

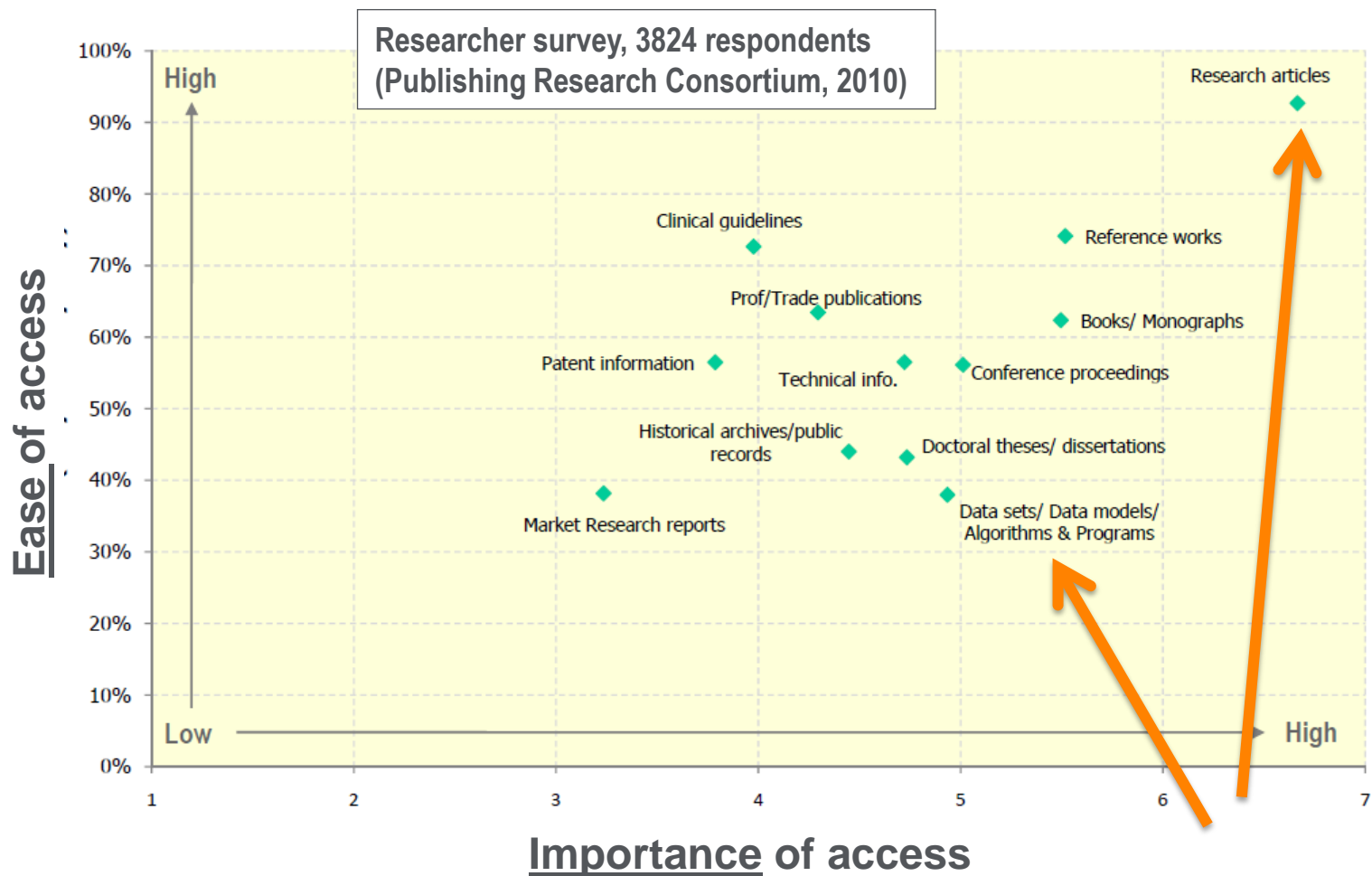
Supporting researchers to store, share, discover and use data



Dr. Hylke Koers, Head of Content Innovation, Elsevier

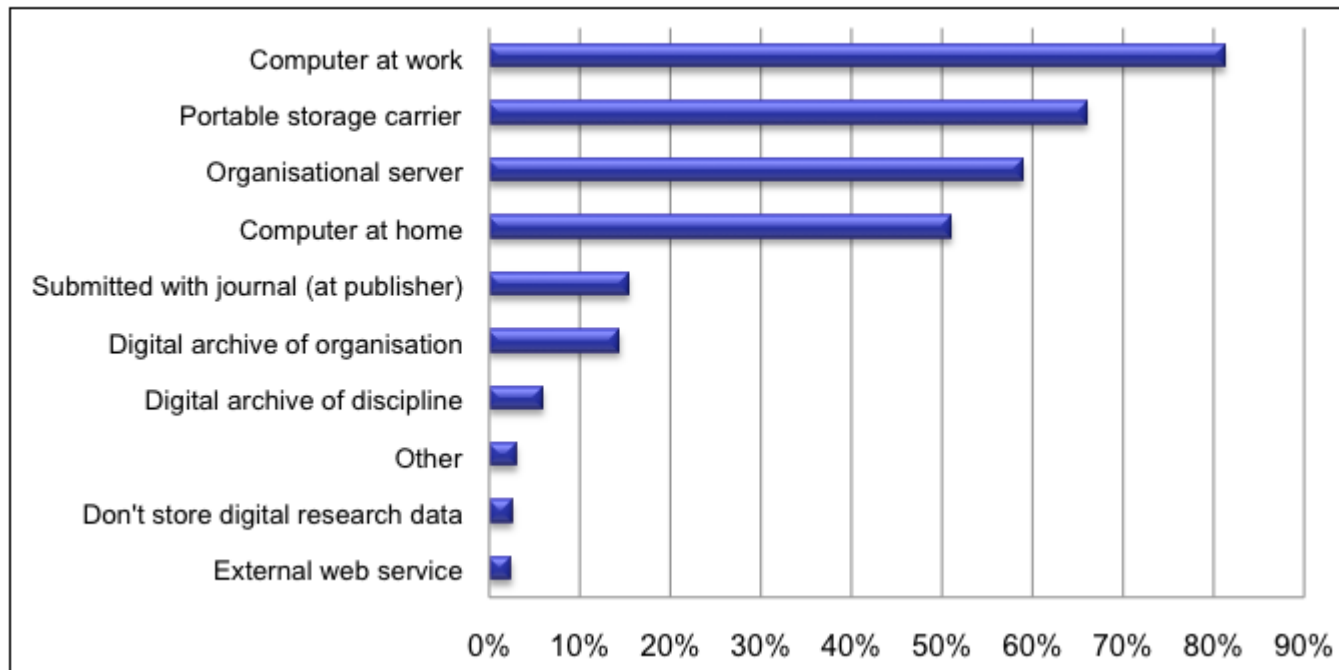
RECODE Conference, Athens, Jan 15, **“Good practices in enabling the re-use of research data”**

Data is important, but hard to access



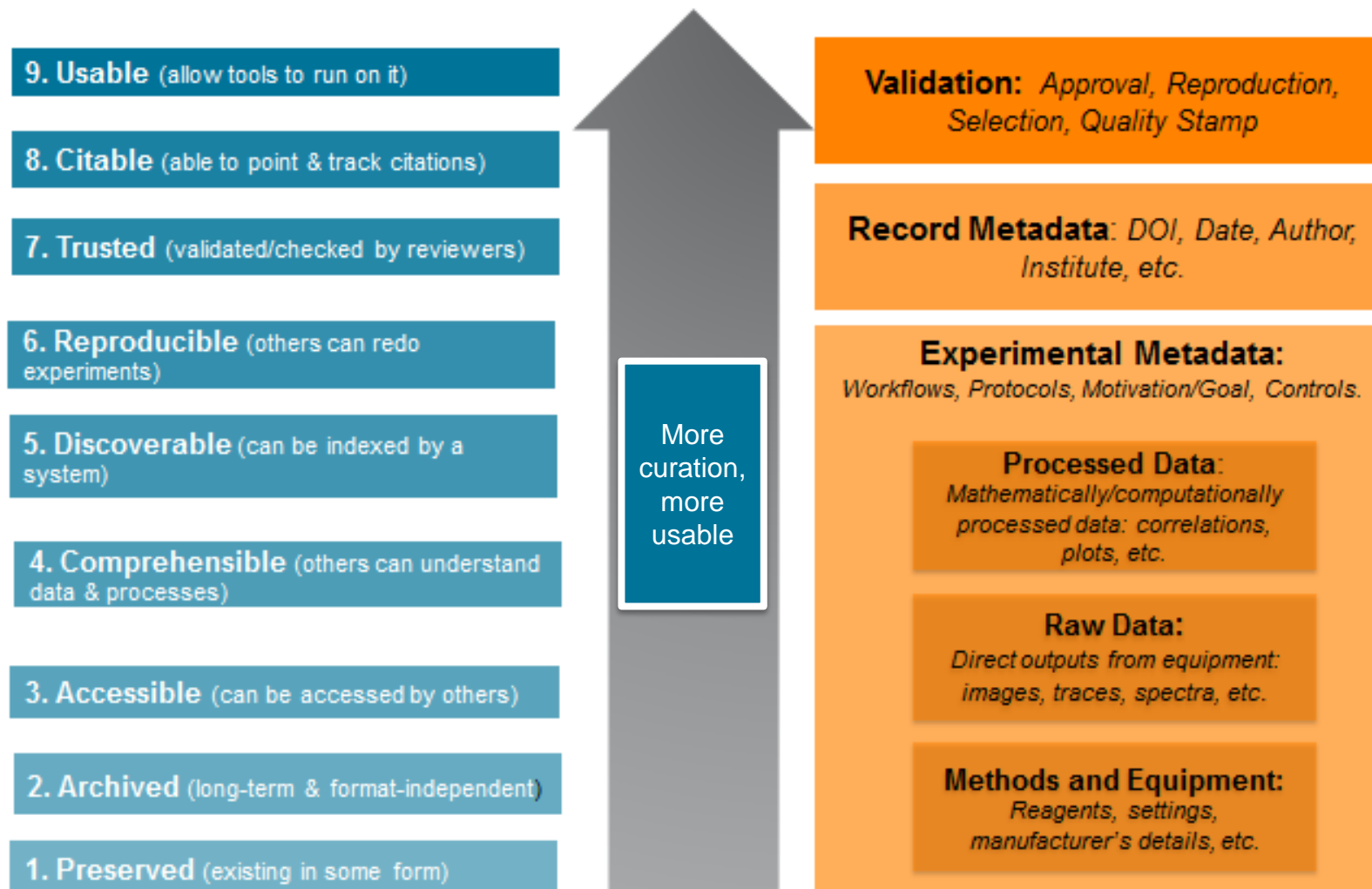
Data storage is very fragmented

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



Researcher survey, 1202 respondents
(PARSE.insight 2010)

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



How is Elsevier supporting researchers to store, share, discover, and use research data?

In a lot of ways, including:

- Linking articles and data
- Embracing data in the Article of the Future
 - In-article data visualization
 - Open Data pilot
- Data journals and microarticles
- Community efforts and standards
 - Data Citation principles
 - RDA / ICSU-WDS working groups

More at: <http://www.elsevier.com/about/research-data>

The 2007 STM Brussels Declaration

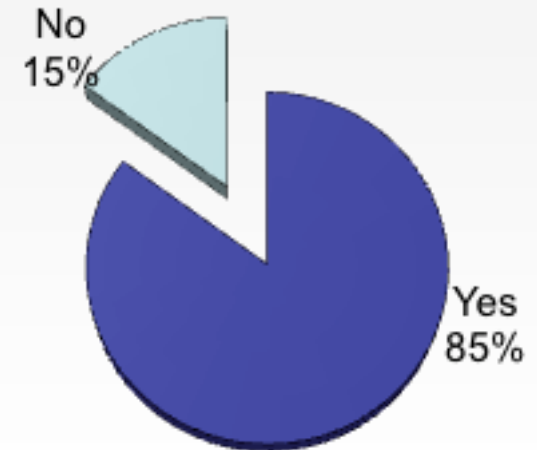
“Raw research data should be made freely available to all researchers. Publishers encourage the public posting of the raw data outputs of research. Sets or sub-sets of data that are submitted with a paper to a journal should wherever possible be made freely accessible to other scholars”

Undersigned by Elsevier and many other publishers

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.

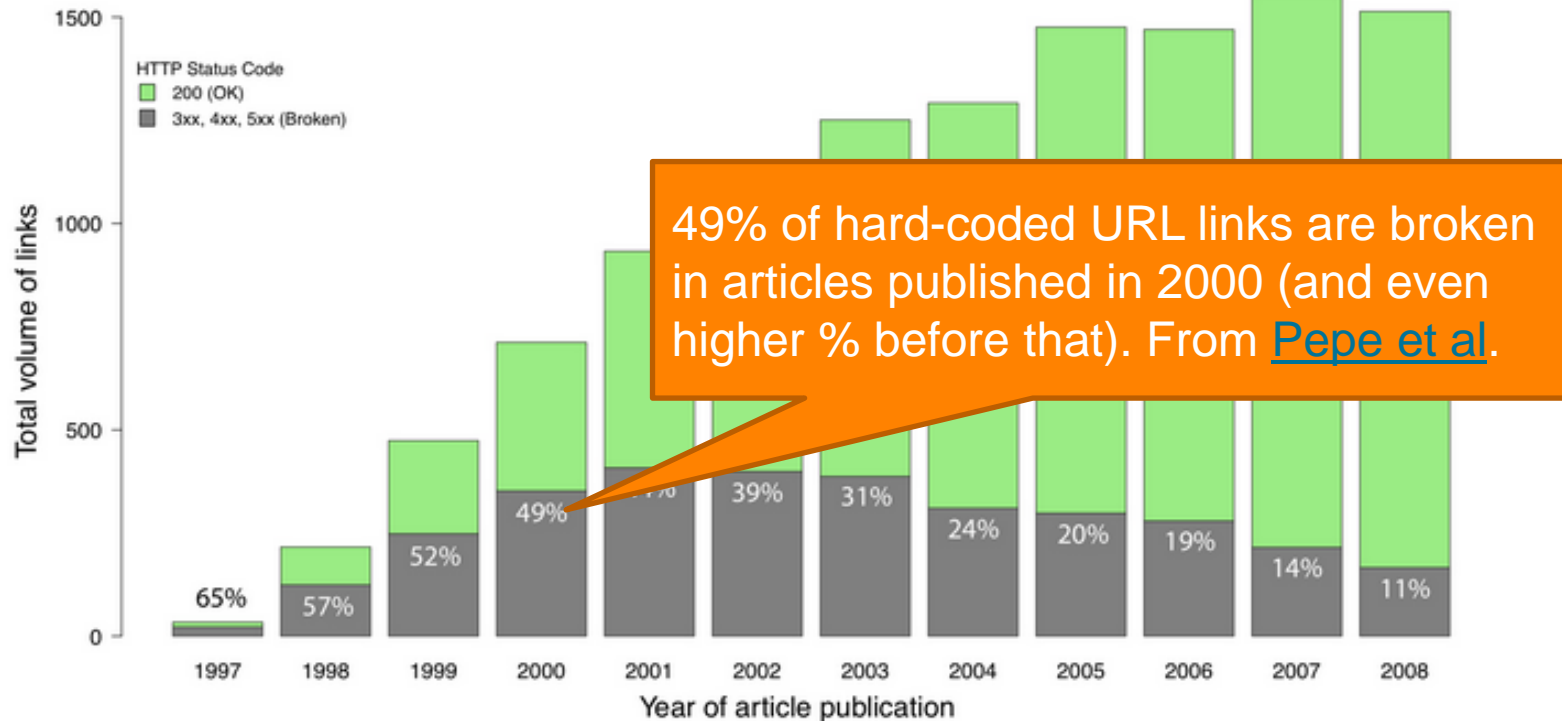
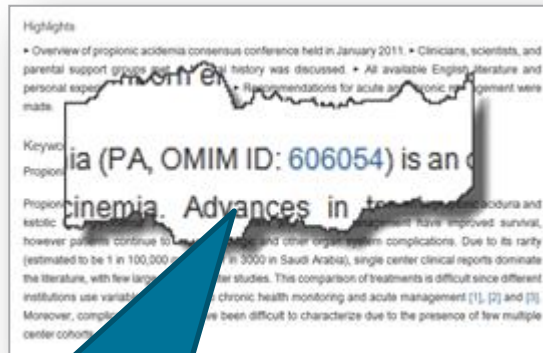


Fig. 2. Volume of potential data links in astronomy publications. Total volume of external links in all articles published between 1997 and 2008 in the four main astronomy journals, color coded by HTTP status code. Green bars represent accessible links (200), grey bars represent broken links.

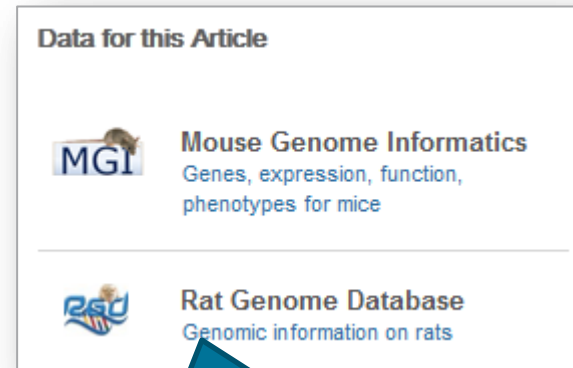
From Pepe et al., *“How do astronomers share data? Reliability and persistence of datasets linked in AAS publications and a qualitative study of data practices among US astronomers.”*
https://authorea.com/users/3/articles/288/show_article

Data-linking at Elsevier

- Elsevier has an extensive program with 40+ 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

See <http://www.elsevier.com/databaselinking>

And beyond linking, there's data integration and visualization



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

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

[Get rights and content](#)

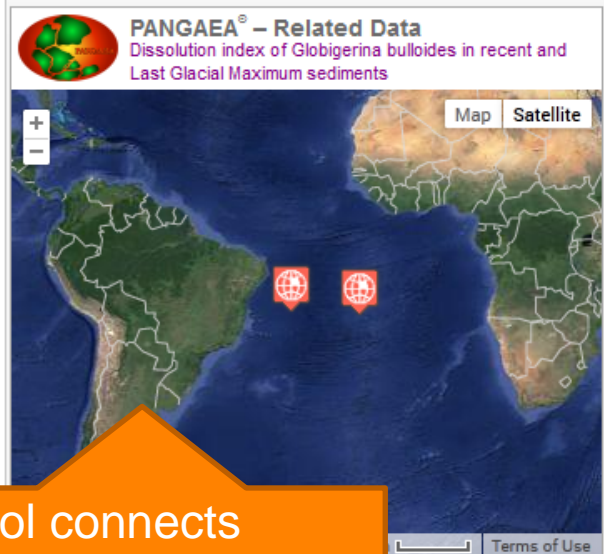
Abstract

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 relative importance of NADW/AABW transition, reflected by the calcite lysocline, was

► Recommended articles

► Citing articles (17)

► Related reference work articles



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

How is Elsevier supporting researchers to store, share, discover, and use research data?

In a lot of ways, including:

- Linking articles and data
- **Embracing data in the Article of the Future**
 - In-article data visualization
 - Open Data pilot
- Data journals and microarticles
- Community efforts and standards
 - Data Citation principles
 - RDA / ICSU-WDS working groups

More at: <http://www.elsevier.com/about/research-data>

Support for (supplementary) data tables

 Download PDF
  Export
  More options...

 Advanced search

Article outline Show full outline

Abstract

Keywords

1. Introduction

2. Iron Age Megiddo

3. Materials and methods

4. Results

5. Discussion

Acknowledgments

References

Figures and tables



A total of 1080 identifiable specimens were recovered (see Table S1). This material includes remains from reptiles (Squamata, mainly snakes and lizards) which shows that numbers of specimens are generally dispersed rather than concentration of the remains. Equation (4.25) shows values of 0.09 in Level Q, indicating an aggregated or uniform distribution of the remains among the

Supplementary data tables offered inline at the point of reference

Inline Supplementary Table S1

Table S1.

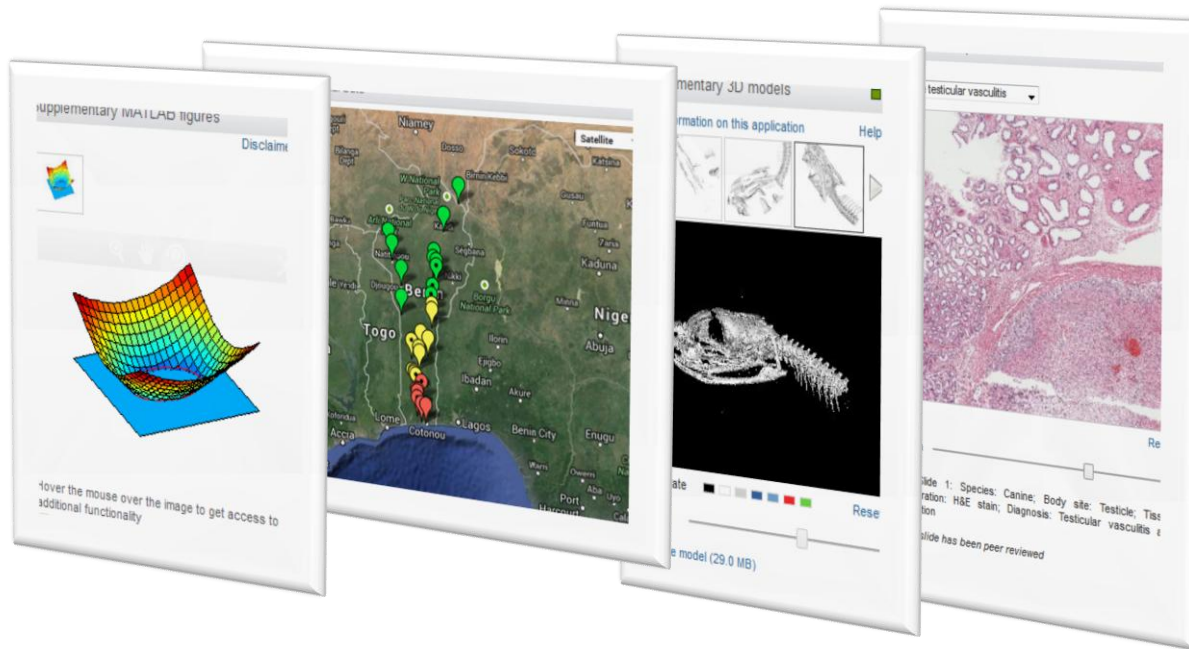
Counts of specimens by taxonomic categories.

| Locus | Basket/level | Bird | Fish | Reptile | Mammal | Rodent | Gerbil | Jird | Field vole | Common mouse | Mole rat | Shrew | Grand total |
|-------|--------------|------|------|---------|--------|--------|--------|------|------------|--------------|----------|-------|-------------|
| 106 | 3 | 1 | 27 | 1 | – | 7 | – | – | – | 6 | – | 1 | 43 |
| 112 | 1 | – | 5 | 2 | – | 5 | – | – | – | 2 | – | – | 14 |
| 157 | 1 | – | 49 | 13 | 2 | 16 | – | 1 | – | 12 | 1 | – | 94 |
| | 2 | 2 | 80 | 17 | – | 25 | – | 1 | – | 15 | – | – | 140 |
| 33 | 15 | – | 7 | 4 | 1 | 8 | – | – | – | 18 | – | – | 38 |
| 46 | 16 | 2 | 3 | – | – | 2 | – | – | – | – | – | – | 7 |
| | 17 | 13 | 21 | 2 | 12 | 13 | – | – | – | 21 | – | 3 | 85 |
| | | 6 | 19 | 2 | – | 8 | – | – | – | 7 | – | 1 | 43 |
| | | 1 | 17 | 2 | – | 14 | – | – | – | 18 | – | – | 52 |

All tables on ScienceDirect can be downloaded as .CSV file

In-article data viewers to allow data exploration with the article

- Extend the format of the article to also capture data (as well as computer code, multimedia files, etc.)
- Offer authors the best possible way to present their work
- Enabling other researchers explore and engage with rich articles



And much more!
See [Content Innovation on Elsevier.com](#)

Interactive plots... Google Maps... 3D models... Virtual Microscope...

Example: 3D neuroimaging data viewer



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

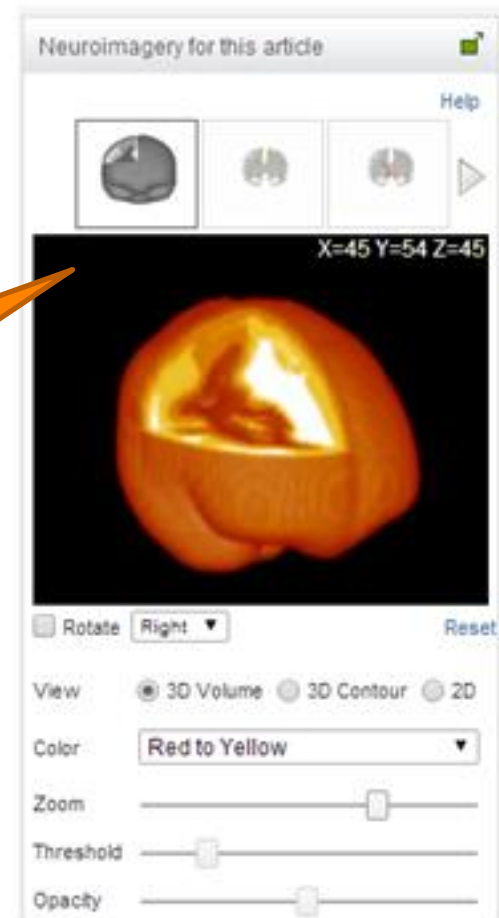
Oscar Bartra¹, Joseph T. N.

<http://dx.doi.org/10.1016/j.neur>

Highlights

- Theories of decision making posit a domain-general *subjective value* signal.
- Numerous fMRI studies have tested for correlates of such a signal.
- We quantitatively meta-analyze findings from 206 studies.
- We identify a "valuation system" in ventromedial PFC and ventral striatum.
- Other regions show a mixture of positive and negative effects across experiments.

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



Example: Interactive (Google) Maps viewer for geospatial data

Article outline Show full outline

Abstract
Keywords

1. Introduction
2. Materials and methods
3. Results
4. Discussion

Acknowledgements
Appendix A. Supplementary data
References

Figures and tables

Table 1




Table 2

upi0010

Factors that influence the distribution of these malaria vectors are discussed. This study underlines the need of further investigations of biological, ecological, and behavioral traits of these species and forms to better appreciate their vectorial capacities. Acquisition of entomological field data appears essential to better estimate the stratification of malaria risk and help improve malaria vector control interventions.

Supplementary Geospatial Data

Map data ©2014 Google Imagery ©2014 TerraMetrics | Terms of Use

In-article data visualization tools for geospatial data (KML / KMZ files)

Open Data

- Pilot project to test implementation and get feedback on researchers' attitudes towards Open Data in live setting
- Authors can submit raw research data with their article → will be made publicly available with their paper on ScienceDirect.

The screenshot shows the article page for "Biological Conservation", Volume 182, February 2015, Pages 177–186. The article title is "Habitat disturbance effects on the physiological stress response in resident Kenyan white-bearded wildebeest (*Connochaetes taurinus*)". The authors listed are J.A. Stabach^a, R.B. Boone^a, J.S. Worden^b, and G. Florant^c. A callout box points to the "Open Data within this article" section, which states: "This article contains open data for download under the CC BY license." Below this, there is a link to "Supplementary data 4" with the text: "This is open data under the CC BY license (http://creativecommons.org/licenses/by/4.0/). Download file (19 KB)".

This supplementary data set is publicly available as Open Data (CC BY) on ScienceDirect

How is Elsevier supporting researchers to store, share, discover, and use research data?

In a lot of ways, including:

- Linking articles and data
- Embracing data in the Article of the Future
 - In-article data visualization
 - Open Data pilot
- **Data journals and microarticles**
- Community efforts and standards
 - Data Citation principles
 - RDA / ICSU-WDS working groups

More at: <http://www.elsevier.com/about/research-data>

Microarticles as a vehicle to disseminate data

The image shows a screenshot of a Genomics Data article page. The page title is "Genomics Data" and the volume is "Volume 1, December 2013, Pages 7–10". The article title is "Genome-wide gene expression profiling to predict resistance to anthracyclines in breast cancer patients". The authors listed are B. Haibe-Kains, C. Desmedt, A. Di Leo, E. Azambuja, D. Larsimont, J. Selleslags, S. Delaloge, C. Duhem, J.P. Kains, B. Carly, M. Maerevoet, A. Vindevoghel, G. Rouas, F. Lallemand, V. Durbecq, F. Cardoso, R. Salgado, R. Rovere, G. Bontempi, S. Michiels, M. Buyse, J.M. Nogaret, Y. Qi, F. Symmans, L. Pusztai, V. D'Hondt, M. Piccart-Gebhart, and C. Sotiriou. A callout box points to a link to primary data stored at domain-specific data repositories. Another callout box points to the "Data in Brief" section, stating that "Data in Brief" articles are a thorough description of data, making datasets truly accessible and reproducible. A sidebar on the right shows "GEO data referred to in this article" with a table of Platform (0), Sample (2), and Series (1). The selected GEO entity is GSM411367, with sample GSM411367, status public on Jan 26, 2010, title XXX_9677, source name XXX_9677, and organism Homo sapiens.

Genomics Data
Volume 1, December 2013, Pages 7–10

ELSEVIER

Recommended articles

Links to primary data stored at domain-specific data repositories

Data in Brief

Genome-wide gene expression profiling to predict resistance to anthracyclines in breast cancer patients ☆

B. Haibe-Kains^{a, c, m}, C. Desmedt^a, A. Di Leo^b, E. Azambuja^a, D. Larsimont^a, J. Selleslags^a, S. Delaloge^d, C. Duhem^e, J.P. Kains^f, B. Carly^g, M. Maerevoet^h, A. Vindevoghelⁱ, G. Rouas^a, F. Lallemand^a, V. Durbecq^a, F. Cardoso^a, R. Salgado^a, R. Rovere^a, G. Bontempi^c, S. Michiels^a, M. Buyse^j, J.M. Nogaret^a, Y. Qi^a, F. Symmans^k, L. Pusztai^l, V. D'Hondt^a, M. Piccart-Gebhart^{a, 1}, C. Sotiriou^a

Show more

<http://dx.doi.org/10.1016/j.gdata.2013.09.001>

Abstract

Validated biomark

GEO data referred to in this article

| Platform (0) | Sample (2) | Series (1) |
|--------------------------------|------------|------------|
| GEO entity: GSM411367 | | |
| Sample GSM411367 | | |
| Status: public on Jan 26, 2010 | | |
| Title: XXX_9677 | | |
| Source Name: XXX_9677 | | |
| Organism: Homo sapiens | | |

GEO
Gene Expression Omnibus

"Data in Brief" articles are a thorough description of data, making datasets truly accessible and reproducible.

How is Elsevier supporting researchers to store, share, discover, and use research data?

In a lot of ways, including:

- Linking articles and data
- Embracing data in the Article of the Future
 - In-article data visualization
 - Open Data pilot
- Data journals and microarticles
- **Community efforts and standards**
 - Data Citation principles
 - RDA / ICSU-WDS working groups

More at: <http://www.elsevier.com/about/research-data>

Research Data Working Groups and Industry Standards

- **Joint Declaration of Data Citation Principles:** best-practices to cite data in articles for better linking and credit
- **Research Data Alliance & ICSU World Data System:** Tackling a broad range of interconnected issues around Data Publication (workflows, bibliometrics, cost recovery, services)



Joint Declaration of Data Citation Principles

- Systematic way to link articles and data using persistent IDs
- Mechanism to give credit & attribution for data
- Help to search for data, and discover from articles
- Recent community effort: Joint Declaration of Data Citation Principles

Data set cited in reference list – treated on similar footing with articles

under their Embryo 2011 programme in collaboration with the York and York (see Hankinson et al., in press, Coplestone et al., in press, Howard et al., 2013 and International Commission on Radiological Protection, (ICRP), 2009). The inputs of N.A. Beresford and M. Muikku to this work were funded under the EURATOM STAR network of excellence in radioecology (www.star-radioecology.org).

References

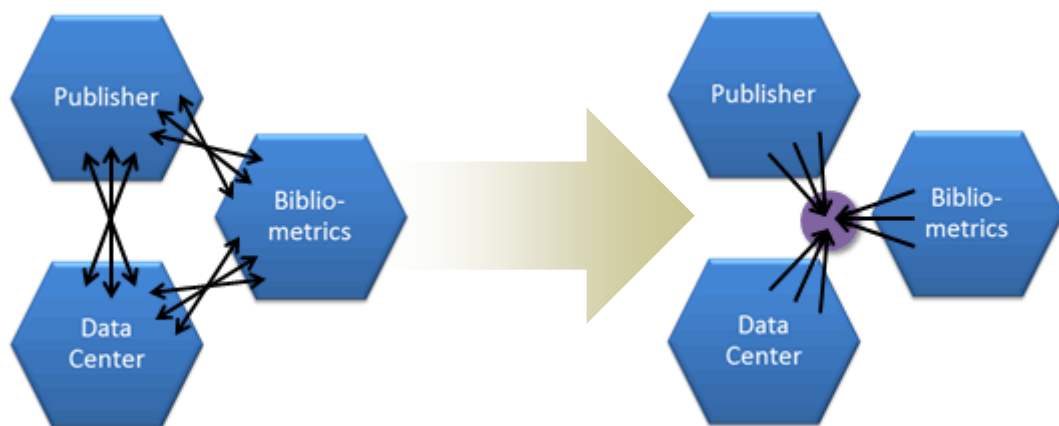
Barnett et al., 2013 C.L. Barnett, N.A. Beresford, L.A. Walker, M. Baxter, C. Wells, D. Coplestone
Element and radionuclide concentrations in representative species of the ICRP's reference animals and plants and associated soils from a forest in North-west England
NERC — Environmental Information Data Centre (2013) <http://dx.doi.org/10.5285/e40b53d4-6699-4557-bd55-10d196ece9ea>

Beresford, 2010 N.A. Beresford
The transfer of radionuclides to wildlife (Editorial)
Radiat Environ Biophys, 49 (2010), pp. 505–508

<http://www.sciencedirect.com/science/article/pii/S0048969713006657>

Data Community Activity: ICSU-WDS/RDA Data Publication Services WG

- Co-chaired by Adrian Burton (ANDS) and Hylke Koers (Elsevier)
- Key deliverable: a common, open and freely accessible article-data cross-referencing service
- Benefits:
 - Efficiency, scalability
 - Better quality & accuracy of links
 - Powering new tools and functionalities to the benefit of researchers



Moving from a system of bilateral arrangements to a standard, central service

Thank you!

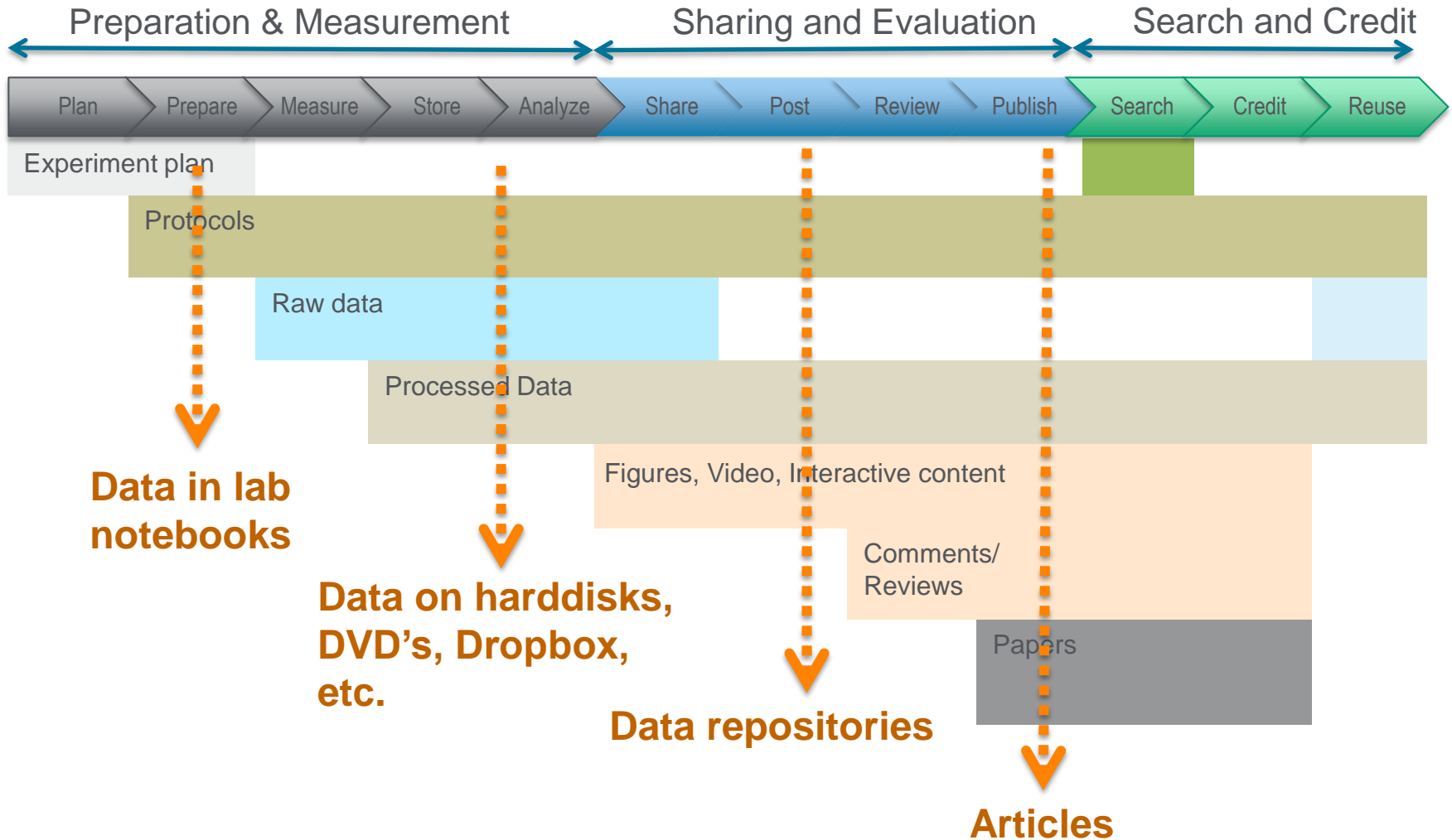
h.koers@elsevier.com

<http://www.elsevier.com/about/research-data>

<http://www.elsevier.com/about/policies/research-data>

<http://www.elsevier.com/about/content-innovation>

Lots of data in the researcher workflow



Another example in-article data viewer: Interactive plot viewer



- 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 program

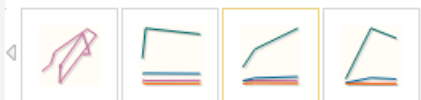
All catalysts were characterized by nitrogen adsorption-desorption isotherm, X-ray diffraction (XRD), temperature programed reduction (TPR), and methanol synthesis activity. The catalyst activity increased with aging time followed by a decrease upon further aging. At a constant aging time of 5 h, the 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_{MeOH}/kg_{cat} 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.

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_{MeOH}/kg_{cat} 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

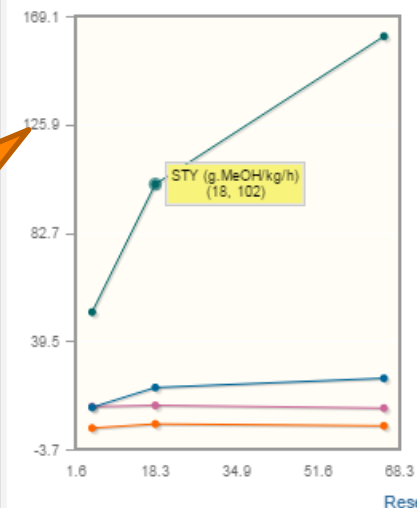
Displaying plot data in CSV format delivered by the author as supplementary material. Allows to access, explore, and download data behind plots.

Interactive plots for this article



Plot

Data table



X axis Aging Time (h)

- Y axis
- CuO crystallite size (nm)
 - Cu surface area (m²/gr)
 - Particle size-D0.5 (micro-m)
 - STY (g.MeOH/kg/h)

Another example of data-linking: Protein Viewer (with PDB)

The image shows a screenshot of an Elsevier article page. At the top, there are navigation options: 'Download PDF' (with an Adobe PDF icon), 'Export', 'More options...', and a search bar for 'ScienceDirect' with an 'Advanced search' button. The article title is partially visible as '...es by ...erase'. The volume information is 'Volume 400, Issue 3, 16 July 2010, Pages 295–308'. The Elsevier logo is on the left. A large orange callout box on the left contains the text: 'Authors include Protein Data Bank accession numbers in their article'. Below the title, the DOI is 'doi:10.1016/j.jmb.2010.05.030' and there is a 'Get rights and content' link. The 'Abstract' section begins with: 'Structure-based protein sequence alignments of family B DNA polymerases revealed a conserved motif that is formed from interacting residues between loops from the N-terminal and palm domains and between the N-terminal loop and a conserved proline residue. The importance of the motif for function of the bacteriophage T4 DNA polymerase was revealed by suppressor analysis. T4 DNA polymerases that form weak replicating complexes cannot replicate DNA when the dGTP pool is reduced. The conditional lethality provides the means to identify amino acid substitutions that restore replication activity under low-dGTP conditions either by correcting the defect produced by the first amino acid substitution or by generally increasing the stability of polymerase complexes; the ... effectively counter the reduced stability caused by a var ... substitutions that increase the stability of polymerase co ... the antiviral drug phosphonoacetic acid. Amino acid sub ... DNA under low-dGTP conditions or drug sensitivity were ...'. On the right side, there are sections for 'Citing articles (5)', 'Related book content', and 'Proteins in this article'. The 'Proteins in this article' section shows two protein structures: '1Q9X' and '1IG9'. Below these is a large 3D protein structure viewer for '1Q9X' with a 'Jmol' logo and 'Rotate' and 'Reset' buttons. A large orange callout box at the bottom right contains the text: 'Data visualization tool pulls in data from PDB and visualized that alongside the article (linking out to PDB for data)'.

Download PDF Export More options... Search ScienceDirect Advanced search

Volume 400, Issue 3, 16 July 2010, Pages 295–308

ELSEVIER

Authors include Protein Data Bank accession numbers in their article

doi:10.1016/j.jmb.2010.05.030 Get rights and content

Abstract

Structure-based protein sequence alignments of family B DNA polymerases revealed a conserved motif that is formed from interacting residues between loops from the N-terminal and palm domains and between the N-terminal loop and a conserved proline residue. The importance of the motif for function of the bacteriophage T4 DNA polymerase was revealed by suppressor analysis. T4 DNA polymerases that form weak replicating complexes cannot replicate DNA when the dGTP pool is reduced. The conditional lethality provides the means to identify amino acid substitutions that restore replication activity under low-dGTP conditions either by correcting the defect produced by the first amino acid substitution or by generally increasing the stability of polymerase complexes; the ... effectively counter the reduced stability caused by a var ... substitutions that increase the stability of polymerase co ... the antiviral drug phosphonoacetic acid. Amino acid sub ... DNA under low-dGTP conditions or drug sensitivity were ...

Citing articles (5)

Related book content

Proteins in this article

1Q9X 1IG9

1Q9X Jmol

Rotate Reset

Data visualization tool pulls in data from PDB and visualized that alongside the article (linking out to PDB for data)