

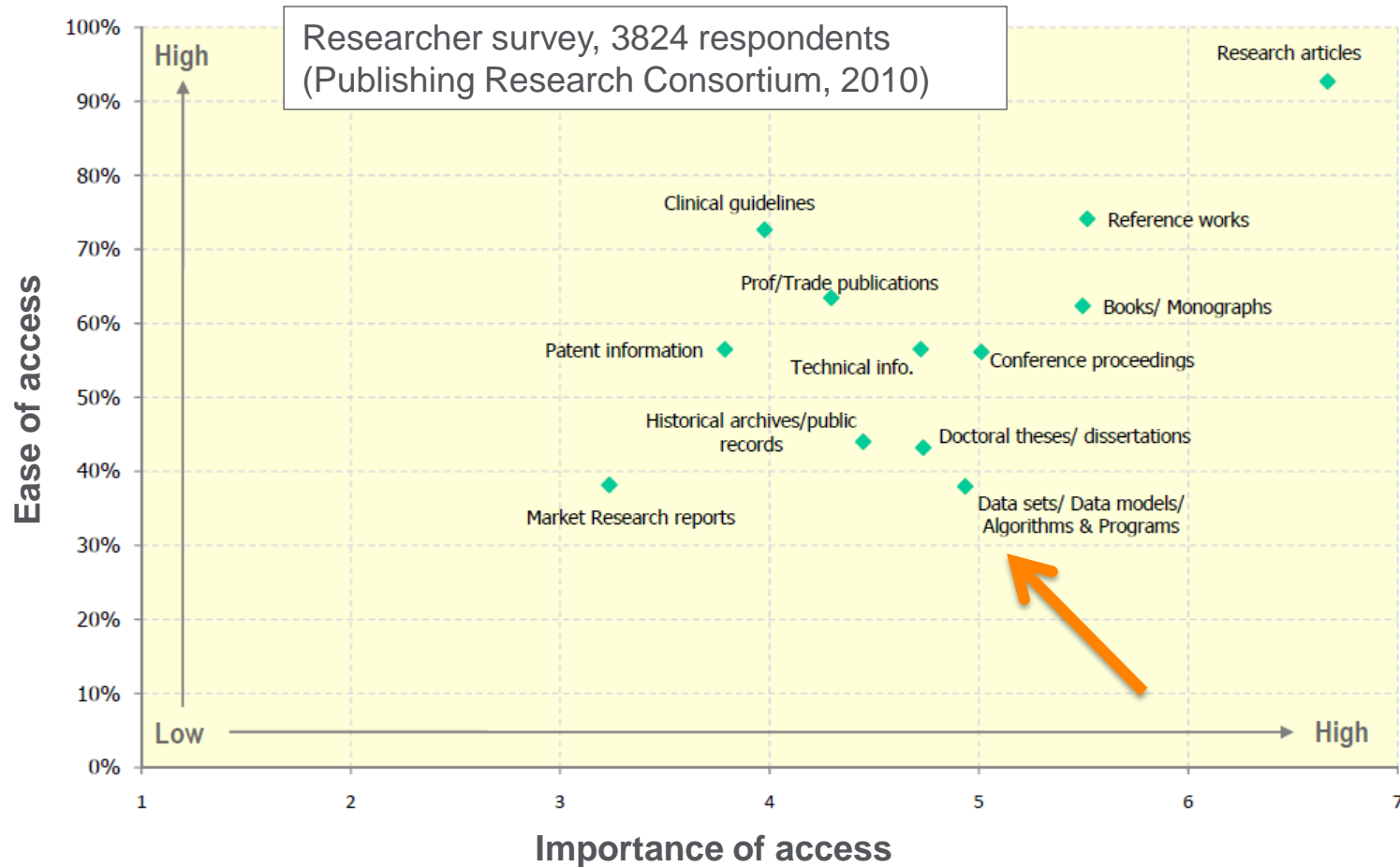
Linking data and publications – the past, present, and future



Dr. Hylke Koers, Head of Content Innovation, Elsevier

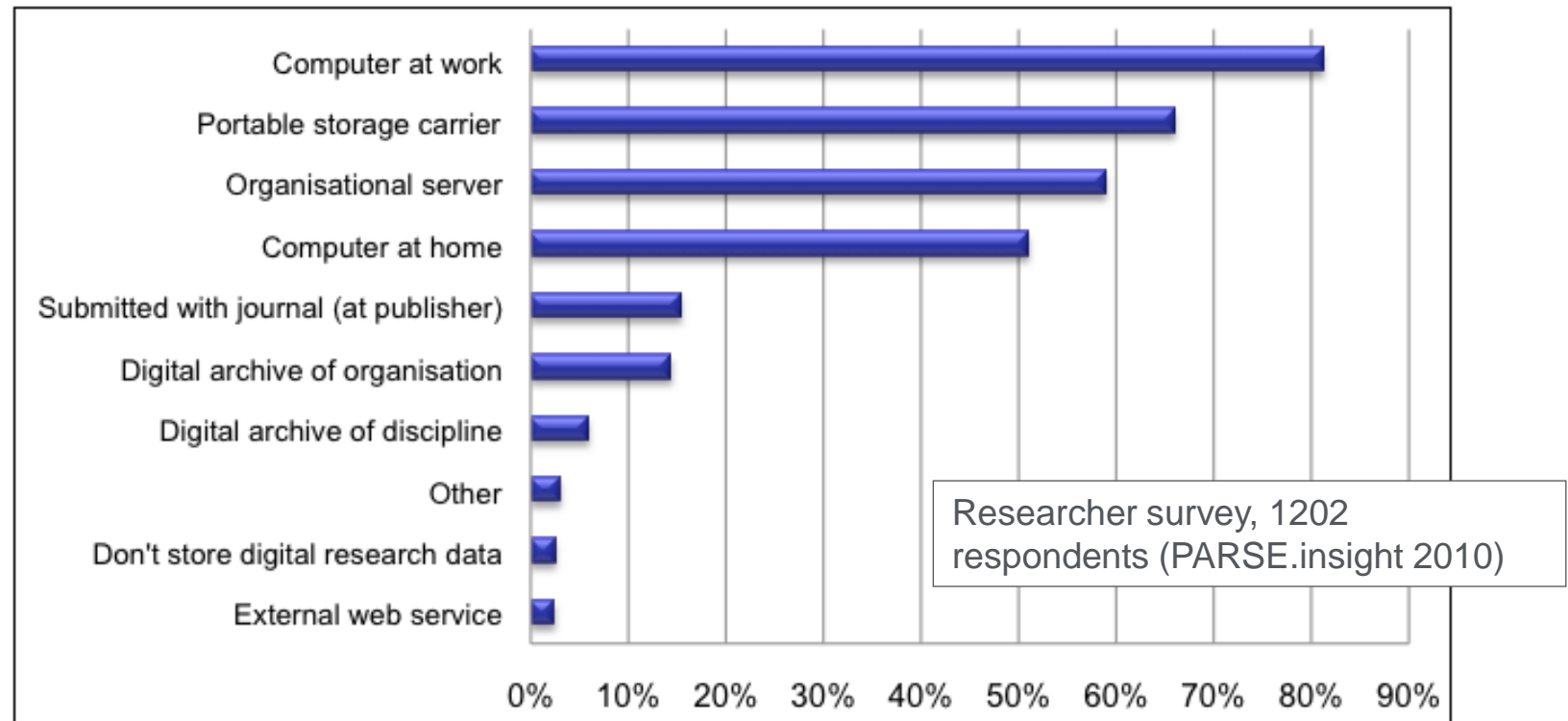
STM Innovations Seminar, 3 Dec 2014

The issue: data is important, but hard to access

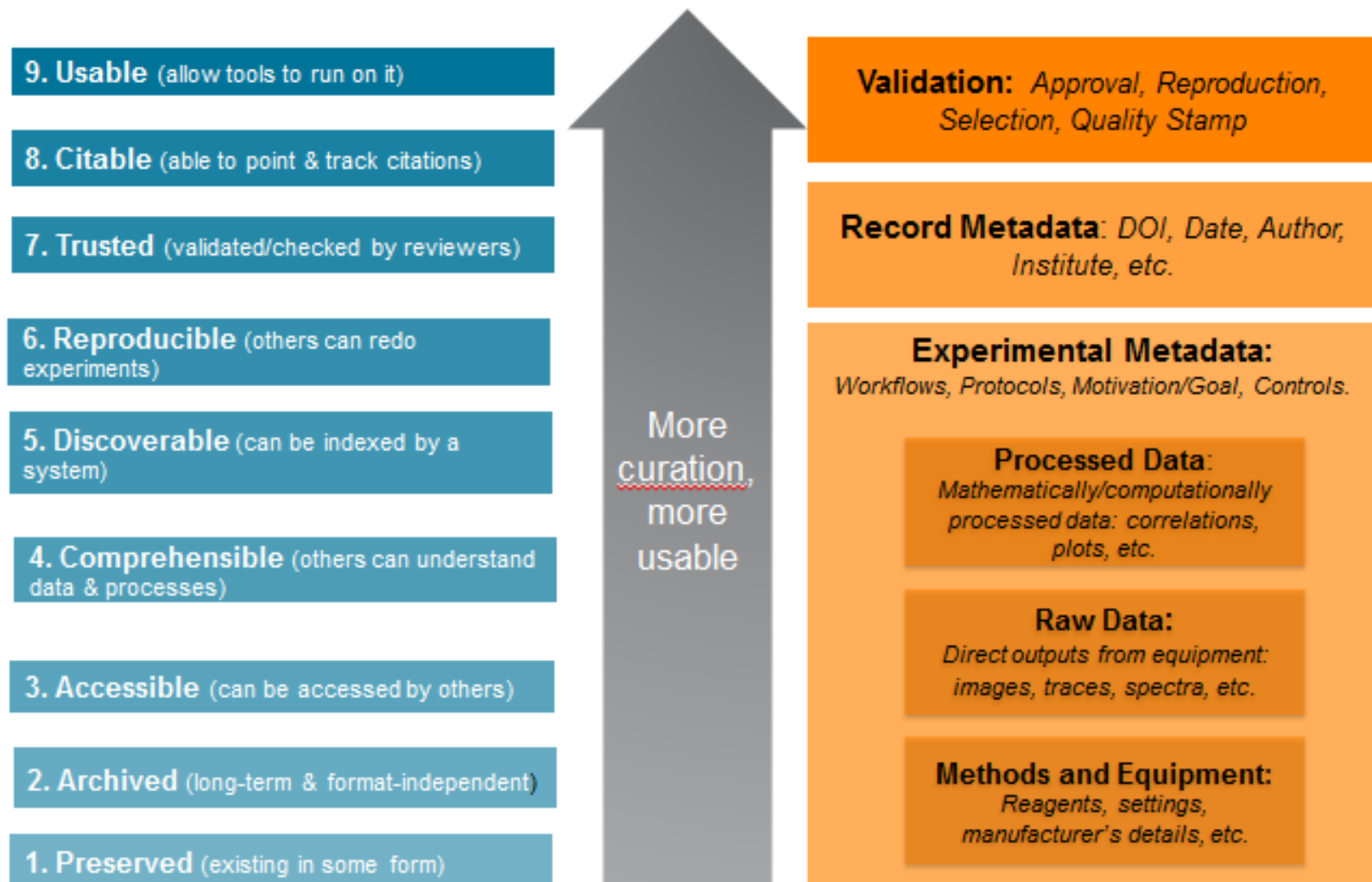


That should not be a surprise, as data is often left in the proverbial drawer.

Where do you currently store your research data? (researchers/multiple answers, N=1202)



And when you find it, you still have to make sense of it!



So... there's a lot of work ahead!



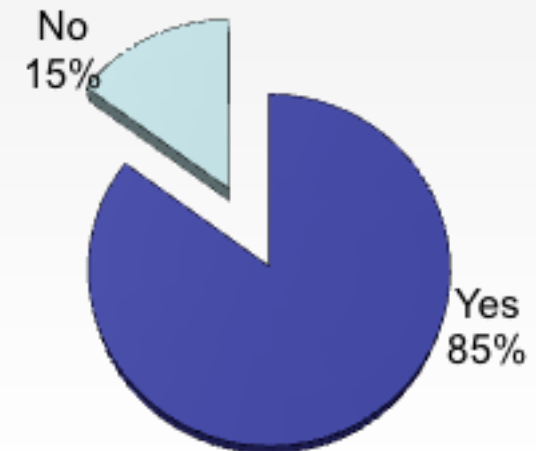
Hercules and the Hydra (ca. 1475) by Antonio del Pollaiuolo.

Incidentally, slaying a Hydra is a lot of work but a finite task – see <http://www.quora.com/What-are-some-of-the-most-counterintuitive-mathematical-results>

Linking articles and data adds value

- Increase visibility, discoverability, and usage of both articles & data
- Provide context, avoid misinterpretation and incorrect usage
- Ensure long-term availability of useful content and context
- Coordinate submission process / deposit mechanism

Question: Do you think it is useful to link underlying research data with formal literature?



Researcher survey, 1202 respondents (PARSE.insight 2010)

But it needs to be done right, and hard-coded URL's is not the way to go.

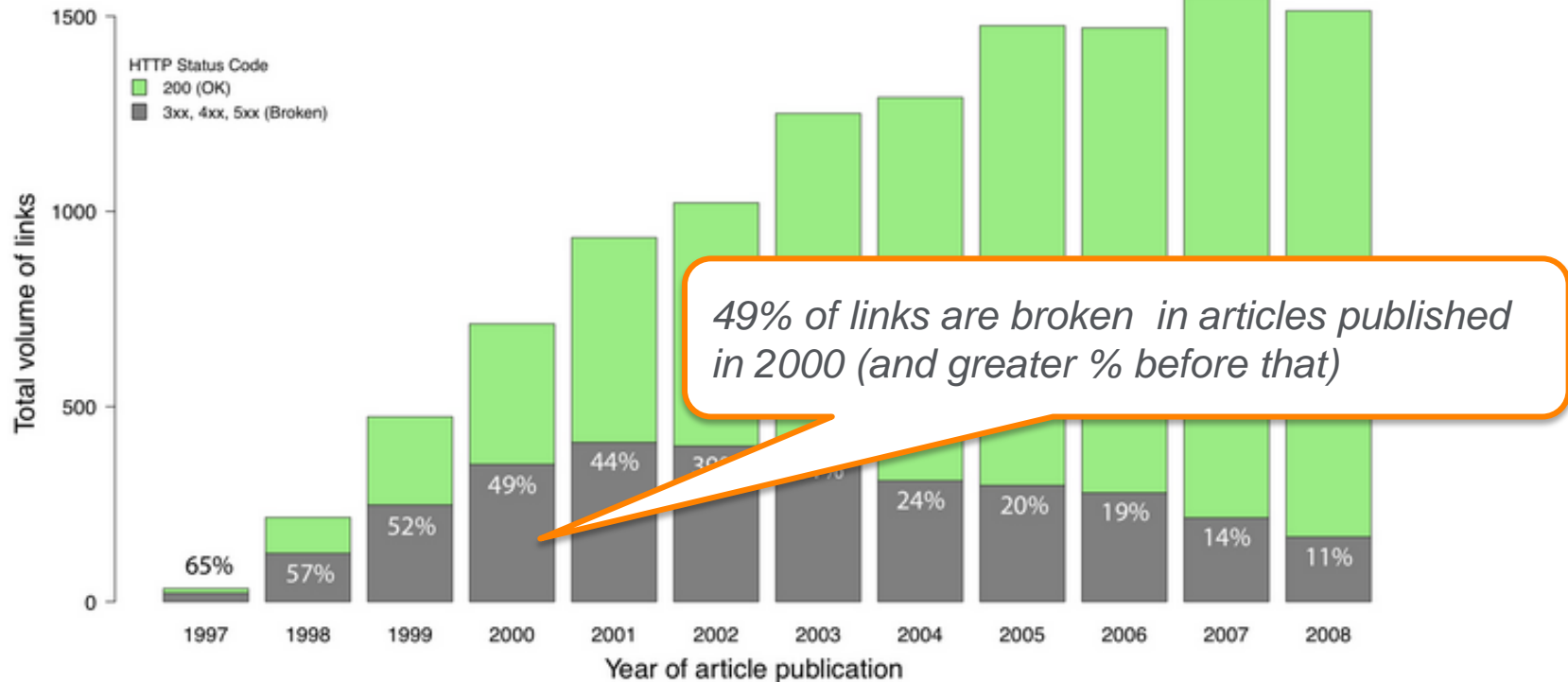


Fig. 2. Volume of potential data links in astronomy publications. Total volume of external links in all articles published between 1997 and 2008 in the four main astronomy journals, color coded by HTTP status code. Green bars represent accessible links (200), grey bars represent broken links.

From Pepe et al., “How do astronomers share data? Reliability and persistence of datasets linked in AAS publications and a qualitative study of data practices among US astronomers.”
https://authorea.com/users/3/articles/288/_show_article

Linking data and publications – ~~the past~~, present, and future







Dr. Hylke Koers, Head of Content Innovation, Elsevier

STM Innovations Seminar, 3 Dec 2014

Data-linking at Elsevier

- Extensive program to set up links between articles on ScienceDirect and relevant data repositories
 - Close to 50 databases linked, including some of the leading domain-specific data repositories
 - Links are bi-directional where possible
 - Links are always specific to data relevant for the article
 - See <http://www.elsevier.com/databaselinking>
- 
- A word cloud in the background of the slide lists various biological databases and repositories. The words are in different colors and orientations, including: Wormbase, NCBI, HepData, CCDC, EarthChem, GenBank, CRAN, GEO, ThermoML, FlyBase, SIMBAD, EMBL, MINT, SGD, OMIM, Antibody Registry, ZFIN, PDB, UniProt, IGSN, ASTM, and NCI Clinical Trials.gov.





Linking articles and data through accession numbers

PDF  Download PDF  Export  More options...  Advanced search

Article outline ☐ Show full outline

Abstract
Keywords
1. Introduction
2. Materials and methods
3. Results
4. Discussion
Acknowledgements
References

Figures and tables


 Table 1
 Table 2


below). Amplification was carried out in a 50 µl reaction mixture containing 2.5 mM MgCl₂, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 2 mM each dNTP, 0.2 µM each primer (Table 2) and 1.25 U of Fast Start Taq DNA polymerase (Roche Diagnostics, Meylan, France). After an initial denaturation step at 95 °C for 10 min, samples were amplified for 45 cycles of denaturation at 95 °C for 30 s, annealing at 60 °C for 30 s, and an extension at 72 °C for 1 min. Specific amplification products were purified and then sequenced using the DiDeoxy Terminator cycle sequencing kit v1.1 protocol (Applied Biosystems, Courtaboeuf, France). The reaction products were run on an ABI PRISM 3130 Genetic Analyzer and analyzed with the Sequence analysis/Seqscape v2.1 software (Applied Biosystems, Courtaboeuf, France).





In total, 14 Nigerien rodent-borne trypanosomes were sequenced. These sequences were included in the full dataset used by Stevens et al. (1999a), the latter matrix is the most recent, complete and well annotated one available to date for *T. brucei* (Stevens, 2008)] SSU rDNA gene. In addition, two sequences of *T. lewisi* isolated from a rat (GenBank accession no. DQ011519) in Senegal/Gambia were included (Stevens, 2006). The management of indels performed on the hand following exactly the strategy (Stevens, 2009) were all considered as 'N' thus making the sequences 2509 bp long, indels included (matrix available upon request).

Maximum parsimony and Maximum Likelihood approaches. Maximum parsimony was preferentially used under parsimony because it can be used on large datasets (Goloboff, 1999). Heuristic searches were performed with 1000 random-addition replicates, CSS, tree ratchet, tree drifting and tree fusing turned on ('new technology search' options of TNT), 100 random-addition replicates and a MaxTree's value of 1000. Gaps were treated as a fifth character in all parsimony analyses. Support of trees was assessed using

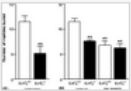
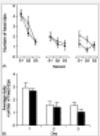
In-text links using data accession numbers – properly tagged in XML


Linking articles and data through data banners






ScienceDirectJournals BooksRemote access Sign in Help

 **Download PDF** **Export** **More options...** **Advanced search**

Article outline
Abstract
Keywords
Acknowledgment
References

Figures and tables





**Neuroscience Letters**
Volume 414, Issue 3, 13 March 2007, Pages 247–251

The 5-HT₇ receptor influences stereotypic behavior in a model of obsessive-compulsive disorder
Peter B. Hedlund   , J. Gregor Sutcliffe
 **Show more**
doi:10.1016/j.neulet.2006.12.054  **Get rights and content**

Abstract

The 5-HT₇ receptor has been suggested as a new putative target for the treatment of neuropsychiatric disorders, especially depression. This hypothesis is based on the finding that antidepressant drugs have relatively high affinity for the 5-HT₇ receptor, and that inactivation or blockade of the receptor leads to an antidepressant-like profile in behavioral models and sleep parameters. Obsessive-compulsive disorder is also believed to involve the [serotonergic](#) system and is treated using antidepressants, thus it is of interest to study the possible role of the 5-HT₇ receptor in this disorder. We have evaluated the effect of inactivation or pharmacological blockade of the 5-HT₇ receptor in three mouse behavioral models that are believed to mimic

Recommended articles
Citing articles (28)
Related book content

Data for this Article
 **Mouse Genome Informatics**
Genes, expression, function, phenotypes for mice
 **Rat Genome Database**
Genomic information on rats

Linking articles and data through data banners

ScienceDirect

Remote access Sign in Help

Search ScienceDirect Advanced search

Article content

Abstract
Keywords
Acknowledgment
References

Figures and tables

Table 1

ELSEVIER

Neuroscience Letters

Volume 423, 13 March 2007, Pages 247–251

The 5-HT₇ receptor influences serotonergic transmission in the rat nucleus accumbens

Peter B. Hedlund, J. Gregor Sutcliffe

Show more

doi:10.1016/j.neulet.2006.12.054

Abstract

The 5-HT₇ receptor has been suggested as a target for the treatment of mood disorders, especially depression. This hypothesis is based on the relatively high affinity for the 5-HT₇ receptor, and antidepressant-like profile in behavioral models are believed to involve the serotonergic system and is the possible role of the 5-HT₇ receptor in this pharmacological blockade of the 5-HT₇ receptor.

Data for this Article

Mouse Genome Informatics
Genes, expression, function, phenotypes for mice

Rat Genome Database
Genomic information on rats

Links to relevant data sets added after publication, curated by data repositories

And beyond linking, there's data integration and visualization



Marine Geology

Volume 204, Issues 1–2, 28 February 2004, Pages 43–57



Calcium carbonate corrosiveness in the South Atlantic during the Last Glacial Maximum as inferred from changes in the preservation of water circulation

Authors have uploaded data to PANGAEA, submitted article for publication to Marine Geology journal

[Get rights and content](#)

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 reconstruct the evolution of the recent NADW/AABW transition, reflected by the calcite lysocline

► Recommended articles

► Citing articles (17)

► Related reference work articles



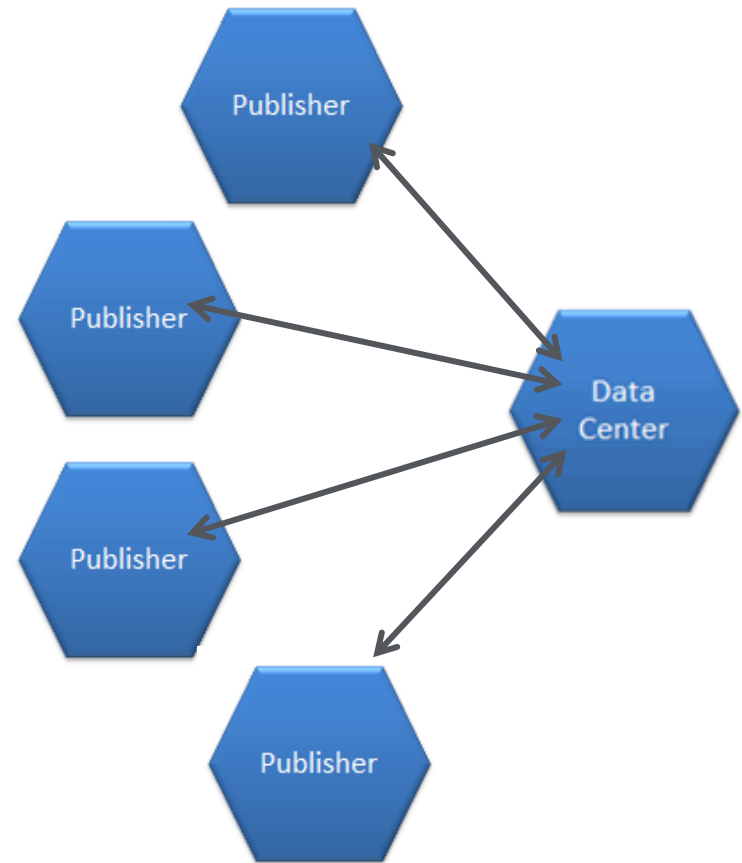
PANGAEA® – Related Data

Dissolution index of *Globigerina bulloides* in recent and Last Glacial Maximum sediments



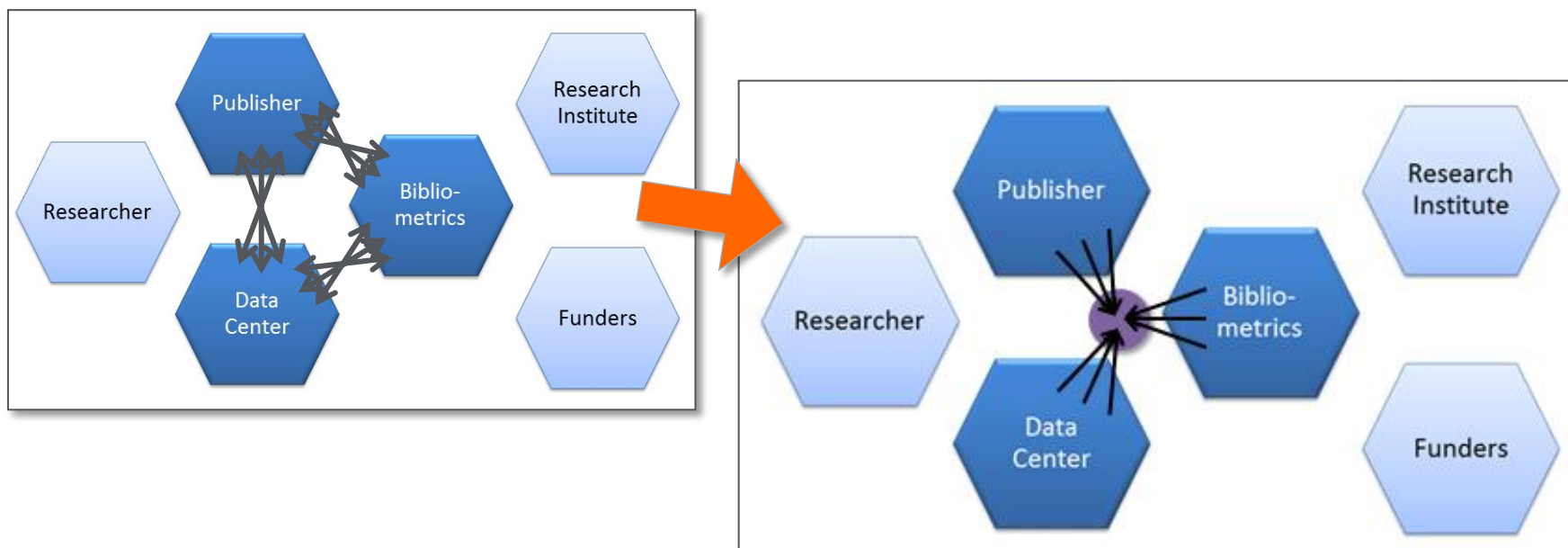
Data visualization tool connects articles and data – pulling in data from PANGAEA for this article and showing to the reader

Linking articles and data is great, but all those bilateral relations don't scale



Introducing the joint ICSU-WDS / RDA Working Group “Data Publication Services”

The challenge in today's data publishing landscape: how do we move from a plethora of bilateral arrangements to a **one-for-all service model** ?



- Increase interoperability
- Decrease systemic inefficiencies
- Power new tools and functionalities to the benefit of researchers

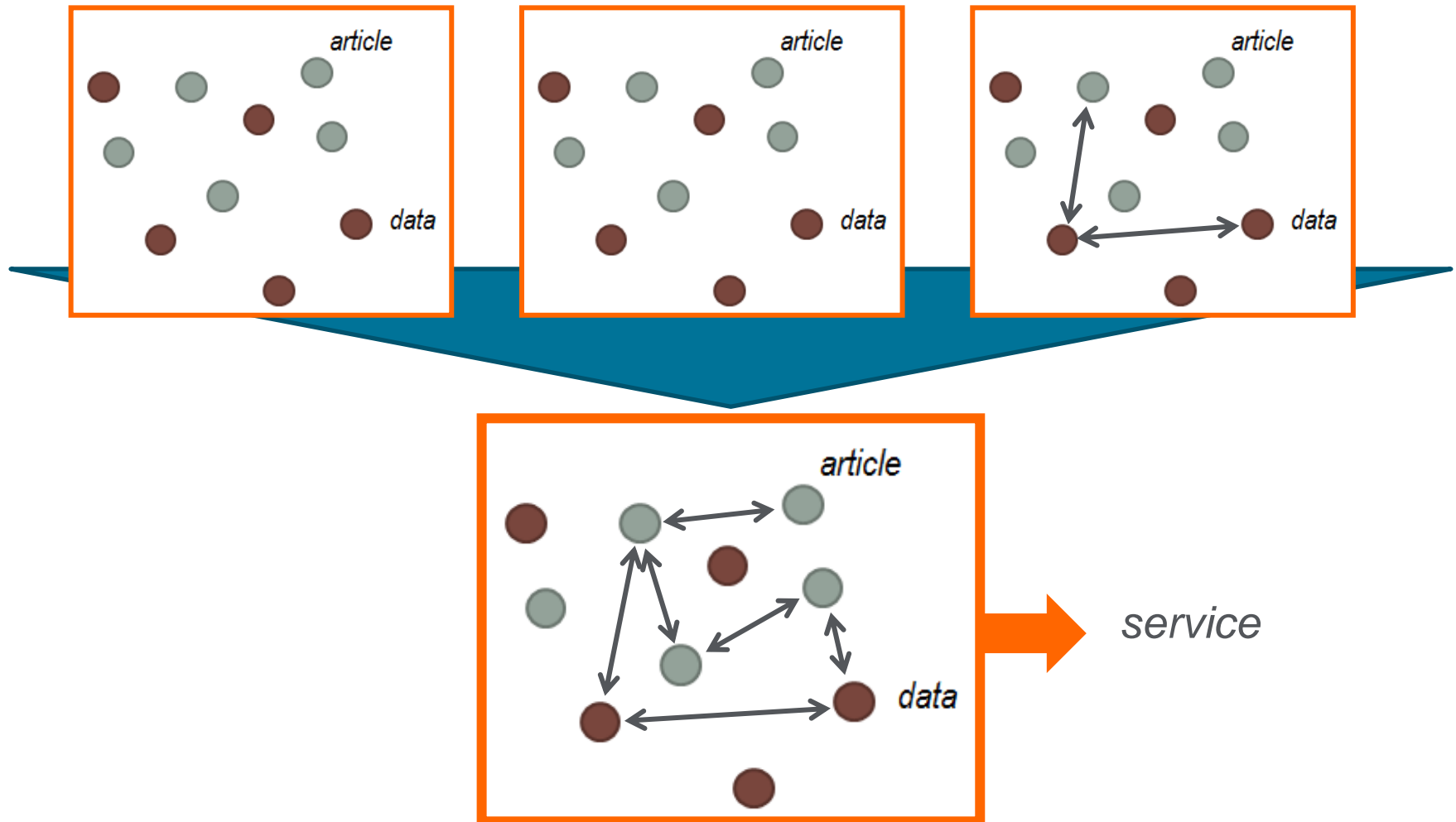
The WG aims to develop a universal article-data cross-referencing service (and then some)

- Address processes, workflows, and solutions that currently exist between individual parties within the data publication landscape, and investigate how these can be lifted to one-for-all service to:
- Primary Focus: universal article – data cross-referencing service
 - Given article A, what relevant data D exists – and vice versa
 - Additional metadata about the nature of the relationship, e.g. supplementary data, related data, etc.
 - Additional metadata for article and/or data set

See full case statement at

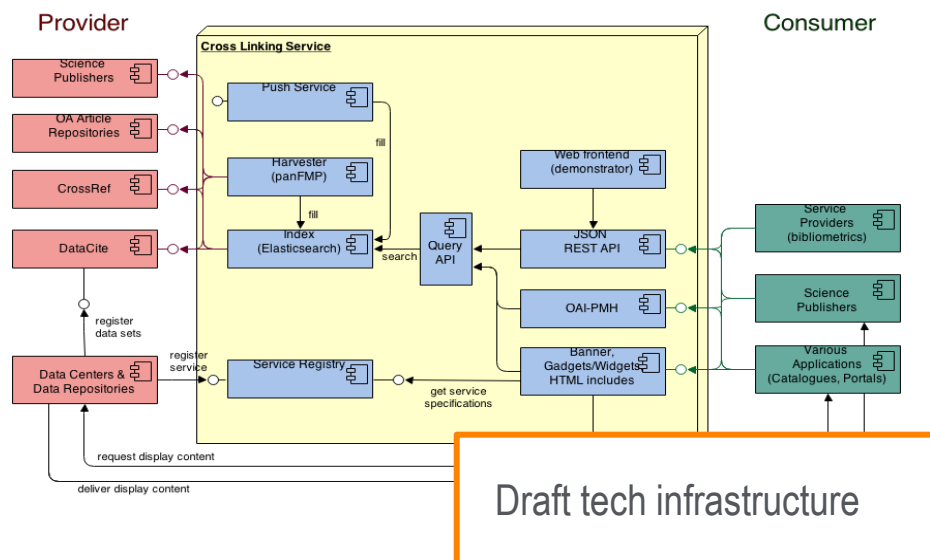
<https://rd-alliance.org/internal-groups/rdawds-publishing-data-ig.html>

It's all about connecting the dots



With links contributed by data repositories, publishers, infrastructure provides, and other key players

- *3TU .Datacentrum*
- *Australian National Data Service (ANDS)*
- *Cambridge Crystallographic Data Center (CCDC)*
- *CrossRef*
- *DataCite*
- *Elsevier*
- *Europe PubMed Central*
- *OpenAire*
- *PANGAEA*
- *Thomson Reuters*
-



Working group status

- Build a wide consortium with different stakeholder groups
- Guiding principles ratified
- Accumulating test corpus – ongoing
- Draft tech infrastructure – now in planning & resourcing phase
- Outreach, webinars, contact with other groups – ongoing

Next milestone: Research Data Plenary 5 in San Diego (March 2015)

Linking data and publications – the past, present, and future

