



# *Scientific Data*

*A new data focused OA peer reviewed platform*

Rigour and Openness in 21<sup>st</sup> 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



# Evolution of research



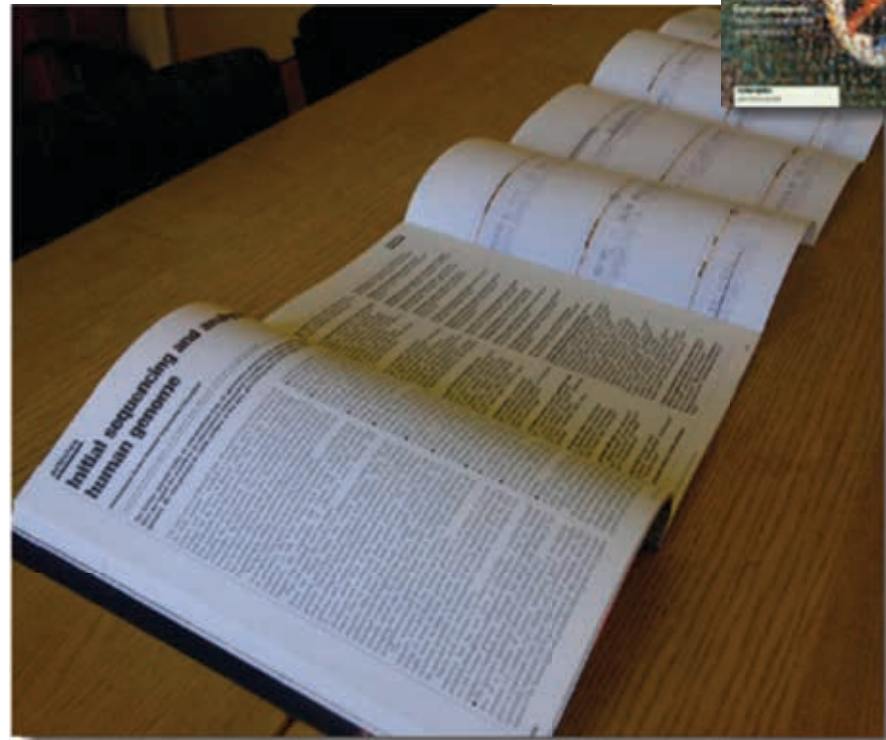
DNA Structure 1953

1 Page  
2 Authors  
1 Figure  
no data

# Evolution of research

## Human Genome 2001

62 Pages, 150 Authors,  
49 Figure, 27 tables





# Evolution of research

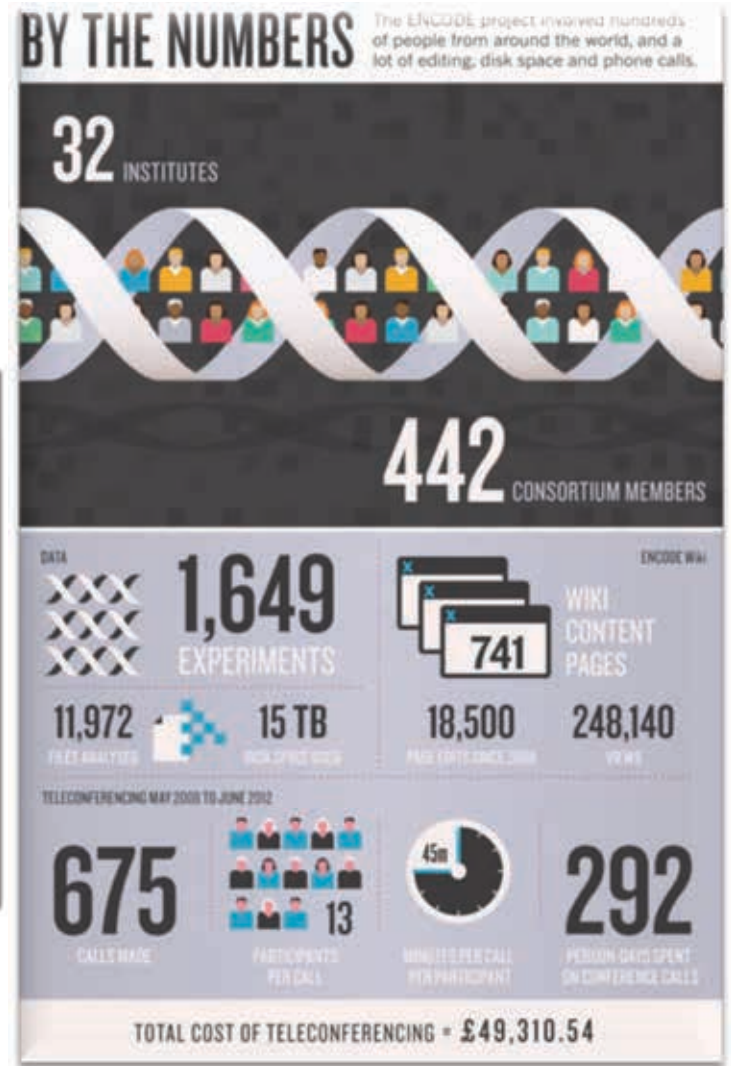
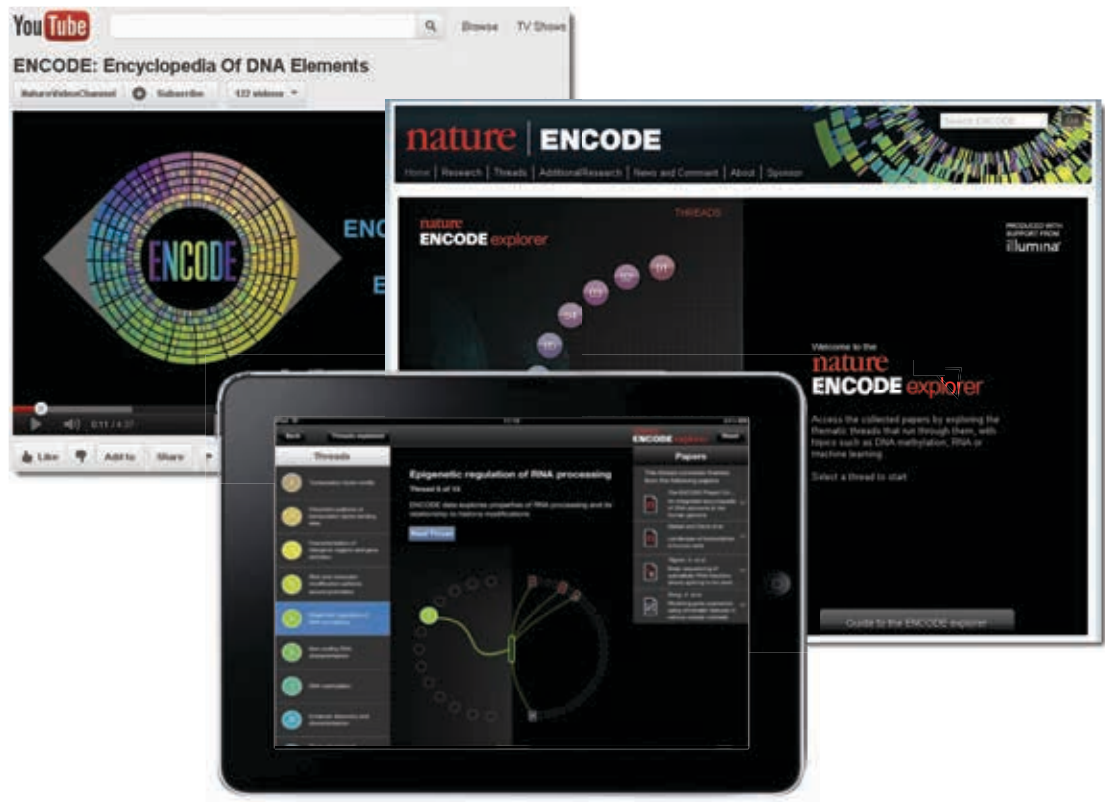
## 1000 Genome Project 2010

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

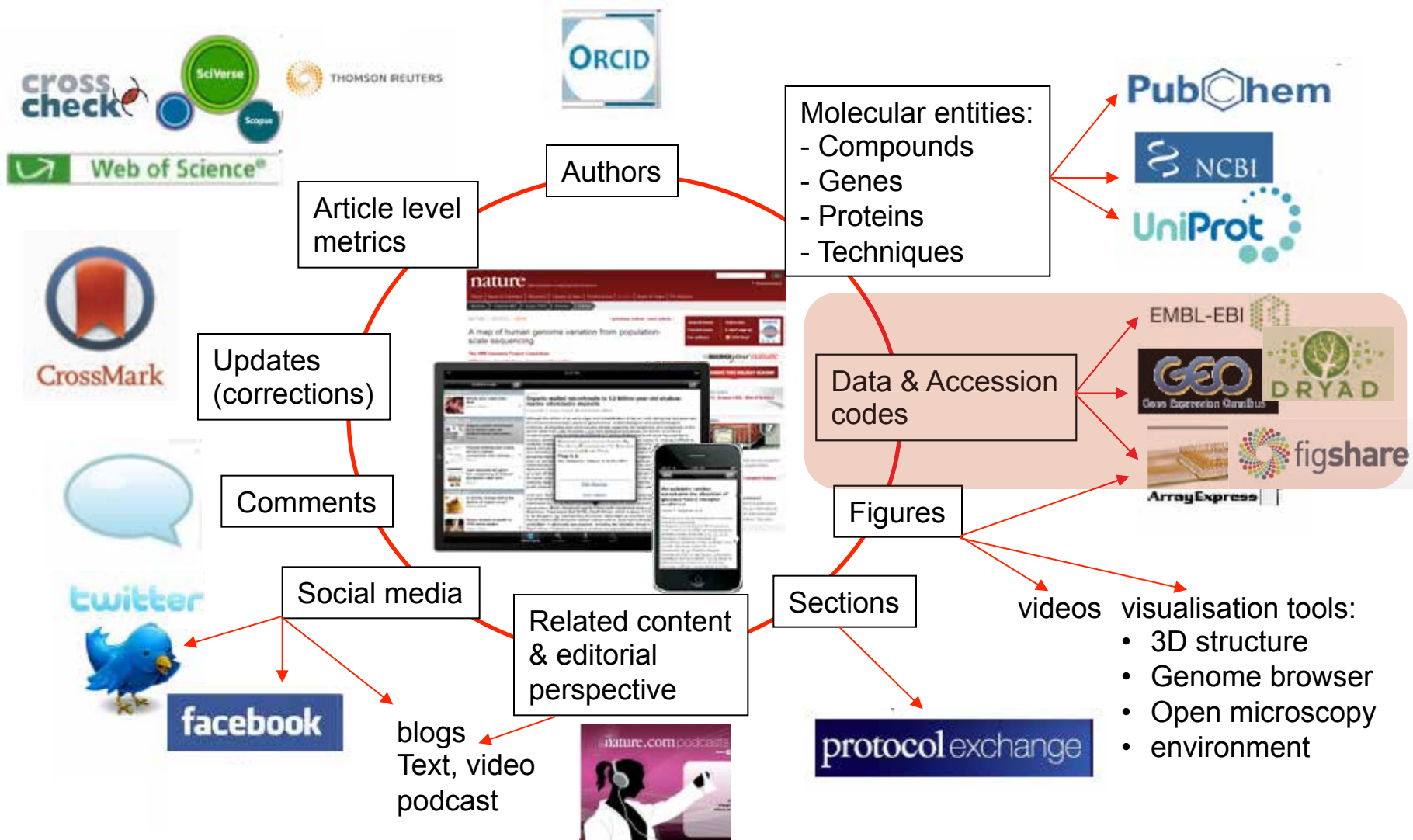


# Evolution of research

Encode Project 2012  
30 papers, 3 Journals

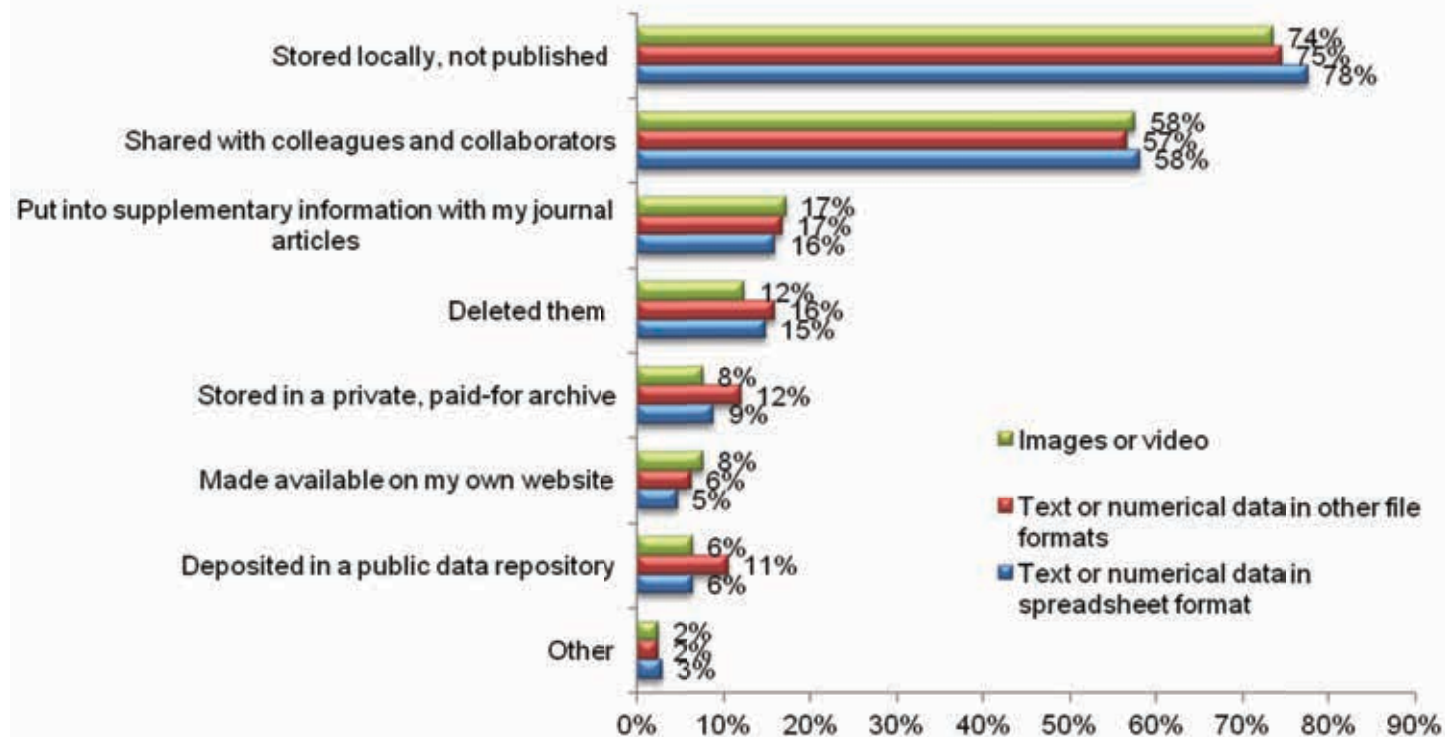


# The article of the future...



# 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

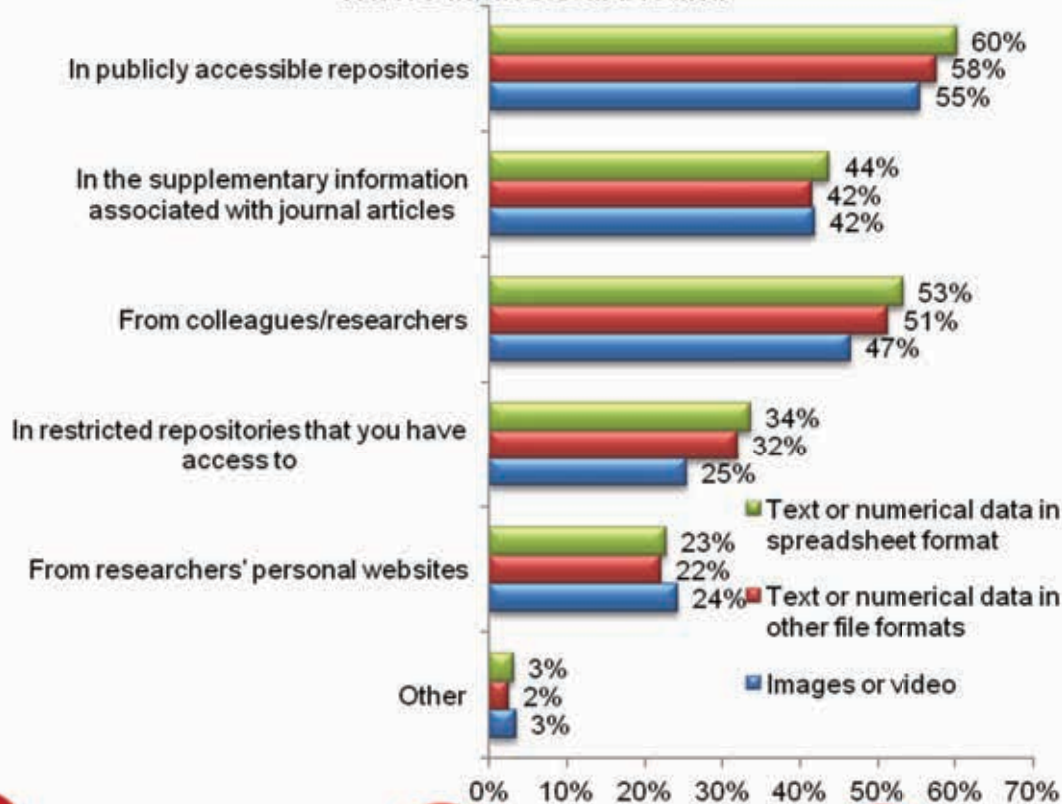




# What do authors and readers say...

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

Where datasets are found



## 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

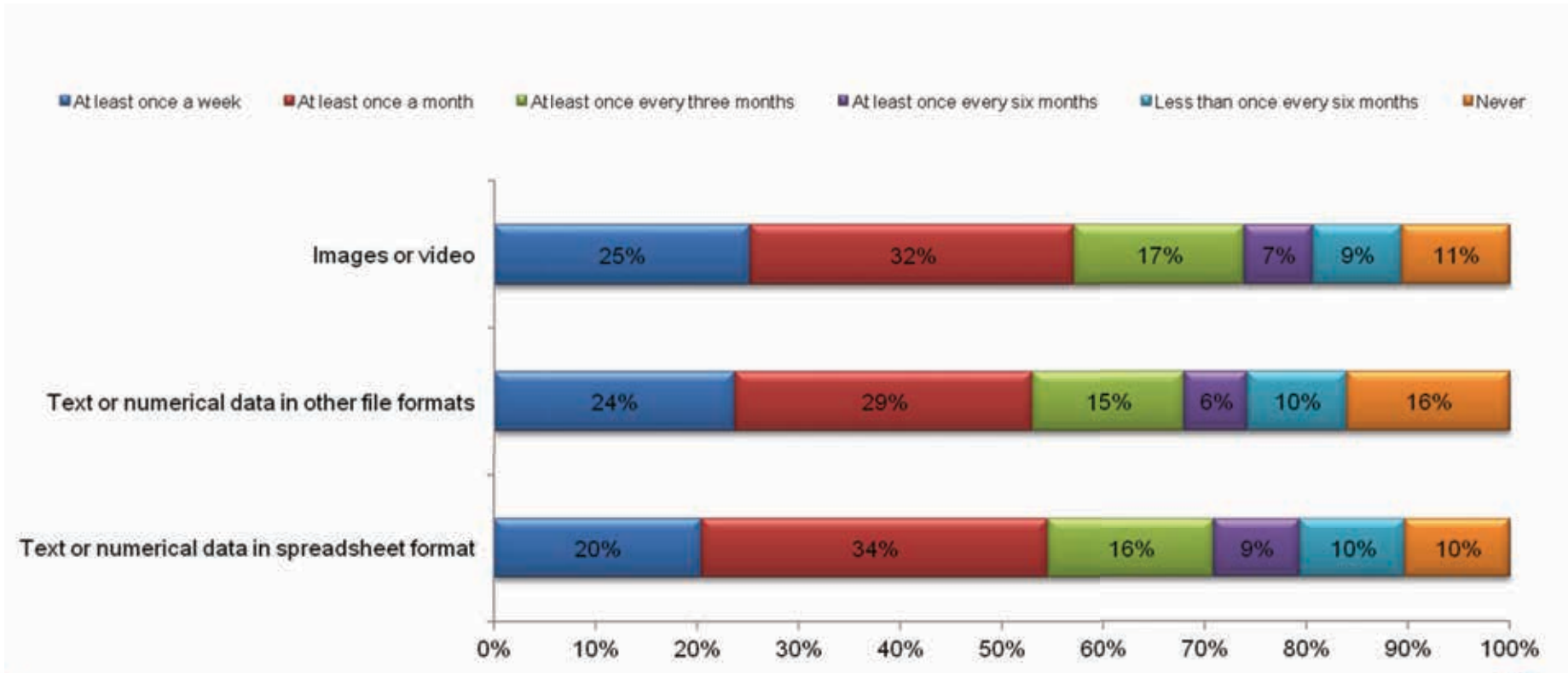
common areas in the disk space

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

[Home](#) | [About](#) | [For Authors](#) | [Advisory and Editorial Board](#) | [Open Access](#) | [Contacts](#) | [FAQ](#)

Helping you publish, discover,  
and reuse research data

 [Sign up for Scientific Data e-alert](#) 

### [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

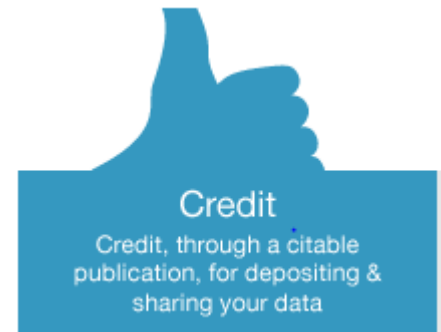
### [Follow Us](#)



 <b>Credit</b> Credit, through a citable publication, for depositing & sharing your data	 <b>Reuse</b> Complete, curated & standardized descriptions enable the reuse of your data	 <b>Quality</b> Rigorous community based peer review
 <b>Discovery</b> Find datasets relevant to your research	 <b>Open</b> Promotes & endorses open science principles & available to all through a Creative Commons license	 <b>Service</b> In-house curation, rapid peer review & publication of your data descriptions

# What is *Scientific Data*?

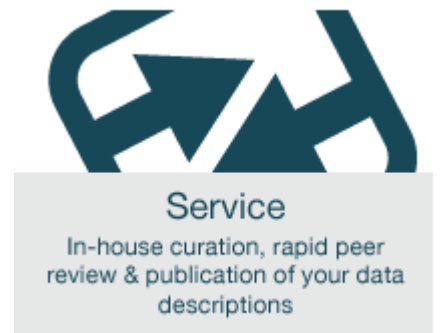
- *Scientific Data* is an Open Access, online-only platform containing data descriptors that describe and explain datasets, supported by an APC model.
- Data descriptors are a **new** type of content and can be viewed as ‘secondary’ material aimed at increasing the visibility and usability of datasets and to aid research reproducibility
  - datasets that relate to a previously published peer-reviewed 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





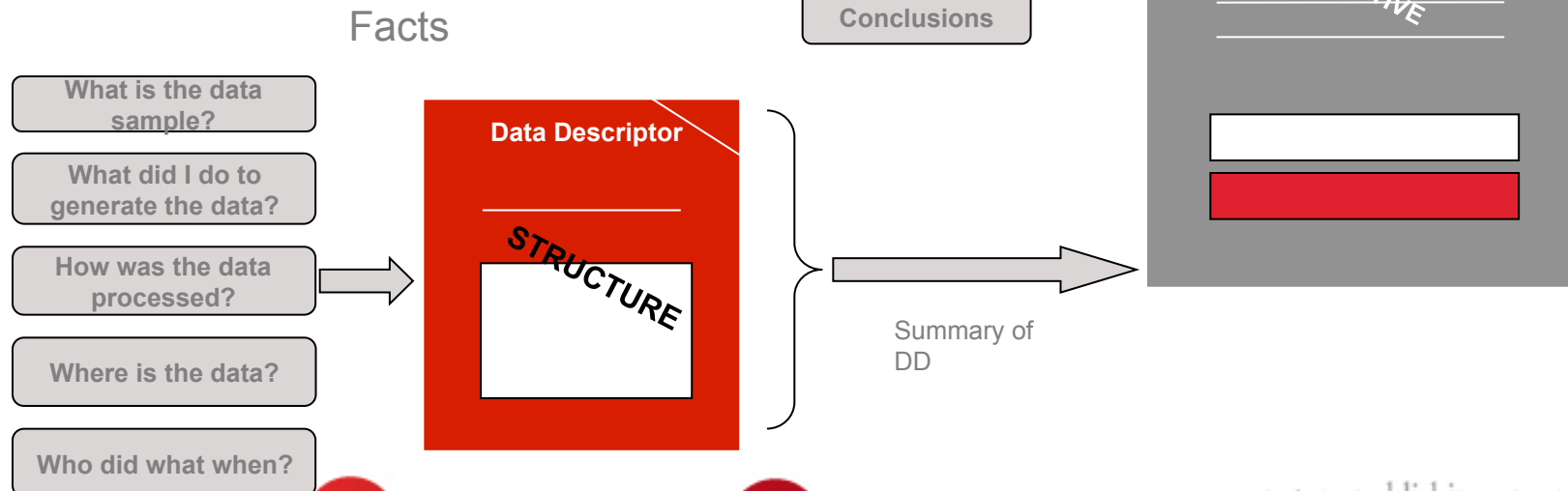
# What is *Scientific Data*..?

- 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
  - submitted at the same time as the journal article
  - submitted after 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 L Hufton** Managing Editor



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.