Scientific Data

A new data focused OA peer reviewed platform

Rigour and Openness in 21st Century Science April 12, 2013

> Jason Wilde Business Development Director Nature Publishing Group





Overview

- The evolution of research
- The article of the future
- What authors and readers say about data
- What is Scientific Data?
- What is a data descriptor?
- License types
- The editors

ns. com April 25, 1953

NATURE

"Brook S. B. Gernet, H., and Jones, W., Feel, Ster., 48, Lat.

* Language Station, St. St., Stor., Std. Stop. Acres, Soc., Studies, Stop., 4, 125 (1982). First St. S. L. Yard Str Paper or Plan Strong Street, 13 10 Calc. Vision S. W., and the Arrest Park (Statebook S.) 10 Cents

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

Will with to suggest a structure for the solt of decays/from market and (D.N.A.). This plantum has novel features which are of considerable

hologonic interest.

A structure for toolsis soul has strong been perpend by Trailing and Circy. They kindly made their management and fall to to be in advanced or subcleament. Their model crossed of three interested offices, with the phosphone term to fiftee some and the boson in the outside. So our opinion, this articular is trained in the model in the subclean Cir. We believe that the material which government of the property of the subclean Cir. We believe that the material which government of the subclean of the subclean Cir. We believe that the material which gives the contract of the subclean the action hydrogen atoms it is not stone what forces would had also attraction regular, repeatedly as the

second hold the attentions together, compossing as the angulation of angular planning more than a suit will repet units extract. If former of this was due Wands of a suit of the contract of the same due Wands Anather three-class experience. In the smooth the general by France in the present. In the smooth the phospharm are on the contract and the beams on the particular of a supplier by high regard by the state and to disability of the present of the due on the attentions on the other hand by the desired, and for structures on the other hand or desired to expense.

reducing different electrons for the soft of deceptions contact acid. This structure has two belong chairs such coaled round the same han two diagrams. We have made the need changed need private temply flar each claim consists of phosphare disolate groups justing \$1.5 decay the disolate makes with \$2.5 to large. The year chains that not their bases; are related by a dyest perpendicular to the like-axis. Dith classes follow right:

equipment, and in Dr. St. E. R. Didson and the ... is a results on each charactery F 4.A. in the delay-mplate and officers of B.B.B. Didsoney If for their time. We have assumed on angle of 30° between part in colling the observations. structure repeats after 20 months in moth chain, that in after \$4 A. The darance of a physphonia area. from the fibre said is \$6.5. As the phospharae accom-the outside, exists been may accome to these.

The direction continues have every mercion to Shape.
The direction is not upon tools, and its waited statistical in relief high. At lower whiter contempts we mould request the house to till on tiled him processes could be extended the mount of the processes and the processes are considerable.

This travel forcing of the atmospher is the measure in which the term chains are field signified by the parties and postulation leave. The places of the base parties and postulation leave. The places of the base are perpendicular to the division, there are julged signified to policy a right bate flows one chain being significant in pairs, a right bate flow one chain being hydrogen-bonded or a mogle fuse from the other state, so that the two lie side by aids work identical p-m-relinator. One of the pair breat he a portion and the other a portraiding for hissing to mean. The hydrogen bands are made as follows: pursue position. 2 to permittee pressure 2; pursue posture 6 to permittee prestrict 6.

If it is assessed that the bases only more in the

otraction in the most pleasible backmann forms (that is, with the laste rather than the expl. co-

that is, with the hore reador than (the right configurations). It is found that only against power of bases one horst repetition. These paper are a function partial configuration with theyence prepared and generalized partial configurations. In Solar records, if on adequate forms one assumption of a pair, on either whents, their on chains assumptions the other member must be when the stellarly first particles and or excellent partial to the experimental configuration and opinions. The sequence of bases on a neighb mission down and approach to be recovered to suit and the sequence of the sequence of the sequence of the forms of the sequence on this other control on general their class of the sequence on this other control on general their class of the sequence on this other class of the sequence on this other class of the sequence on this other class of the sequence on the sequence of the sequence rate obsits to given, then the sequence on the other class, is not mentacedly determined.

It has been found experimentally. that the ratio of the amounts of advance to thyrana, and the ratio of guantite to rytomes, are slowers very clear to empy for decayothese, madeir and

It is probably impossible to build this expenses, with a others engine in place of the decay-close, as the extra rapper store would reste too store a ran-for Waste request.

The previously published X-ray data's on dravy

about matter wild are invalidious for a rigorous too of our elevation. So far as we can sell, it is simply compacible with the expressional data. Soil is store to require our compacible as compared total if has been effected. the regionises are compressed upon 10 data from Attentional against terms outside security. Hence of Hause are given in the filling and experimental control of the solution and assume of the solution of the mainting presentant there exhaus no deviced out determined, within those country through, and settlingly our problemed conjunctional design them are determined.

main. Both visions filling eights standed believe the experimental date and some-timated believes, but sweep to the except and expensives. The dynd the authorized of the system of the contract is the two visions are present and expensive and the resolution of the product of property assertable products from the product of the product of the product of the contract from the product of the contract of the contract of the contract of the contract of the product of the contract of the contract of the product of the contract of the contract of the product of the contract of

the bases are on the hearts of the brief and the physicisms on simulations for the atomic, will be preliminary on simulations of the square and the solutions seems included to Dr. Jorey Donedon for the square and the solutions considered and the square and the solutions of the square and th

STATE Makes Published Groups

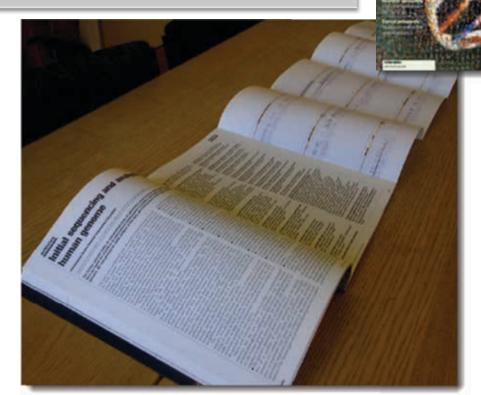
DNA Structure 1953

1 Page 2 Authors 1 Figure no data

Human Genome 2001

62 Pages, 150 Authors, 49 Figure, 27 tables

that disease being being a being a best

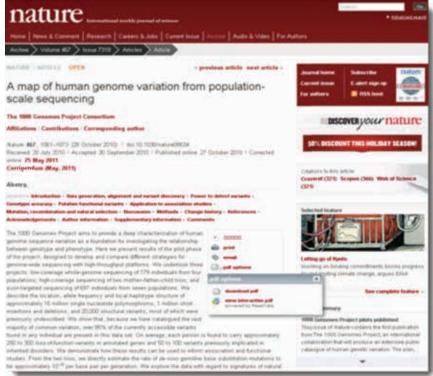


nature

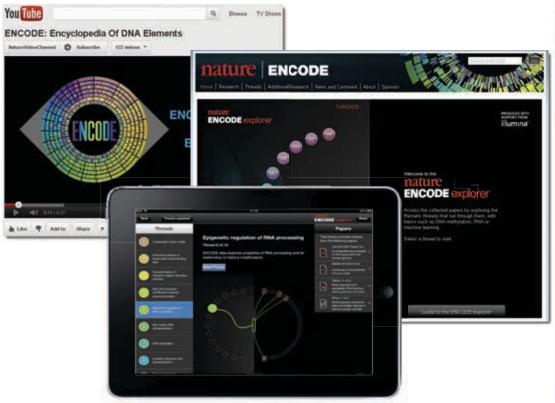
1000 Genome Project 2010

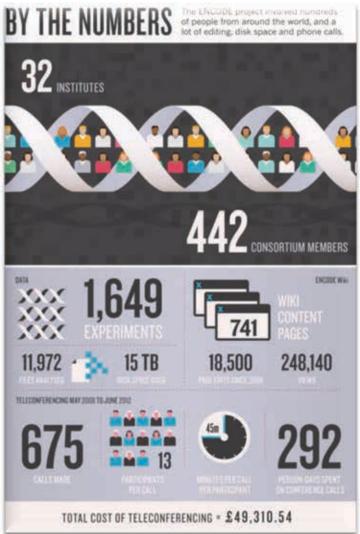
12 page, 76 Institutions, 12,145 SRA run ids



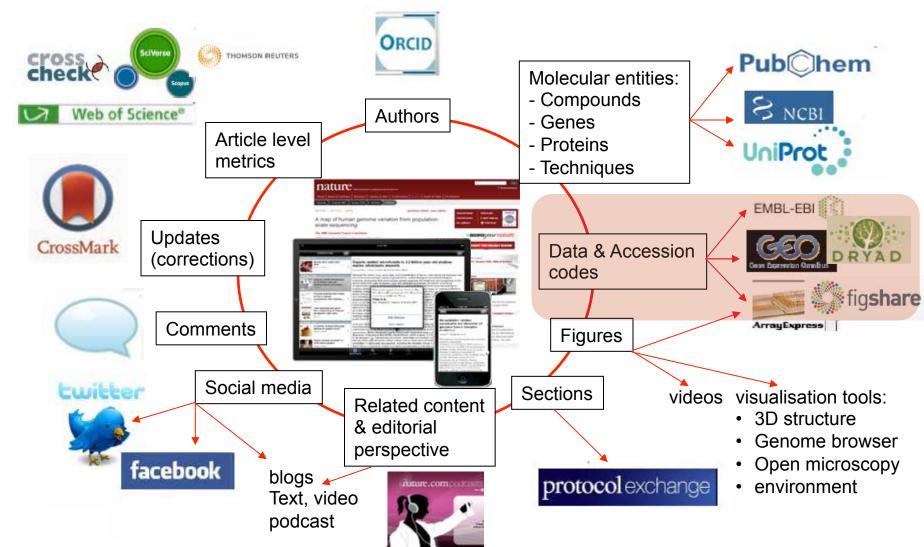


Encode Project 2012 30 papers, 3 Journals



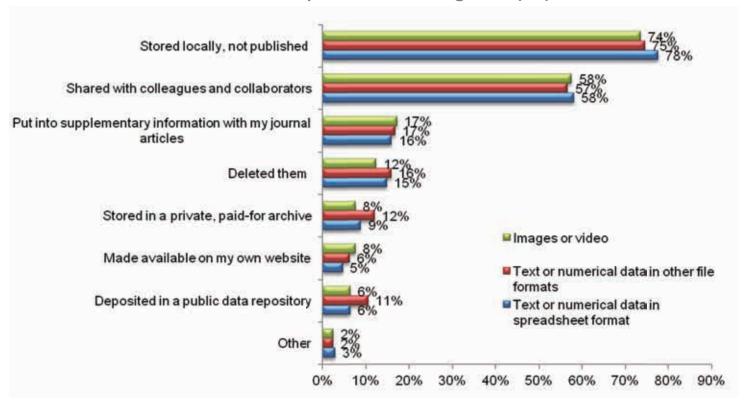


The article of the future...



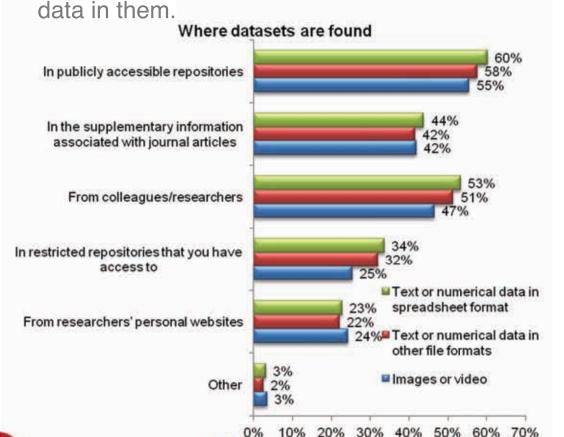
What do authors and readers say...

Storing Data: The majority of participants who produce each type of data say that some or all of it is stored locally and not published. Almost as many researchers delete their data as publish it alongside papers



What do authors and readers say...

Finding Data: Respondents find the data they are looking for in publicly available repositories despite a low proportion actually depositing their own



Other places datasets are found:

From my colleagues

tables, articles, books, etc

Webpages of the organizations, UN, CERN, ...

Universities etc

In my school s library

Reports and data compiled from experiments performed by technicians and students

through our library data bases

net

Usually I don't use text or numerical data in file formats other than spreadsheet format.

My ex-colleagues or job partners who do scientific jobs actually.

use self made procedures

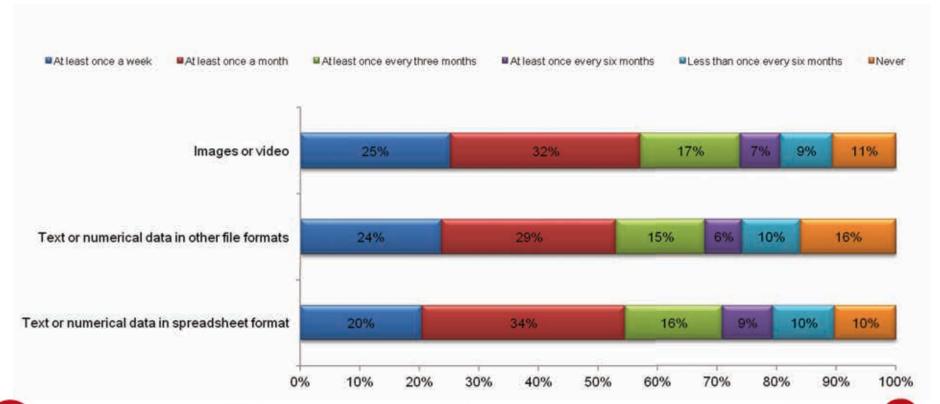
common areas in the disk space

Journals such as Nature (Biological), New Scientist and Scientific American

scientific journals . collections of data like "Landolt" , "Nuclear Data Tables" etc. monographs and other printed texts

What do authors and readers say...

Using Data: The majority of participants look for other researchers datasets, with more than half doing so once a month or more for each data type, and between a fifth and a quarter doing so once a week or more frequently.



What is Scientific Data?

SCIENTIFIC DATA

Search

Go

Home About For Authors Advisory and Editorial Board Open Access Contacts FAQ

Helping you publish, discover, and reuse research data







Credit

Credit, through a citable publication, for depositing & sharing your data

Reuse

Complete, curated & standardized descriptions enable the reuse of your data

Quality

Rigorous community based peer review



Discovery

Find datasets relevant to your research



Promotes & endorses open science principles & available to all through a Creative Commons license



Service

In-house curation, rapid peer review & publication of your data descriptions



Scientific Data Updates

Introducing Scientific Data Updates April 4, 2013

If not now then when - my view from within April 3, 2013

[Press Release] NPG to launch Scientific Data to help scientists publish and reuse research data April 3, 2013

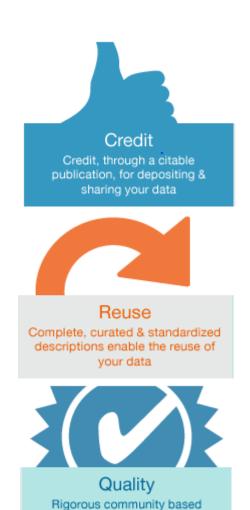
Follow Us





What is Scientific Data?

- Scientific Data is an Open Access, online-only platform containing data descriptors that describe and explain datasets, supported by an APC model.
- Data descriptors are a new type of content and can be viewed as 'secondary' material aimed at increasing the visibility and usability of datasets and to aid research reproducibility
 - datasets that relate to a previously published peerreviewed article, forming secondary material to further understand of the published work
 - datasets that would not otherwise be published in a peer-review article but that are of interest to the community (e.g. negative data)
- For all types of data the descriptor will be peer reviewed

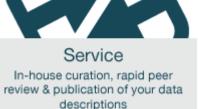


peer review

What is Scientific Data..?

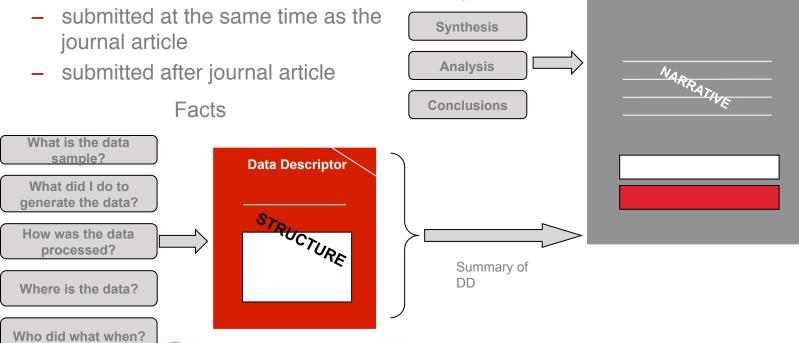
- As part of the peer review process we will check that the data is publically available in an approved data repository and follows community guidelines
- All content will be published open access with the author able to select from a number of options. In addition the descriptor metadata will be available under CC0 to enable TDM
- An in-house editorial team and new authoring tools are being developed to ensure the creation, submission, curation and publication of data descriptors is as simple as possible
- The external advisory board will represent different stakeholder views and provide feedback on key services.





What is a Data Descriptor?

- The data descriptor is only concerned with the facts behind the methodology of data generation/collection and processing.
- Contains both narrative sections and structured, curated information (metadata).
- A data descriptor can be:
 - submitted prior to journal article



Interpretation

Journal article

License types



Data: the raw datasets will reside in public repositories and we would expect this to be CC0 similar to Figshare and Dryad etc...



DATA DESCRIPTOR

Metadata: as NPG has already done with its existing Linked Data Portal the metadata about data descriptors in *Scientific Data* will be **CC0**



Narative/Figures: the narrative describing the methodology of data generation/collection and processing will be licensed under either of the following, by author choice:







The editors



Susanna-Assunta Sansone Honorary Academic Editor

Susanna focuses on the strategic development of *Scientific Data*, particularly by fostering relationships with the scientific community and helping to define standards for data reuse.

PhD at Imperial College London. In 2010, Susanna joined the University of Oxford e-Research Centre, where she is an Associate Director and Principal Investigator.





Andrew is responsible for developing the editorial policies of *Scientific Data*, in consultation with the Honorary Editor and Advisory Panel, and will work with the Editorial Board to ensure a fair and thorough peer review process for all submissions.

PhD at Stanford University in 2006. Postdoctoral work at the Max Planck Institute for Molecular Genetics in Berlin.