stm

INTERNATIONAL ASSOCIATION OF SCIENTIFIC, TECHNICAL & MEDICAL PUBLISHERS

Integrating Data and Publications

how to make things better

Integration of Research Data and Publications

Opportunities for Data Exchange

Eefke Smit
International Association of STM Publishers
Director, Standards and Technology

Amsterdam, Preparde Workshop at IDCC 2013,
17 January 2012

A famous paper in Nature: DNA structure - 1953

No. 4356 April 25, 1953 NATURE 737

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

¹ Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
² Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
³ Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
⁴ Ekman, V. W., *Arkiv. Mat. Astron. Fysik.* (Stockholm), **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[3,4] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

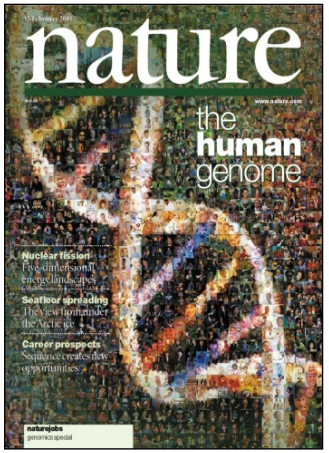
- 1 page
- 2 authors
- 1 figure
- no data

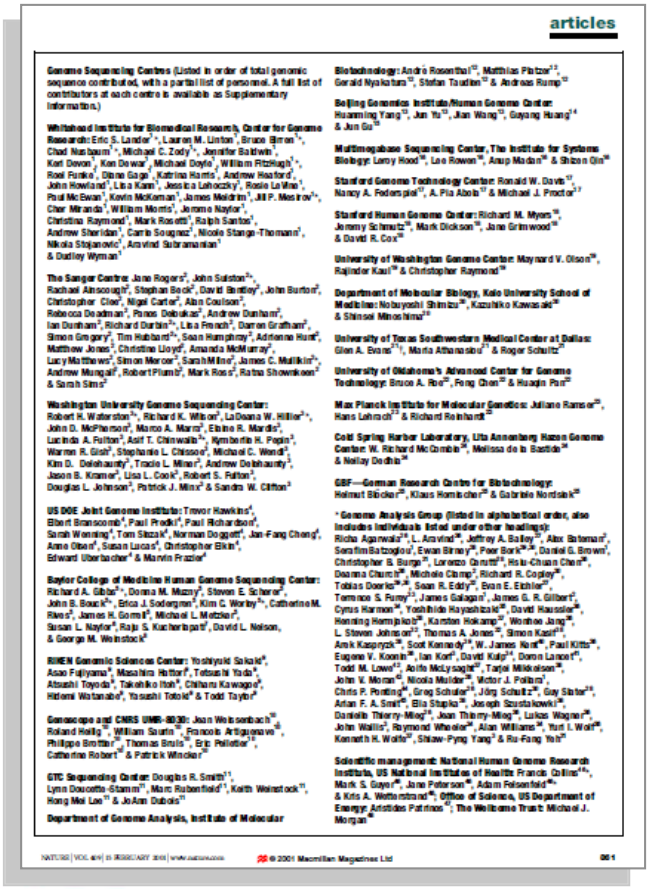
Source: V. Kiermer, Nature Publishing Group, 2011

Opportunities for Data Exchange

Nature in 2001:
The human genome issue

- 62 pages, 49 figures, 27 tables




articles

Genome Sequencing Centres (Listed in order of total genomic sequence contributed, with a partial list of personnel. A full list of contributors at each centre is available as Supplementary Information.)

Whitehead Institute for Biomedical Research, Center for Genome Research: Eric S. Lander*†, Lauren M. Linton†, Bruce Birren*†, Chad Nusbaum*†, Michael C. Zody*†, Jennifer Baldwin†, Keri Devon†, Ken Dewar†, Michael Doyle†, William FitzHugh†, Rosl Funke†, Diane Gage†, Katrina Harris†, Andrew Heaford†, John Howland†, Lisa Kann†, Jessica Lehoczky†, Rosio LeVine†, Paul McEwan†, Kevin McKernan†, James Meldrim†, Jill P. Mesirov*†, Cher Miranda†, William Morris†, Jerome Naylor†, Christina Raymond†, Mark Rosetti†, Ralph Santos†, Andrew Sheridan†, Carrie Sougnez†, Nicole Stange-Thomann†, Nikola Stojanovic†, Aravind Subramanian†, & Dudley Wyman†

The Sanger Centre: Jane Rogers*, John Sulston*†, Rachael Ainscough†, Stephan Beck†, David Bentley†, John Burton†, Christopher Clee†, Nigel Carter†, Alan Coulson†, Rebecca Deadman†, Panos Deloukas†, Andrew Dunham†, Ian Dunham†, Richard Durbin*†, Lisa French†, Darren Grafham†, Simon Gregory†, Tim Hubbard*†, Sean Humphray†, Adrienne Hunt†, Matthew Jones†, Christine Lloyd†, Amanda McMurray†, Lucy Matthews†, Simon Mercer†, Sarah Milne†, James C. Mullikin*†, Andrew Mungall†, Robert Plumb†, Mark Ross†, Ratna Shownkeen*†, & Sarah Sims†

Washington University Genome Sequencing Center: Robert H. Waterston*†, Richard K. Wilson†, LaDeana W. Hillier*†, John D. McPherson†, Marco A. Marra†, Elaine R. Mardis†, Lucinda A. Fulton†, Asif T. Chinwalla*†, Kymberlie H. Pepin†, Warren R. Gish†, Stephanie L. Chissoe†, Michael C. Wendl†, Kim D. Delehaunty†, Tracie L. Miner†, Andrew Delehaunty†, Jason B. Kramer†, Lisa L. Cook†, Robert S. Fulton†, Douglas L. Johnson†, Patrick J. Minx†, & Sandra W. Clifton†

US DOE Joint Genome Institute: Trevor Hawkins*, Elbert Branscomb*, Paul Predki*, Paul Richardson*, Sarah Wenning*, Tom Slezak*, Norman Doggett*, Jan-Fang Cheng*, Anne Olsen*, Susan Lucas*, Christopher Elkin*, Edward Uberbacher* & Marvin Frazier*

Baylor College of Medicine Human Genome Sequencing Center: Richard A. Gibbs*†, Donna M. Muzny†, Steven E. Scherer†, John B. Bouck†, Erica J. Sodergren†, Kim C. Worley†, Catherine M. Rives†, James H. Gorrell†, Michael L. Metzker†, Susan L. Naylor†, Raju S. Kucherlapati†, David L. Nelson, & George M. Weinstock†

RIKEN Genomic Sciences Center: Yoshiyuki Sakaki*, Asao Fujiyama*, Masahira Hattori†, Tetsushi Yada†, Atsushi Toyoda†, Takehiko Itoh†, Chiharu Kawagoe†, Hidemi Watanabe†, Yasushi Totoki† & Todd Taylor†

Genoscope and CNRS UMR-8030: Jean Weissenbach*, Roland Heilig*†, William Saurin*†, Francois Artiguenave*†, Philippe Brottier*†, Thomas Bruls*†, Eric Pelletier*†, Catherine Robert*† & Patrick Wincker*†

GTC Sequencing Center: Douglas R. Smith*†, Lynn Doucette-Stamm*†, Marc Rubenfield*†, Keith Weinstock*†, Hong Mei Lee*† & JoAnn Dubois*†

Department of Genome Analysis, Institute of Molecular Biotechnology: André Rosenthal*†, Matthias Platzer*†, Gerald Nyakatura*†, Stefan Taudien*† & Andreas Rump*†

Beijing Genomics Institute/Human Genome Center: Huanming Yang*†, Jun Yu*†, Jian Wang*†, Guyang Huang*† & Jun Gu*†

Multimegabase Sequencing Center, The Institute for Systems Biology: Leroy Hood*†, Lee Rowen*†, Anup Madan*† & Shizen Qin*†

Stanford Genome Technology Center: Ronald W. Davis*†, Nancy A. Federspiel*†, A. Pia Abola*† & Michael J. Proctor*†

Stanford Human Genome Center: Richard M. Myers*†, Jeremy Schmutz*†, Mark Dickson*†, Jane Grimwood*† & David R. Cox*†

University of Washington Genome Center: Maynard V. Olson*†, Rajinder Kaul*† & Christopher Raymond*†

Department of Molecular Biology, Keio University School of Medicine: Nobuyoshi Shimizu*†, Kazuhiko Kawasaki*† & Shinsei Minoshima*†

University of Texas Southwestern Medical Center at Dallas: Glen A. Evans*†, Maria Athanasiou*† & Roger Schultz*†

University of Oklahoma's Advanced Center for Genome Technology: Bruce A. Roe*†, Feng Chen*† & Huaqin Pan*†

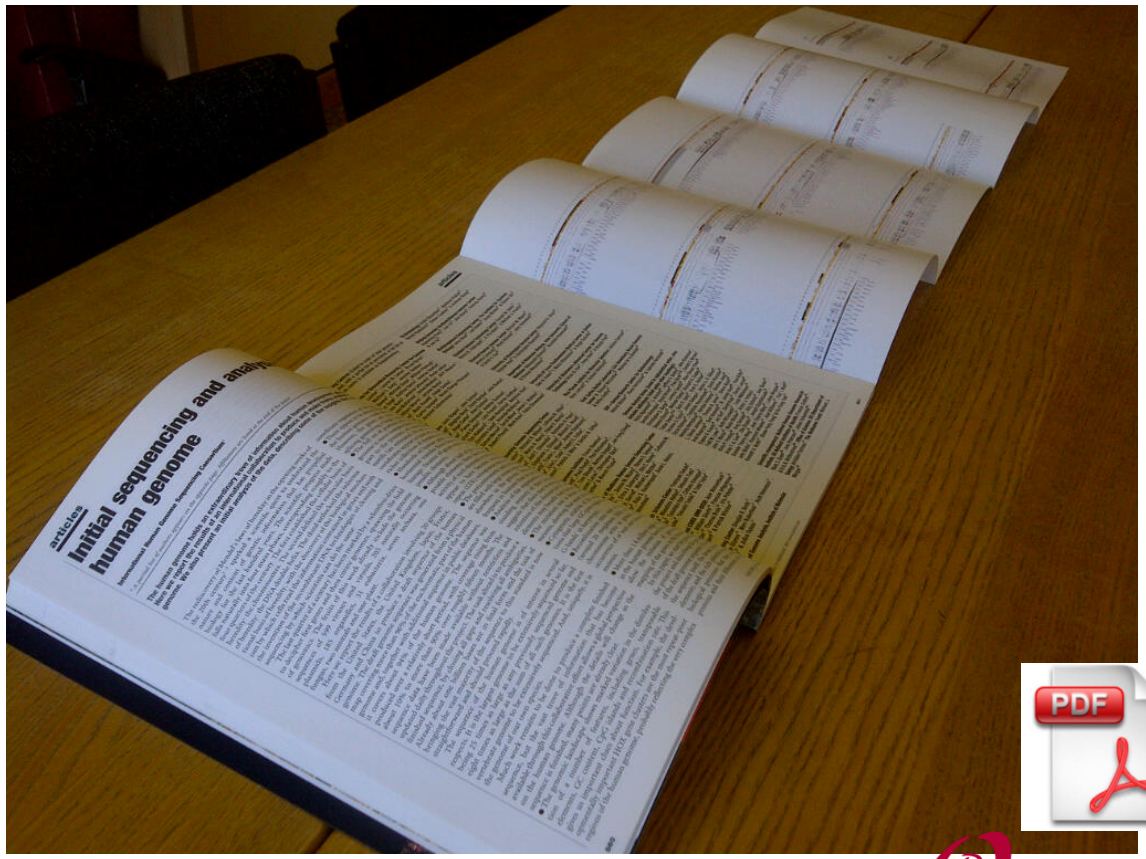
Max Planck Institute for Molecular Genetics: Juliane Ramser*†, Hans Lehrach*† & Richard Reinhardt*†

Cold Spring Harbor Laboratory, Uta Ansorge Hazen Genome Center: W. Richard McCombie*†, Melissa de la Bastide*† & Nelloy Dedhia*†

GBF—German Research Centre for Biotechnology: Helmut Blöcker*†, Klaus Homischer*† & Gabriele Nordsiek*†

* Genome Analysis Group (listed in alphabetical order, also includes individuals listed under other headings): Richa Agarwala*†, L. Aravind*†, Jeffrey A. Bailey*†, Alex Bateman†, Serafim Batzoglou†, Ewan Birney*†, Peer Bork*†‡, Daniel G. Brown†, Christopher B. Burge*†, Lorenzo Cerutti*†, Hsiu-Chuan Chen*†, Deanna Church*†, Michele Clamp†, Richard R. Copley*†, Tobias Doerks*†, Sean R. Eddy*†, Evan E. Eichler*†, Terrance S. Furey*†, James Galagan†, James G. R. Gilbert†, Cyrus Harmon*†, Yoshihide Hayashizaki*†, David Haussler*†, Henning Hermjakob*†, Karsten Hokamp*†, Wonhee Jang*†, L. Steven Johnson*†, Thomas A. Jones*†, Simon Kasif*†, Arek Kaspryzk*†, Scot Kennedy*†, W. James Kent*†, Paul Kitts*†, Eugene V. Koonin*†, Ian Korf†, David Kulp*†, Doron Lancet*†, Todd M. Lowe*†, Aoife McLysaght*†, Tarjei Mikkelsen*†, John V. Moran*†, Nicola Mulder*†, Victor J. Pollara†, Chris P. Ponting*†, Greg Schuler*†, Jörg Schultz*†, Guy Slater*†, Arian F. A. Smit*†, Elia Stupka*†, Joseph Szustakowki*†, Danielle Thierry-Mieg*†, Jean Thierry-Mieg*†, Lukas Wagner*†, John Wallis*†, Raymond Wheeler*†, Alan Williams*†, Yuri I. Wolf*†, Kenneth H. Wolfe*†, Shiaw-Pyng Yang*† & Ru-Fang Yeh*†

Scientific management: National Human Genome Research Institute, US National Institutes of Health: Francis Collins*†‖, Mark S. Guyer*†, Jane Peterson*†, Adam Felsenfeld*†‖ & Kris A. Wetterstrand*†; Office of Science, US Department of Energy: Aristides Patrinos*†; The Wellcome Trust: Michael J. Morgan*†


articles
Initial sequencing and analysis of the human genome

stm

Source: V. Kiermer, Nature Publishing Group, 2011

Opportunities for Data Exchange

The human genome at 10 – 2010
Nature now in an iPad edition:

Source: V. Kiermer, Nature Publishing Group, 2011

A thousand genomes – 2010

http://www.nature.com/nature/journal/v467/n7319/full/nature09534.html

Source: V. Kiermer, Nature Publishing Group, 2011

Opportunities for Data Exchange

new publishing models

author information

doi

live updates

article-level metrics

Tool box to print, download reference, share: email, social media, bookmark

Related content

Figure previewer

Collapsible sections

Source: V. Kiermer, Nature Publishing Group, 2011

Opportunities for Data Exchange

Data Repositories and Journals

PANGAEA®
Data Publisher for Earth & Environmental Science

Not logged in (log in or sign up)

Always quote citation when using data!

Data Description

Show Map | Google Earth | RIS | BibTeX

Citation: **Indermühle, Andreas; Monnin, Eric; Stauffer, Bernhard; Stocker, Thomas F; Wahlen, Martin (2000):** A high-resolution record of the atmospheric CO2 concentration from 60-20 kyr BP from the Taylor Dome ice core, Antarctica. doi:10.1594/PANGAEA.710905

Reference(s): **Indermühle, Andreas; Monnin, Eric; Stauffer, Bernhard; Stocker, Thomas F; Wahlen, Martin (2000):** Atmospheric CO2 concentration from 60 to 20 kyr BP from the Taylor Dome ice cores, Antarctica. *Geophysical Research Letters*, **27(5)**, 735-738, doi:10.1029/1999GL010960

Abstract: A high-resolution record of the atmospheric CO2 concentration from 60 to 20 thousand years before present (kyr BP) based on measurements on the ice core of Taylor Dome, Antarctica is presented. This record shows four distinct peaks of 20 parts per million by volume (ppmv) on a millennial time scale. Good correlation of the CO2 record with temperature reconstructions based on stable isotope measurements on the Vostok ice core (Antarctica) is found.

Project(s): **European Project for Ice Coring in Antarctica** (EPICA)

Coverage: *Latitude:* -77.783330 * *Longitude:* 158.716670

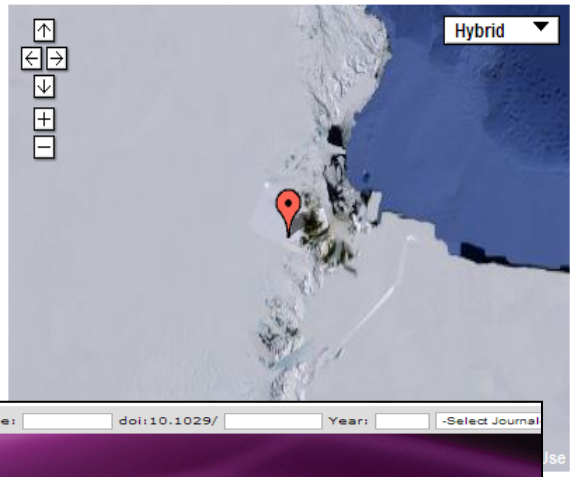
Minimum DEPTH, ice/snow: 380.8 m

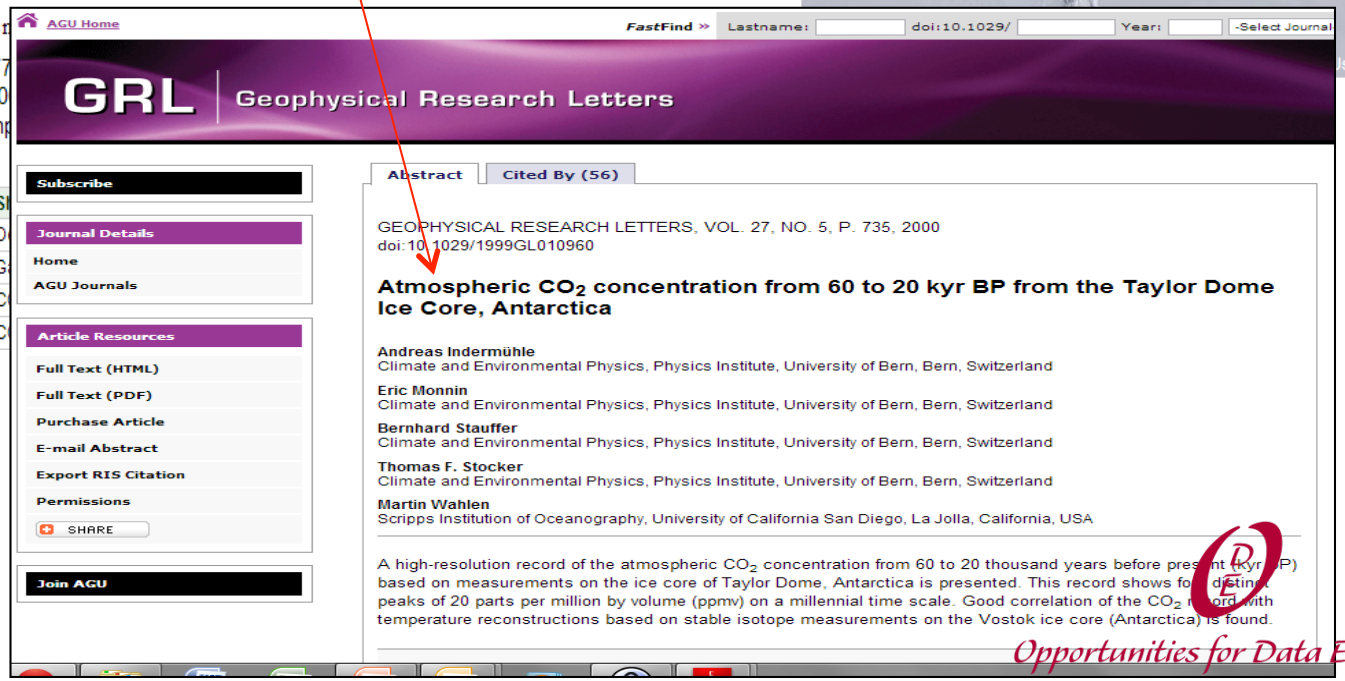
Event(s): **Taylor_Dome** (TD) * *Latitude:* -77
Date/Time End: 1994-12-31T00:0
TAYLOR_DOME * *Basis:* Samp

Comment: Depth is top of sample interval (m)

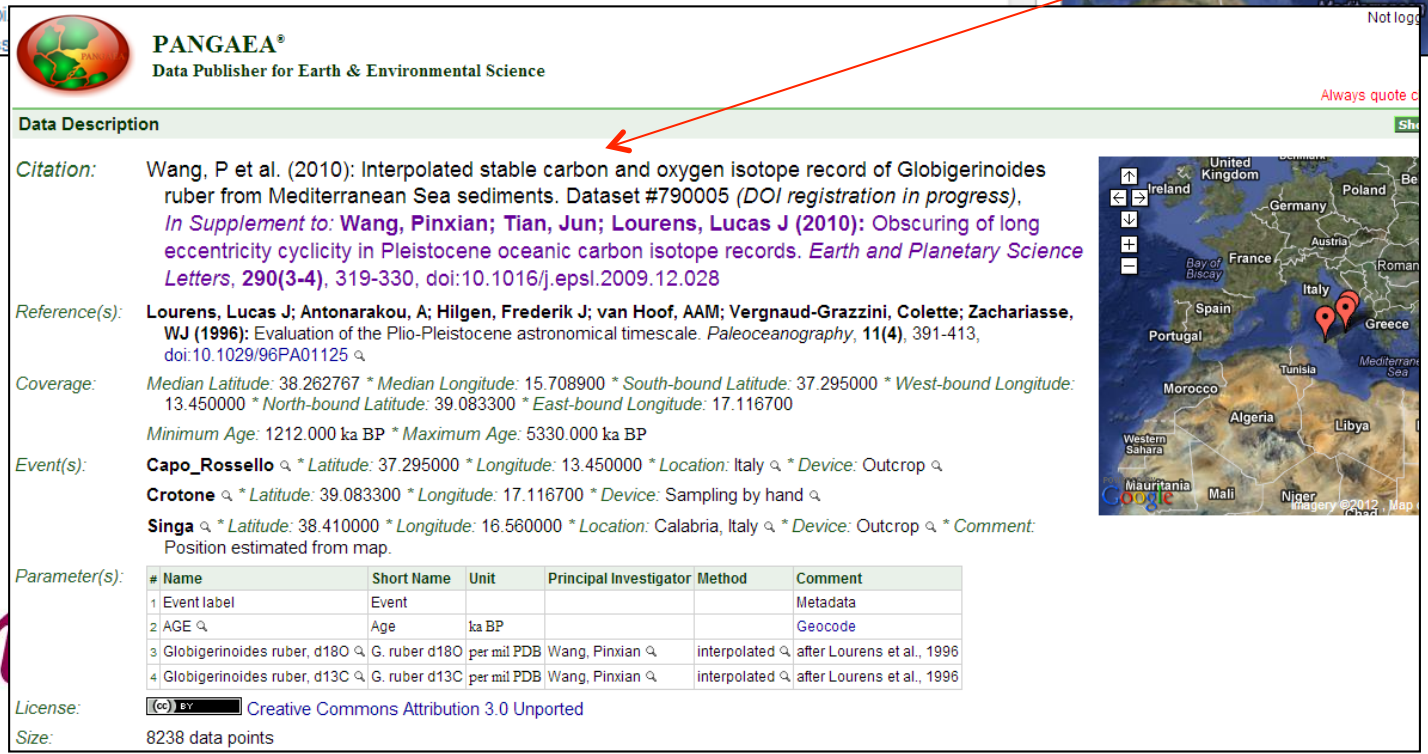
Parameter(s):

#	Name	
1	DEPTH, ice/snow	D
2	Gas age	G
3	Carbon dioxide	C
4	Carbon dioxide, standard deviation	C





From the Journal Article to the Data Repository



Datarepositories and Journals

DRYAD

Login

Search Data

Submit Data Now!

See how to submit

My Account

Login or Register

Browse

Authors

Journal Title

Information

Depositing Data

Using Data

Dryad Members

Journal Archiving Policy

About Dryad

Dryad Blog

Dryad Documentation

Data from: Synchrotron X-ray tomographic microscopy of fossil embryos

When using this data, please cite the original article:

Donoghue PCJ, Bengtson S, Dong X, Gostling NJ, Huldtgren T, Cunningham JA, Yin C, Yue Z, Peng F, Stampanoni M (2006) Synchrotron X-ray tomographic microscopy of fossil embryos. Nature 442: 680-683. doi:10.1038/nature04890

Additionally, please cite the Dryad data package:

Donoghue PCJ, Bengtson S, Dong X, Gostling NJ, Huldtgren T, Cunningham JA, Yin C, Yue Z, Peng F, Stampanoni M (2006) Data from: Synchrotron X-ray tomographic microscopy of fossil embryos. Dryad Digital Repository. doi:10.5061/dryad.fn78r

Cite | Share

Dryad Package Identifier doi:10.5061/dryad.fn78r 38 views

Abstract Fossilized embryos from the late Neoproterozoic and earliest Phanerozoic have caused much excitement because they preserve the earliest stages of embryology of animals that represent the initial diversification of metazoans. However, the potential of this material has not been fully realized because of reliance on traditional, non-destructive methods that allow an
interior
complet
impregn
distribut
cement
blastom
as a sc

Subscribe

Login Cart

nature International weekly journal of science

Search go Advanced search

Access

To read this story in full you will need to login or make a payment (see right).

nature.com > Journal home > Table of Contents

Letter

Nature **442**, 680-683 (10 August 2006) | doi:10.1038/nature04890; Received 21 February 2006; Accepted 10 May 2006

Synchrotron X-ray tomographic microscopy of fossil embryos

Philip C. J. Donoghue[1], Stefan Bengtson[2], Xi-ping Dong[3], Neil J. Gostling[1], Therese Huldtgren[2], John A. Cunningham[1,4], Chongyu Yin[5], Zhao Yue[2,5], Fan Peng[3] & Marco Stampanoni[6]

1. Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK
2. Department of Palaeozoology, Swedish Museum of Natural History, Stockholm, Sweden
3. School of Earth and Space Sciences, Peking University, Beijing, China
4. Department of Earth and Ocean Sciences, 4 Brownlow Street, Liverpool L69

ARTICLE LINKS

- Figures and tables
- Supplementary info

SEE ALSO

- Editor's Summary

ARTICLE TOOLS

- Send to a friend
- Export citation
- Export references

I want to purchase this article

Price: €30

In order to purchase this article you must be a registered user.

Register now

I want to subscribe to *Nature*

Price: US$199

This includes a free subscription to *Nature News* together with *Nature Journal*.

Subscribe now

Personal subscribers to *Nature* can view

stm

Data Journals

(GIGA)n DB Beta

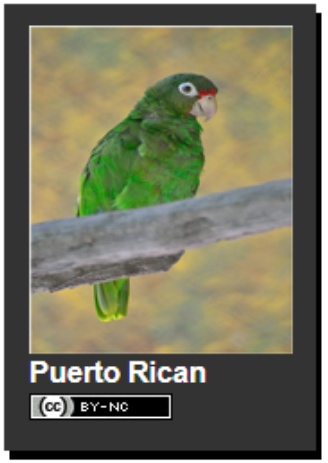
Revolutionizing data dissemination, organization, and use

Home | About | Contact | Terms of use

search

Puerto Rican parrot

These data represent the first assembly of a genome sequence for a critically endangered parrot (*Amazona vittata*) endemic to the United States, and also the first genome of a species from the diverse and ecologically important genus *Amazona* native to South America and the Caribbean. One sample has been selected from the non-reproductive female at Rio Abajo Breeding Facility in Puerto Rico (IACUC#201109.1), and sequenced on Illumina HiSeq platform with both fragment and paired-end sequencing approaches, resulting in a total of 42,479,499,706 bases. We predicted a total coverage depth of 26.89X of the parrot's genome: 17.08X coverage for the short fragment reads, and 9.8X coverage for the mate pairs. The sequencing was initiated with the construction of two genome libraries: a short fragment library (~300 bp inserts) for sequencing the majority of the genome, and a long fragment library (~2.5 Kb inserts) to generate scaffolds to be used to order and assemble contigs derived from the short fragment library. The Illumina paired-end and mate-pairs reads were assembled together with Ray (http://denovoassembler.sourceforge.net), with the k-mer defined iteratively. In total, given that the genome size is predicted to be 1.58Gb, with the total scaffold length of 1,184, 594,388 bp, the overall coverage of the genome is around 76%, a value that might be slightly overestimated given that some of the scaffolds may be overlapping but could not be assembled. Filtering followed by assembly resulted in 259,423 contigs

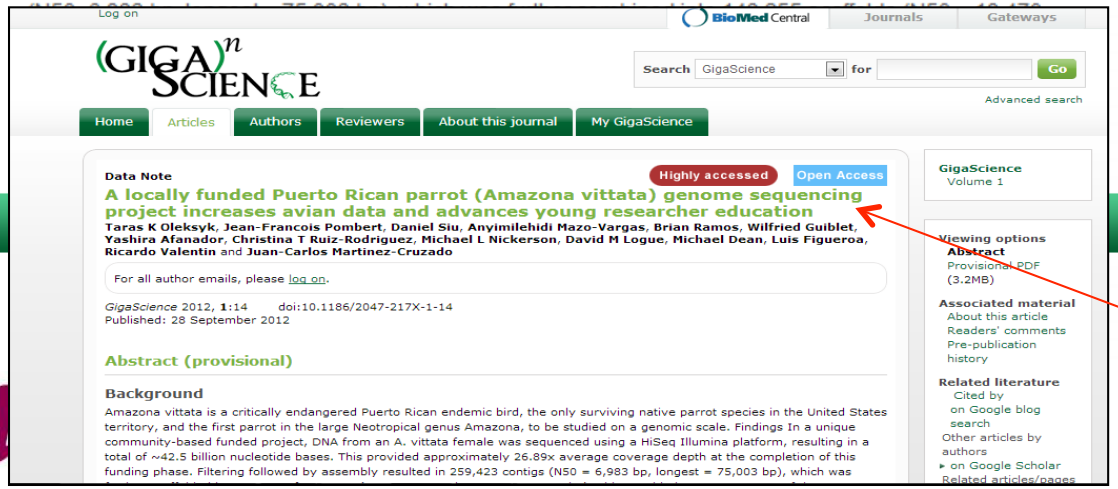


Citation

In accordance with our terms of use, please cite this dataset as:

Oleksyk, TK; Guiblet, W; Pombert, JF; Valentin, R; Martinez-Cruzado, JC (2012): Genomic data of the Puerto Rican Parrot (*Amazona vittata*) from a locally funded project. GigaScience. http://dx.doi.org/10.5524/100039

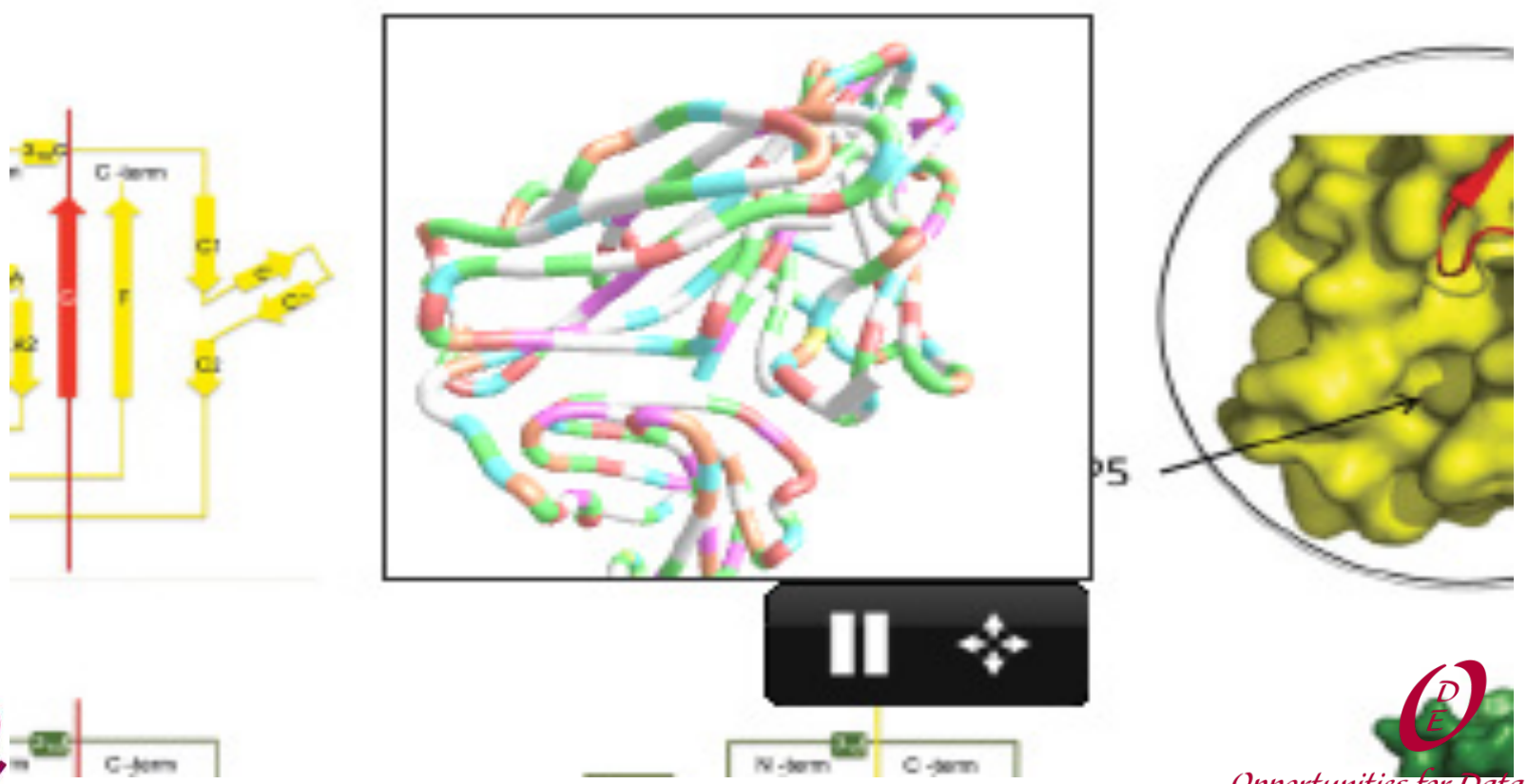
Related manuscript available at: doi:10.1186/2047-217X-1-14



From The BioChemical Journal, Portland Press:

Every wanted to inspect data referenced in articles? Utopia Documents allows you to interact directly with curated database entries. Play with molecular structures; edit sequence and alignment data; even plot curated tabular data yourself.

http://www.biochemj.org/bj/semantic_faq.htm



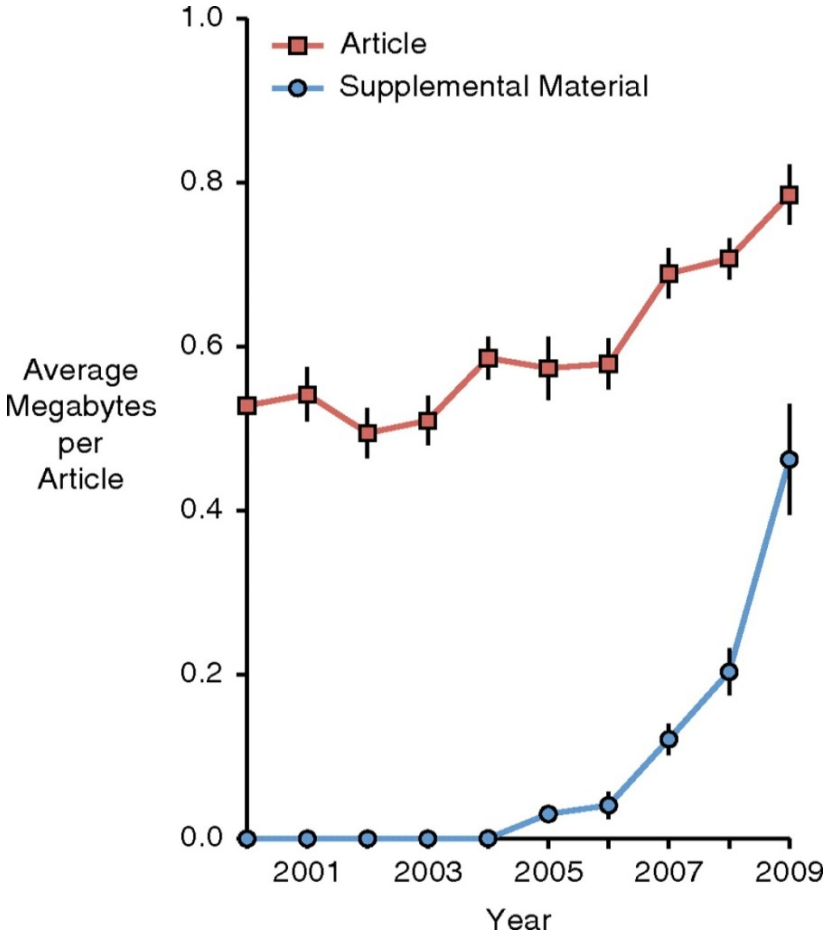
stm

Opportunities for Data Exchange

Elsevier offers data-viewers from within the article, to data stored elsewhere:

Opportunities for Data Exchange

While Supplementary Journal Files tend to grow out off hand:



Maunsell J J. Neurosci. 2010;30:10599-10600

The Graph depicts the average size of a Journal of Neuroscience article and supplemental material in megabytes.

As a consequence, the Journal no longer accepts supplementary files to manuscripts, soon the supplementary material would outgrow the article volume. The burden on the peer review process became simply to large.

Editors suspect researchers to treat supplements as data dumping grounds (Emily Marcus, Cell)

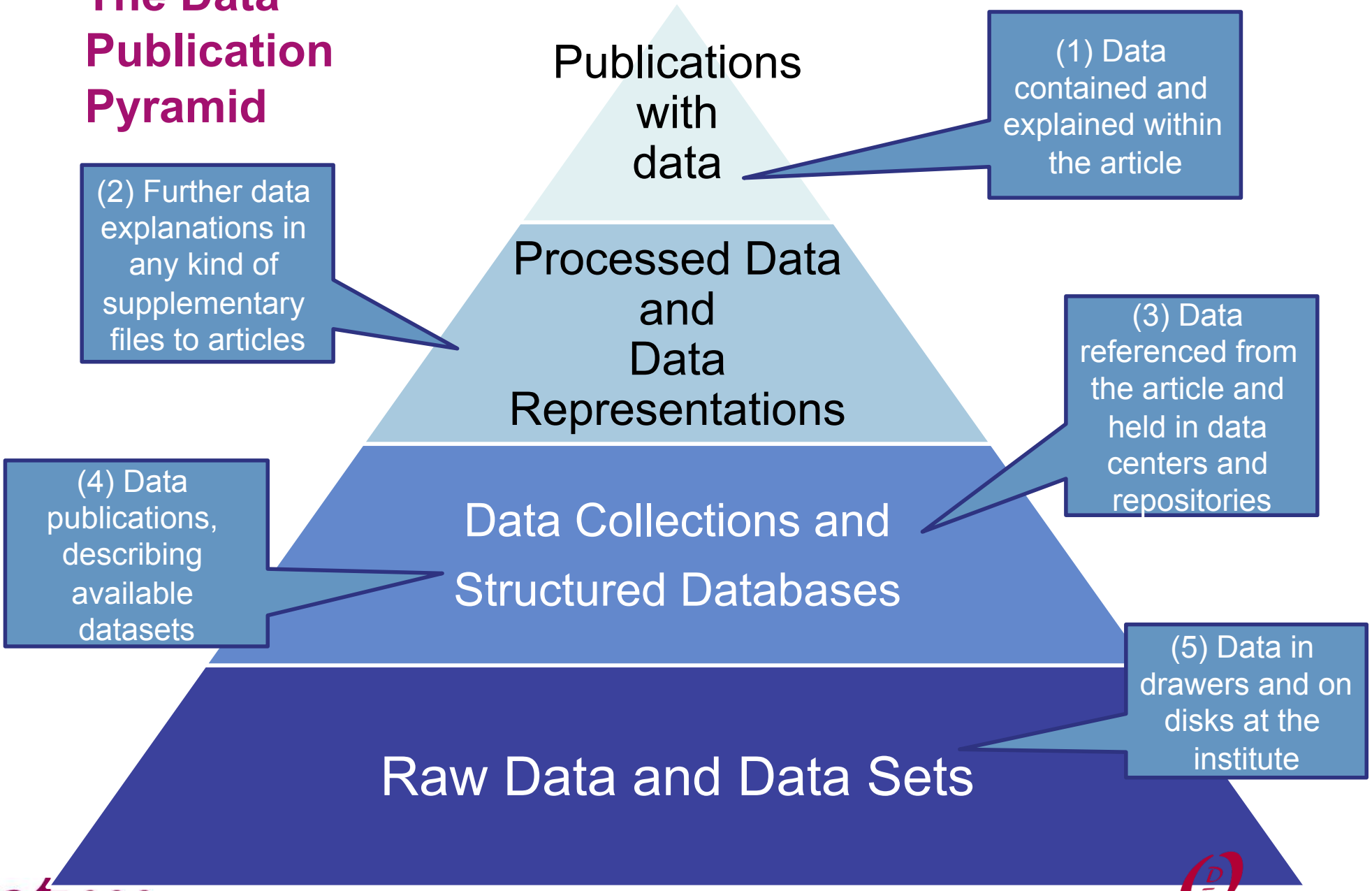
Publishers cannot guarantee proper preservation and future accessibility of supp files.

stm

SFN Society for Neuroscience The Journal of Neuroscience

DE Opportunities for Data Exchange

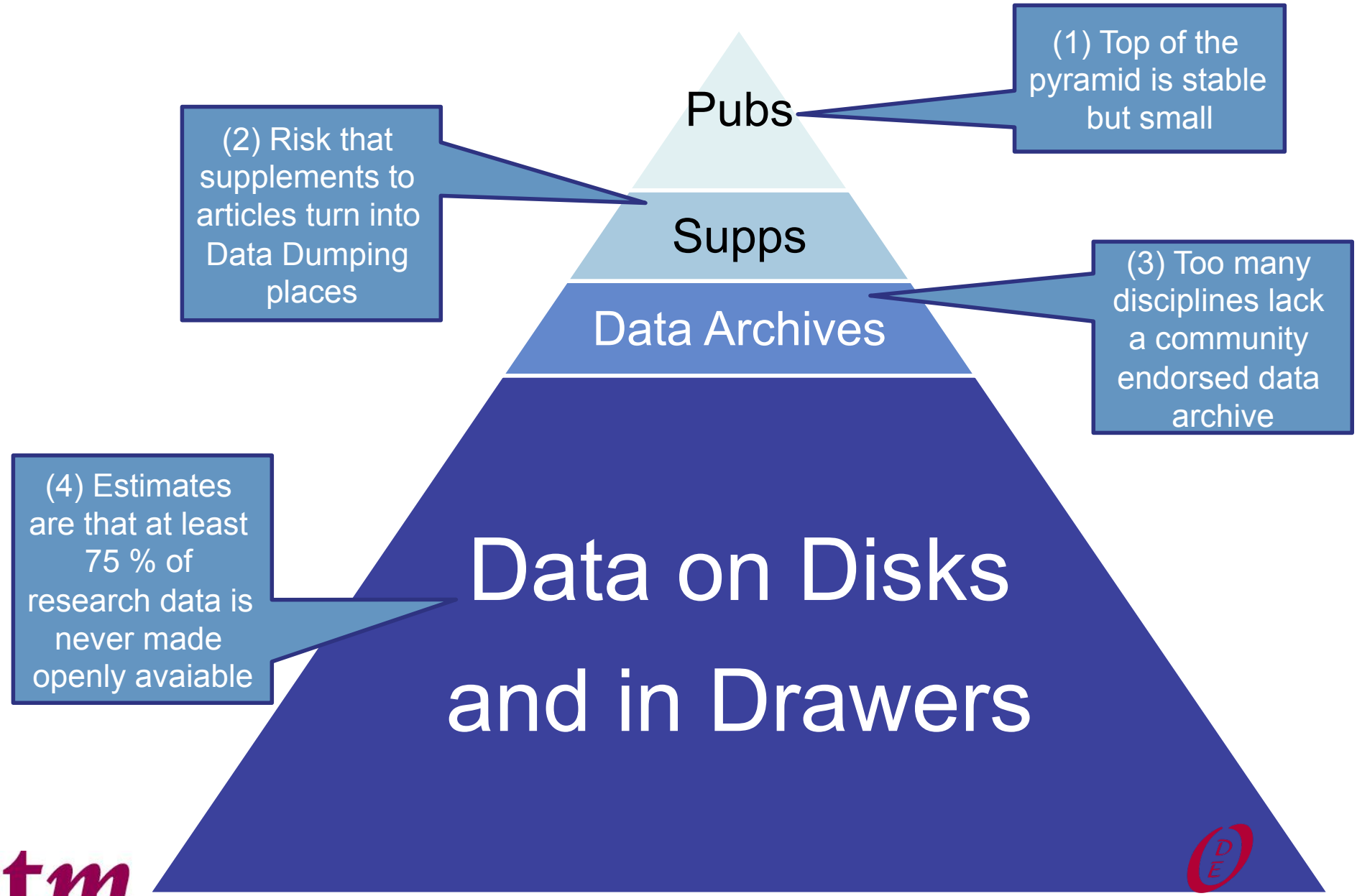
The Data Publication Pyramid



stm

Opportunities for Data Exchange

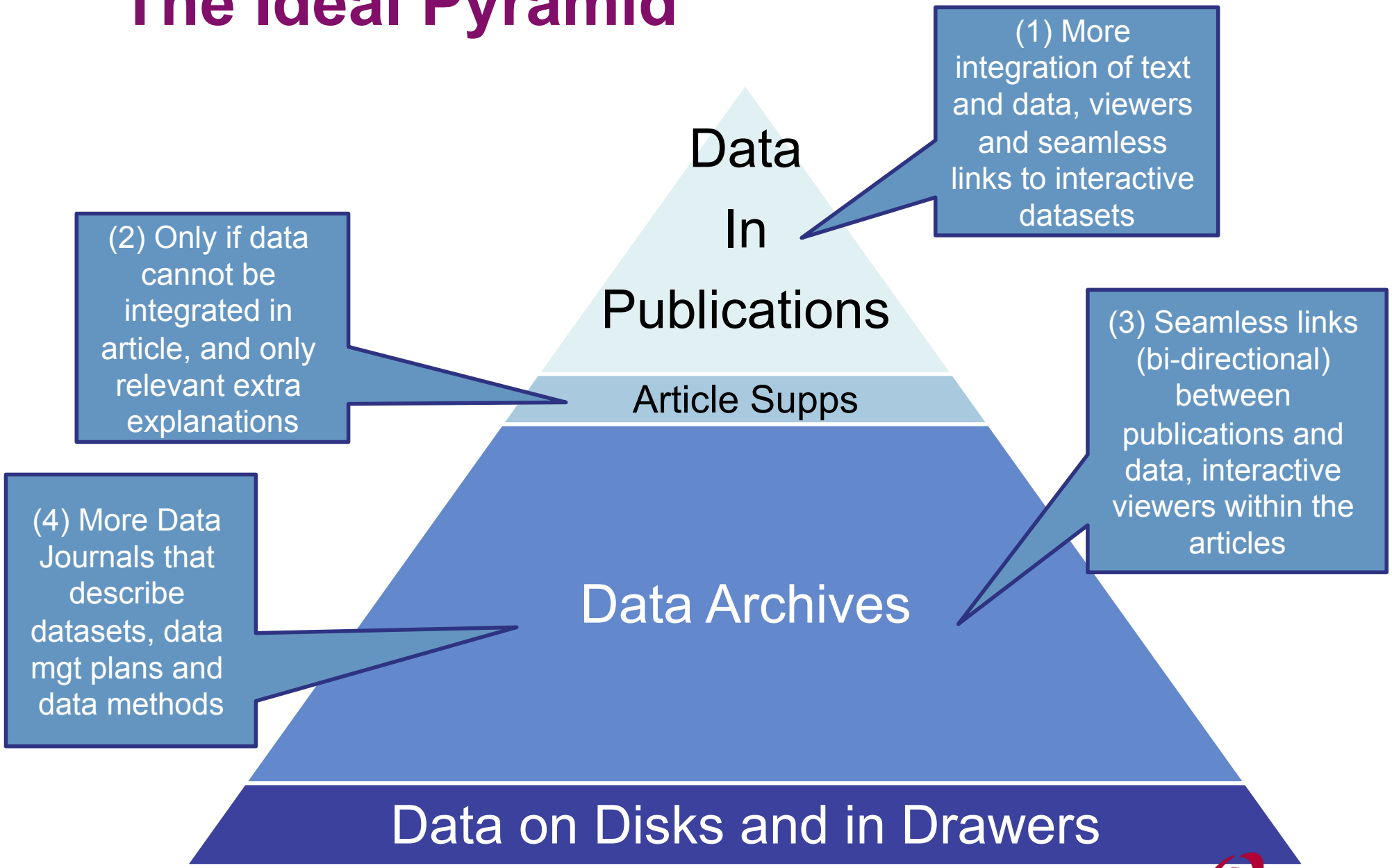
The Pyramid's likely short term reality:



stm

Opportunities for Data Exchange

The Ideal Pyramid



stm

Opportunities for Data Exchange

ODE recommendations: How can publishers help to make things better

- Clearer editorial policies on the availability of underlying data
- Recommend reliable and trustworthy Data Archives to authors
- Enhance articles for better integration of underlying data
- Endorse guidelines for proper citation of data
- Launch and sponsor Data Journals
- Ensure persistent identifiers and bi-directional linking
- Partner with reliable Data Archives for further integration of Data and Publications, including interactivity for re-use.

stm

Opportunities for Data Exchange

Questions ?

Eefke Smit
International Association of STM Publishers
Director, Standards and Technology
smit@stm-assoc.org

stm

Opportunities for Data Exchange