Ocean, dominated by the interactions of North Atlantic Deep Water (NADW) and Antarctic Bottom Water (AABW), plays a key role in redistributing heat from the Southern to the Northern Hemisphere. In order to understand the evolution of the relative importance of these two water masses...



- Supplementary data at PANGAEA
- Bidirectional links between PANGAEA & ScienceDirect
- Data visualized next to the article



# In-article data visualization: iPlots



Download PDF

Export

More options...

Search ScienceDirect



Advanced search

- Achieving smaller particle size during precipitation and aging is critical to obtain high performance catalysts.
- Higher reduction temperatures (of larger CuO crystallites) promote sintering of Cu crystallites.

## Abstract

Binary Cu/ZnO methanol synthesis catalysts were prepared by the co-precipitation of copper and zinc hydroxycarbonates using different initial solution concentrations, stirring rates and aging times and temperatures during precipitation, and different calcination temperatures of the precipitated hydroxycarbonates. The precipitates (catalyst precursors), fresh catalyst and were characterized at appropriate stages by nitrogen adsorption-desorption isotherm, X-ray diffraction (XRD), temperature

prog  
All ca  
of co  
ratio  
to gre  
after  
by in  
but th  
of ag

0.75 h of aging, and the catalyst activity increased with aging time followed by a decrease upon further aging. At a constant aging time of 5 h, the catalyst activity increased with temperature in the range of 40–60 °C and then decreased when temperature rose further from 60 to 80 °C. The highest methanol synthesis activity (555 g<sub>MeOH</sub>/kg<sub>cat</sub> h) was observed for the catalyst prepared from 1 M initial solution, 500 rpm stirring rate and aged at 60 °C. This was attributed to the small CuO crystallite size and large Cu surface area of the resulting catalyst.

Graphical abstract

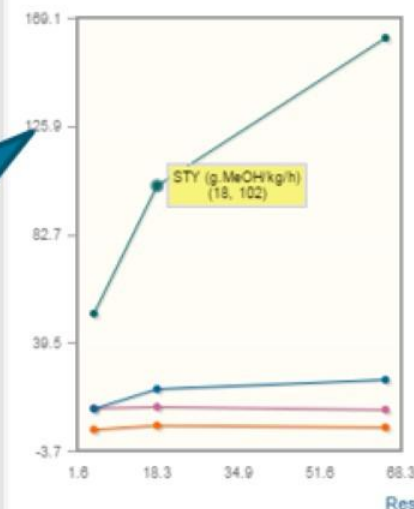
Displaying plot data in CSV format delivered by the author as supplementary material. Allows to access, explore, and download data behind plots.

## Interactive plots for this article



Plot

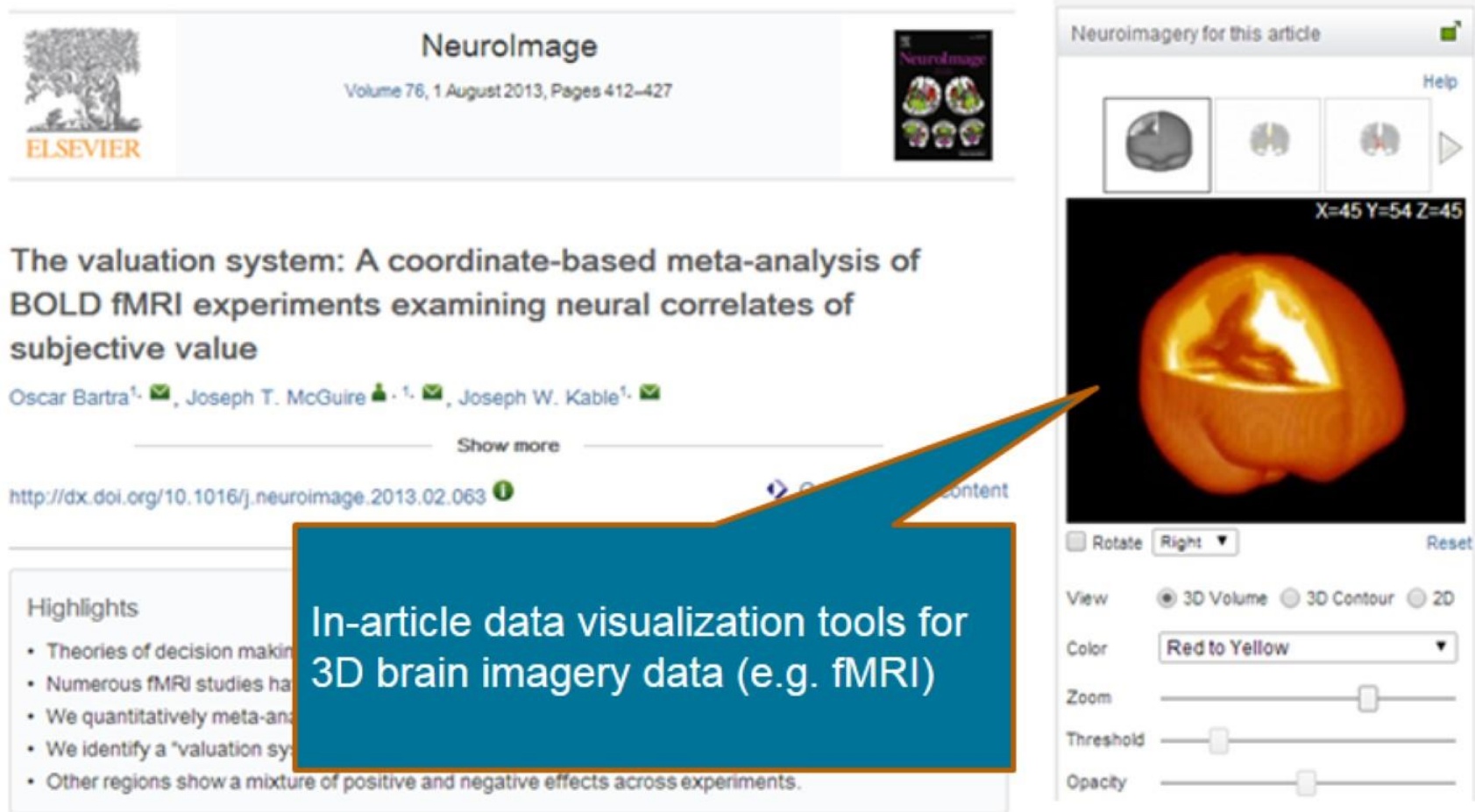
Data table



X axis Aging Time (h)

- Y axis
- CuO crystallite size (nm)
  - Cu surface area (m<sup>2</sup>/gr)
  - Particle size-D0.5 (micro-m)
  - STY (g MeOH/kg/h)

# Data visualization: in-article data viewers to allow data exploration with the article



The screenshot shows a NeuroImage article page. The article title is "The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value". The authors listed are Oscar Bartra, Joseph T. McGuire, and Joseph W. Kable. A URL is provided: <http://dx.doi.org/10.1016/j.neuroimage.2013.02.063>. A "Highlights" section is visible on the left, and a "Neuroimaging for this article" viewer is on the right. The viewer displays a 3D brain volume with a color scale from red to yellow. The viewer includes controls for rotation, view (3D Volume, 3D Contour, 2D), color, zoom, threshold, and opacity. A blue callout box points to the viewer with the text: "In-article data visualization tools for 3D brain imagery data (e.g. fMRI)".

NeuroImage  
Volume 76, 1 August 2013, Pages 412–427

The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value

Oscar Bartra<sup>1</sup>, Joseph T. McGuire<sup>1</sup>, Joseph W. Kable<sup>1</sup>

<http://dx.doi.org/10.1016/j.neuroimage.2013.02.063>

Highlights

- Theories of decision making
- Numerous fMRI studies have
- We quantitatively meta-anal
- We identify a "valuation sy
- Other regions show a mixture of positive and negative effects across experiments.

Neuroimaging for this article

X=45 Y=54 Z=45

Rotate Right

View  3D Volume  3D Contour  2D

Color Red to Yellow

Zoom

Threshold

Opacity

In-article data visualization tools for 3D brain imagery data (e.g. fMRI)

# Elsevier Data Search

Beta  
DataSearch

*Search for research data across domains and types, from many domain-specific, cross-domain and institutional data repositories.*

Find research data



Or Try: cosmic microwave radiation, frog phylogeny or protein structure prediction



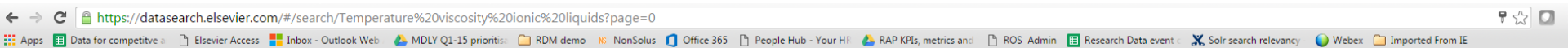
DataSearch.Elsevier.com

1. Across repositories
2. (Deep) indexing of data, so not just metadata
3. Data preview



# Elsevier Data Search

## E.g. search for "Temperature viscosity ionic liquids"



DataSearch  datasearch

- Filter Results** reset
- 1620284 results for *Temperature viscosity ionic liquids*
- Types**
- Image (1415019)
  - Tabular Data (872183)
  - Document (55138)
  - File Set (2985)
  - Raw Data (1331)**
  - Video (1174)
  - Slides (825)
  - Software (8)
  - Statistical Data (7)

- Solvent Properties of Functionalized Ionic Liquids for CO<sub>2</sub> Absorption**  
*L.M. Galán Sánchez, G.W. Meindersma & A.B. de Haan - 2006-10-24*  
 Ionic liquids can be used as solvents for gas absorption operations in order to improve the process economy and general efficiency of gas separations. This work investigates solvent properties of ionic liquids and compares them to amine solutions used for absorption of carbon dioxide (CO<sub>2</sub>). The CO<sub>2</sub> solubility into six different room temperature...
- Ion conductive characteristics of ionic liquids prepared by neutralization of alkyimidazoles**  
*Hiroyuki Ohno & Masahiro Yoshizawa - 2002-03-20*  
 A wide variety of ionic liquids was prepared by the neutralization of five kinds of imidazole derivatives and nine kinds of acids. Their physical and chemical properties such as melting points, glass transition temperature, viscosity, and ionic conductivity were studied. Among these, imidazoles neutralized with imide-type acids were revealed to have...
- CO<sub>2</sub> removal with 'switchable' versus 'classical' ionic liquids**  
*E. Privalova, M. Nurmi, M.S. Marañón, E.V. Murzina, P. Maki-Arvela, K. Eränen, D.Yu. Murzin & J.-P. Mikkola*  
 ► Comparison of ionic liquid systems acting as chemical vs. physical solvents in practical terms. ► Recycling and reuse issues of ionic liquids in carbon dioxide capture. ► Introduction of new types of switchable ionic liquids. ► Studying the capture-release cycle behavior of the aforementioned ionic liquids. ► In essence, we have focused on engineerin...

**Temperature dependence of viscosity for room temperature ionic liquids**

- Description
- Table 4** >
- Table 8
- Fig. 5
- Table 3
- Table 2

RTIL	T/ °C									
	10	15	20	25	30	35	40	50	60	70
[Triethylsulfonium][Imide]	56	46	39	33 (30) [49]	28	24	21	17	13	11
[1-Butyl-3-methylimidazolium][Imide]	84	69	58 (52) [36]	47	41	34	29	22	17	14
[1-Butyl-1-methylpyrrolidinium][Imide]	128	106	89	76 (70) [50]	64	54	46	34	26	20
[1-Butyl-3-methylimidazolium][CF <sub>3</sub> SO <sub>3</sub> ]	131	107	90 (90) [36]	74	61	52	44	33	25	19

**Table 4**  
  
 Absolute viscosities (in mPas) of air/moisture-tolerant RTILs at 10 temperatures

# Hivebench – an Electronic Laboratory Notebook



Plan and store experiments in one place



Manage and share protocols



Link research data and protocols to your experiments

The screenshot displays the Hivebench web interface. At the top, there is a search bar and navigation tabs for 'Notebook', 'Protocol', 'Inventory', and 'Data'. The main content area is titled 'Notebook' and shows a list of entries under 'My Notebook'. The selected entry is 'Anita's Test' by 'anitawaard123' from 03/26/2014. The details for this entry include the author, laboratory (DeWaard Labs), date (26-03-2014), and a list of duplicates. Below this, there is a section for 'Objectives' with a text input field and a list of steps, including 'Step #1: ie. Mix gently and incubate for 5 minutes at room temperature.'

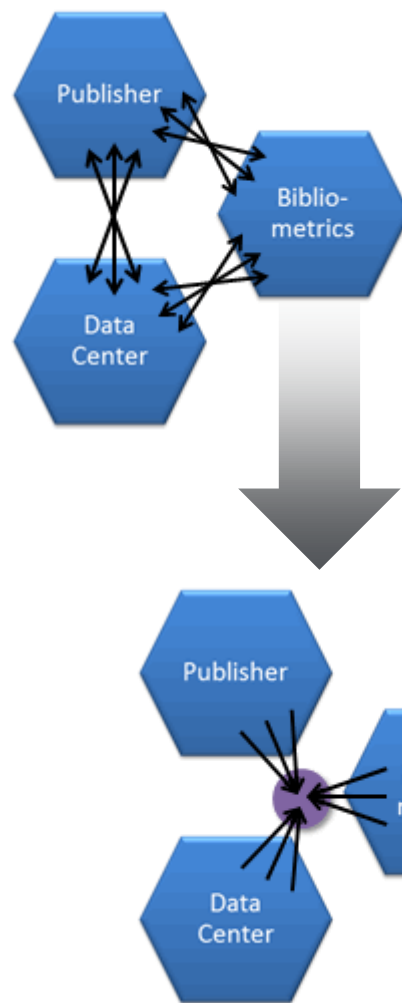


# Development of Industry Standards



- ICSU/WDS/RDA [Publishing Data Service Working group](#)
- Creating linked-data model for exposing DOI to DOI links outside publisher's firewall
- Merged with National Data Service pilot with the same goal
- Collaboration between CrossRef, DataCite, Europe PubMed Central, ANDS, Thompson Reuters, Elsevier, OpenAire

[www.scholix.org](http://www.scholix.org)



Objective: move from

a plethora of (mostly) bilateral arrangements between the different players...

.. to ..

.. **a one-for-all cross-referencing service** for articles and data



**5000+ data repositories!**

## Research Data Policy

Elsevier will:

- Encourage and support researchers and research institutions to **share data where appropriate and at the earliest opportunity**.
- Provide **guidance to authors regarding the deposit and sharing** of data.
- **Encourage and enable two-way linking** of relevant datasets and publications **using permanent standard identifiers**.
- Encourage and **support proper data citation practices** so that researchers can be cited and credited for their work.
- Work closely with the scientific community to **establish data review practices** to ensure that published research data is valid, properly documented and can be re-used.
- **Develop tools and services** to support researchers to **discover, use and reuse** data to further their research.

*“Raw research data should be made freely available to all researchers wherever possible” – STM Brussels Declaration 2007*

**Thank you ! Questions?**

**Contact me at [i.stadt@Elsevier.com](mailto:i.stadt@Elsevier.com)**



## What are we really after: malaria



# WWARN

**WWARN, the first malaria data sharing network, has used pooled analysis of shared data to provide evidence to help improve dosing regimens of malaria treatments**

- **research partners from over over 260 institutions globally have worked with WWARN, and over 120,000 individual patient records have been contributed to the WWARN Data Centre. That equates to around 80% of all the available artemisinin combination therapy trial data.**
- **Based on the results from the the Dihydroartemisinin-Piperaquine (DP) Dose Impact Study Group and pharmacometric modelling of piperaquine, the World Health Organization has revised the recommended dose of DP, a commonly used antimalarial for young children. These revised dose regimens are predicted to provide similar piperaquine concentrations across all age groups.**
- **Similarly, a meta-analysis undertaken by the Artesunate Amodiaquine (ASAQ) Dose Impact Study Group, based on 9,106 patients, found that although the overall efficacy of ASAQ is adequate in most settings, efficacy varies with the formulation and is affected by a range of risk factors including age. The Artemether lumefantrine (AL) Dose Impact Study Group found that cure rates were lowest in young children from Asia, especially those with high parasitemia, and young underweight children from Africa.**