



**HAL**  
open science

## Trends in scholarly publishing - the impact of information technologies

Odile Hologne

► **To cite this version:**

Odile Hologne. Trends in scholarly publishing - the impact of information technologies. Meeting of the editors of "Annals of forest science", Jul 2015, Nancy, France. hal-02799198

**HAL Id: hal-02799198**

**<https://hal.inrae.fr/hal-02799198v1>**

Submitted on 5 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Trends in scholarly publishing

The impact of information technologies

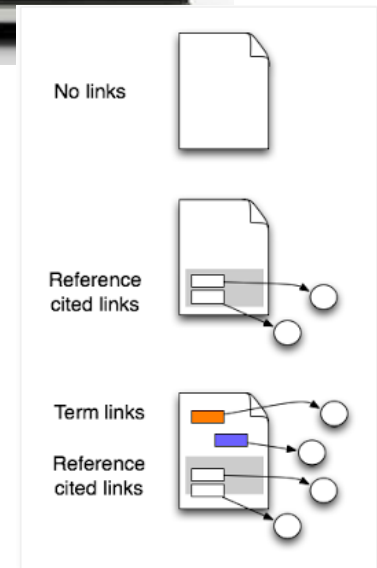
Odile Hologne, INRA, Head of the department of scientific information

[Odile.hologne@versailles.inra.fr](mailto:Odile.hologne@versailles.inra.fr)

Twitter : Holo\_08



# The end of Gutenberg era ...



<http://iphylo.blogspot.fr/2009/04/semantic-publishing-towards-real.html>

# A Quick Tour with in 6 questions

- ❖ What means « open » ?
- ❖ What is a journal ?
- ❖ What is an article ?
- ❖ What can we measure ?
- ❖ How to control text mining ?
- ❖ How to link data and articles ?
- ❖ Conclusion – What else ?



# Open access

But what means « open » ?

# Introduction

- ❖ What means « open »
  - ✓ put free PDF on line ?
  - ✓ Business models or APC
- ❖ We forget other aspects of openness
  - ✓ technical aspects
  - ✓ legal aspects

# Journals : How open is it ?

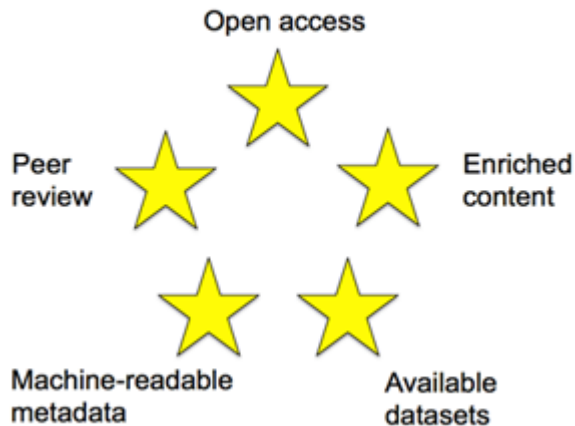
| Access        | Reader Rights   | Reuse Rights  | Copyrights   | Author Posting Rights   | Automatic Posting  | Machine Readability  | Access        |
|---------------|---|---|--|---|--|--|---------------|
| OPEN ACCESS   | Free readership rights to all articles immediately upon publication                             | Generous reuse & remixing rights (e.g., CC BY license)  | Author holds copyright with no restrictions  | Author may post any version to any repository or website  | Journals make copies of articles automatically available in trusted third-party repositories (e.g., PubMed Central) immediately upon publication | Article full text, metadata, citations, & data, including supplementary data, provided in community machine-readable standard formats through a community standard API or protocol | OPEN ACCESS   |
|               | Free readership rights to all articles after an embargo of no more than 6 months                | Reuse, remixing, & further building upon the work subject to certain restrictions & conditions (e.g., CC BY-NC & CC BY-SA licenses) | Author holds copyright, with some restrictions on author reuse of published version              | Author may post final version of the peer-reviewed manuscript ("postprint") to any repository or website        | Journals make copies of articles automatically available in trusted third-party repositories (e.g., PubMed Central) within 6 months              | Article full text, metadata, citations, & data, including supplementary data, may be crawled or accessed through a community standard API or protocol                              |               |
|               | Free readership rights to all articles after an embargo greater than 6 months                   | Reuse (no remixing or further building upon the work) subject to certain restrictions and conditions (e.g., CC BY-ND license)       | Publisher holds copyright, with some allowances for author and reader reuse of published version | Author may post final version of the peer-reviewed manuscript ("postprint") to certain repositories or websites | Journals make copies of articles automatically available in trusted third-party repositories (e.g., PubMed Central) within 12 months             | Article full text, metadata, & citations may be crawled or accessed without special permission or registration   |               |
|               | Free and immediate readership rights to some, but not all, articles (including "hybrid" models) | _____   | Publisher holds copyright, with some allowances for author reuse of published version            | Author may post submitted version/draft of final work ("preprint") to certain repositories or websites          | _____  | Article full text, metadata, & citations may be crawled or accessed with permission  |               |
| CLOSED ACCESS | Subscription, membership, pay-per-view, or other fees required to read all articles             | No reuse rights beyond fair use/ limitations & exceptions to copyright (all rights reserved copyright) to read                      | Publisher holds copyright, with no author reuse of published version beyond fair use             | Author may not deposit any versions to repositories or websites   | No automatic posting in third-party repositories   | Article full text & metadata not available in machine-readable format  | CLOSED ACCESS |

\*HowOpenIsIt™ Open Access spectrum\*, © 2013 SPARC and PLOS, licensed under CC BY

[http://sparc.arl.org/sites/default/files/hoii\\_guide\\_rev4\\_web.pdf](http://sparc.arl.org/sites/default/files/hoii_guide_rev4_web.pdf) (2013)

# Articles : openness and quality

## The Five Stars of Online Journal Articles



The Five Stars of Online Journal Articles — a Framework for Article Evaluation  
David Shotton  
Department of Zoology, University of Oxford  
david.shotton@zoo.ox.ac.uk  
doi:10.1045/january2012-shotton

- ❖ **Peer review**  
Ensure your article is peer reviewed, to provide assurance of its scholarly value, quality and integrity.
- ❖ **Open access**  
Ensure others have cost-free open access both to read and to reuse your published article, to ensure its greatest possible readership and usefulness.
- ❖ **Enriched content**  
Use the full potential of Web technologies and Web standards to provide interactivity and semantic enrichment to the content of your online article.
- ❖ **Available datasets**  
Ensure that all the data supporting the results you report are published under an open license, with sufficient metadata to enable their re-interpretation and reuse.
- ❖ **Machine-readable metadata**  
Publish machine-readable metadata describing both your article and your cited references, so that these descriptions can be discovered and reused automatically.
- ❖ Ex : cardioshare <http://biordf.net/cardioSHARE/queries.html>





# OA to publications mandate in H2020

**Each beneficiary must ensure OA to all peer-reviewed scientific publications relating to its results:**

- **Deposit** a machine-readable copy of the published version or final peer-reviewed manuscript accepted for publication in a repository of the researchers choice (possibly OpenAIRE compliant)
- **Ensure OA** on publication or at the latest within 6 months (12 for SSH)
- **Aim to deposit** at the same time **the research data needed to validate the results ("underlying data")**
- Ensure **OA to the bibliographic metadata** that identify the deposited publication, via the repository

Celina Ramjoue (Head of Sector "Open Access to scientific Publications and Data", EC DG CNECT)

Agenda

# Data publication

Linked Data ou  
publication  
Web de  
données

Données sur le Web, dans un format structuré et non propriétaire, données du fichier identifiées par des URIs et reliées à d'autres



Données sur le Web, dans un format structuré et non propriétaire, données du fichier identifiées par des URIs



Publication  
Web

Données sur le Web, identifié ou pas, dans un format structuré et non propriétaire

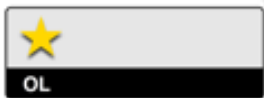


`http://data...`

Fichier sur le Web, identifié ou pas dans un format structuré



Fichier sur le Web, identifié ou pas, quel que soit le format



from <http://5stardata.info/>



# What is a journal ?

From a technical point of  
view

# From dissemination to data processing

## ❖ Old

- ✓ dissemination
- ✓ to be readable by human

## ❖ Now

- ✓ metadata processing
- ✓ data access and data processing
- ✓ to be readable by computers

# A platform – a software

5

BioMed Central  
The Open Access Publisher

Springer

# (GIGA)<sup>n</sup> SCIENCE

Journal, data-platform and database for large-scale data

华大基因  
BGI

in conjunction with

BioMed Central  
The Open Access Publisher

Open API to retrieve information from (GIGA)<sup>n</sup> DB

Open-Paper



DOI:10.1186/2047-217X-1-18

Highly accessed >11000 accesses

Linked to  
DOI

(GIGA)<sup>n</sup> DB  
Data sets

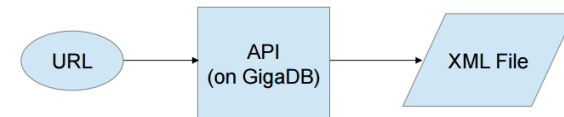
Open-Data  
DOI:10.5524/100038

78GB CC0 data

Linked to  
DOI

(GIGA)<sup>n</sup> Galaxy  
by CBIIT  
Analyses

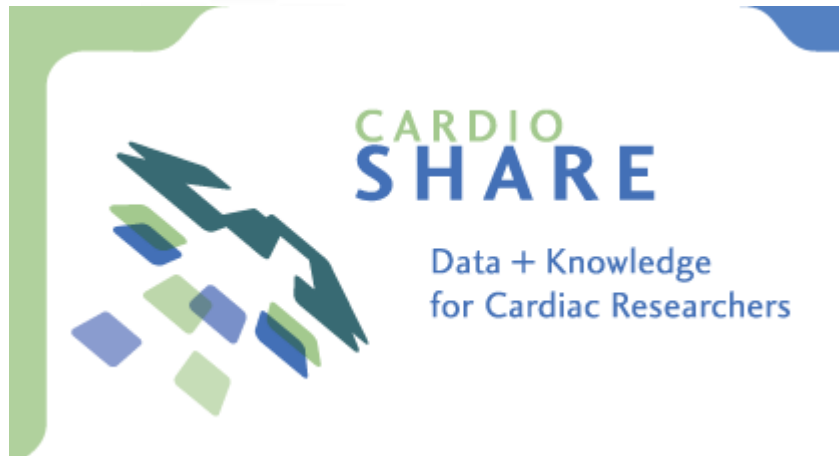
Open-Pipelines  
Open-Workflows  
DOI:10.5524/100044



- Only journal where data and source code behind article + article can be mined through open API
- Only journal offering a home for complex image data (like fMRIs, eg) right next to article

Amye Kenall [http://www.gfii.fr/uploads/docs/GFII\\_Springer.pdf](http://www.gfii.fr/uploads/docs/GFII_Springer.pdf)

# Be able to do queries on articles and data



*What proteins is Pubmed article 9207092 written about? What organisms do these proteins belong to?*

```
PREFIX sad: <http://sadiframework.org/ontologies/properties.owl#>
PREFIX ss: <http://semanticscience.org/resource/>
PREFIX pubmed: <http://lsrn.org/PMID:>
SELECT ?protein ?organismName
WHERE {
    pubmed:9207092 ss:SIO_000252 ?protein .
    ?protein sad:fromOrganism ?organism .
    ?organism sad:hasName ?name .
    ?name ss:SIO_000300 ?organismName .
}
```

# Journal and data processing

## An Analysis of the Viola-Jones Face Detection Algorithm

Yi-Qing Wang

article demo archive

published reference • 2014-06-26 -- BibTeX  
• Yi-Qing WANG, *An Analysis of the Viola-Jones Face Detection Algorithm*, Image Processing On Line, 4 (2014), pp. 128–148.  
<http://dx.doi.org/10.5201/ipol.2014.104>

Communicated by Jose-Luis Lisani  
Demo edited by Yi-Qing Wang

### Abstract

In this article, we decipher the Viola-Jones algorithm, the first ever real-time face detection system. There are three ingredients working in concert to enable a fast and accurate detection: the integral image for feature computation, AdaBoost for feature selection and an attentional cascade for efficient computational resource allocation. Here we propose a complete algorithmic description, a learning code and a learned face detector that can be applied to any color image. Since the Viola-Jones algorithm typically gives multiple detections, a post-processing step is also proposed to reduce detection redundancy using a robustness argument.

### Download

- full text manuscript: PDF low-res. (433K) PDF (6M) [?]
- source code: TAR/GZ

### Preview

Loading takes a few seconds. Images and graphics are degraded here for faster rendering. See the downloadable PDF documents for original high-quality versions.

LOW RESOLUTION PDF: Images may show compression artifacts. A full resolution PDF is available at [www.ipol.im](http://www.ipol.im).



Published in Image Processing On Line on 2014-06-26.  
Submitted on 2013-08-31, accepted on 2014-05-09.  
ISSN 2105-1232 © 2014 IPOL & the authors CC-BY-NC-SA  
This article is available online with supplementary materials,  
software, datasets and online demo at  
<http://dx.doi.org/10.5201/ipol.2014.104>

## An Analysis of the Viola-Jones Face Detection Algorithm

Yi-Qing Wang

CMLA, ENS Cachan, France ([yiqing.wang@cmla.ens-cachan.fr](mailto:yiqing.wang@cmla.ens-cachan.fr))

## An Analysis of the Viola-Jones Face Detection Algorithm

article demo archive

Please cite the reference article if you publish results obtained with this online demo.

### Select Data

Click on an image to use it as the algorithm input.



image credits

### Upload Data

Upload your own image files to use as the algorithm input.

input image  Aucun fichier choisi

<http://www.ipol.im/>



# What is an article



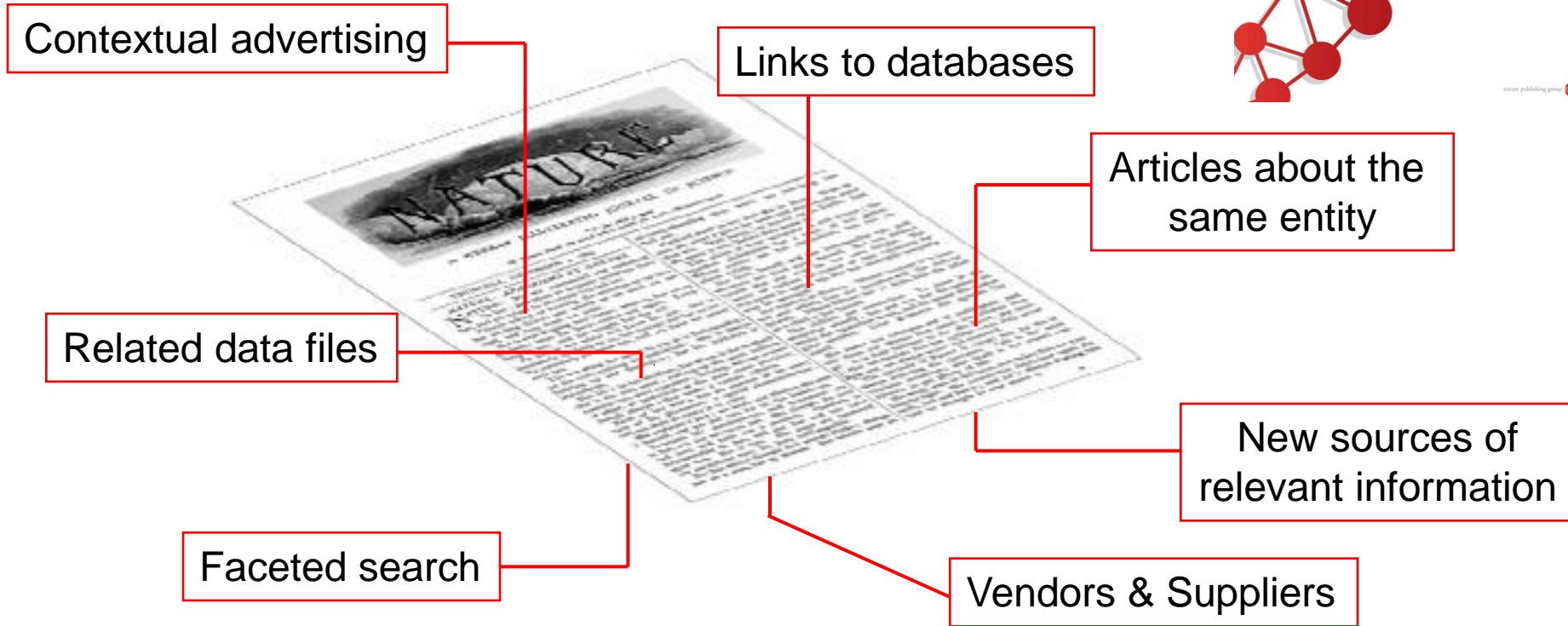
# Evolution of scientific publishing

From articles to particles:  
How Nature's semantically-enriched content  
boosts scientific discovery

Dr Jason Wilde  
Director, Business Development  
Nature Publishing Group  
30<sup>th</sup> November 2011



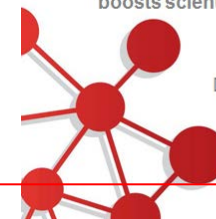
nature publishing group 



Articles need to move from being 2D into a 3D entity in the information space.



# Behind the scenes - interface



**Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm**

Prdm14 is a PR-domain and zinc-finger protein whose expression is restricted to the pluripotent cells of the early embryo, embryonic stem cells (ESCs) and germ cells. Here, we show that Prdm14 safeguards mouse ESC (mESC) maintenance by preventing induction of extraembryonic endoderm (ExEn) fates. Conversely, Prdm14 occupies and represses genomic loci encoding ExEn differentiation factors, while also binding to and promoting expression of genes essential for ESC self-renewal. Prdm14-associated genomic regions substantially overlap those occupied by Nanog and Oct4, are enriched in a chromatin signature associated with pluripotency, and contain a unique DNA-sequence motif recognized by Prdm14 *in vitro*. Our work identifies a new member of the mESC transcriptional network, Prdm14, which plays a dual role as a context-dependent transcriptional repressor or activator.

In the mammalian embryo, each cell of the inner cell mass (ICM) of the blastocyst is destined to choose between two prospective fates: that of the epiblast, which gives rise to all tissues of the embryo proper, and that of the primitive endoderm, which develops into the visceral and parietal extraembryonic endoderm, providing nutrient exchange and inductive signals for the embryo.

Segregation of the epiblast and the primitive endoderm begins in the early ICM, where Nanog and Gata6, master regulators of pluripotency and endoderm formation, respectively, are expressed in a progressively mutually exclusive manner. Indeed, within the early ICM, two cell subpopulations with distinct expression profiles exist: one characterized by the upregulation of certain primitive endoderm genes (for example, Gata4 and Gata6) and the other characterized by expression of certain early epiblast genes (for example, Nanog and Fgf4). The FGF-Grb2-MAPK signaling pathway is important for breaking of the expression symmetry by upregulation of Gata6 and downregulation of Nanog, although even after the initial establishment of this asymmetry, cells of the ICM retain the ability to switch their lineage choice. Eventually, however, Nanog-positive cells form the epiblast, whereas Gata6-positive cells commit to the primitive endoderm fate, induce transcription of specific adhesion molecules such as Lamb1 and Dab2, and migrate to the ICM surface, where they later diversify into parietal and visceral extraembryonic endoderm lineages.

Because of its biochemical, genomic and genetic tractability, the mESC model provides a useful tool for studies of molecular mechanisms governing early cell fate transitions and can be used to investigate the fate choice between the pluripotent epiblast and ExEn. Similar to its function in the transcription factors Gata4, Gata6, Sox7 or Sox17 is sufficient to induce mESC differentiation into ExEn fates, whereas ectopic expression of any of the

In search of new regulators of mESC self-renewal and repression of the ExEn differentiation program, we identified Prdm14, a PR-domain protein family of putative transcriptional regulators, which is characterized by the presence of tandem zinc fingers and a PR (PRD1-BF1 and RIZ) domain. The PR domain is a sequence-dependent transcriptional repressor and a potent regulator of cell fate decisions in various tissues. Moreover, the critical domain for Prdm14 function is the PR domain. Prdm14 is a sequence-dependent transcriptional repressor and a potent regulator of cell fate decisions in various tissues. Moreover, the critical domain for Prdm14 function is the PR domain. Prdm14 is a sequence-dependent transcriptional repressor and a potent regulator of cell fate decisions in various tissues. Moreover, the critical domain for Prdm14 function is the PR domain.

Here, we demonstrate that Prdm14 is a sequence-dependent transcriptional regulator that protects mESCs from entering ExEn fates. Using a combination of cell biology, genomic and biochemical approaches, we identified Prdm14-regulated genes, Prdm14's genomic binding patterns and its unique DNA-sequence specificity.

## Results

### Differentiation to ExEn fates upon Prdm14 knockdown

To examine whether Prdm14 contributes to the maintenance of the mESC state, we treated mESCs with Prdm14-targeting short interfering RNA (siRNA), leading to a partial Prdm14 mRNA and protein depletion (Supplementary Fig. 1a), and cultured cells under conditions compatible with self-renewal. Within 3 d of transfection, Prdm14 siRNA-treated cells were morphologically indistinguishable from cells transfected with control nontargeting siRNA or from untreated, wild-type (WT) mESCs (not shown). However, 5–7 d after initial treatment and specifically in the Prdm14 siRNA-transfected samples, we noted the appearance of dispersed refractive cells, which morphologically resembled previously described ExEn cells (Fig. 1a) and which lost alkaline phosphatase and Nanog expression (Fig. 1b,c). Consistent with ExEn differentiation, the vast majority—if not all—of these dispersed cells stained positively for ExEn markers, such as the transcription factor Gata4 and surface molecules Lamb1 and Dab2 (Fig. 1c; Supplementary Fig. 1b). The described knockdown phenotype was recapitulated with an independent, non-overlapping Prdm14 siRNA (Supplementary Fig. 2). In contrast, control siRNA-treated cells did not exhibit differentiated morphology or ExEn marker staining (Fig. 1a–c; Supplementary Fig. 1b; Supplementary Fig. 2). Of note, a substantial subpopulation of cells treated with Prdm14 siRNA retained ESC morphology and Nanog expression and stained negatively for Gata4, Lamb1 and Dab2 (Fig. 1 and Supplementary Fig. 2). This suggests that a partial knockdown of Prdm14 is insufficient to abrogate self-renewal in some cells and/or a cell subset that retains ESC characteristics corresponds to the cells that remained untransfected. Although, given those caveats, we could not determine whether Prdm14 is absolutely required for self-renewal, our data suggest that Prdm14 protects mESCs from spontaneous differentiation to ExEn fates.

### Prdm14 depletion leads to the induction of ExEn genes

Gene expression profiling by RNA-seq identified 480 upregulated and 173 downregulated genes in the Prdm14 knockdown sample, as compared to the control siRNA-treated sample (at 1.4-fold change cutoff and criteria representing statistical significance, false discovery rate (FDR) < 0.0001; genes listed in Supplementary Data 1). Examination of upregulated genes showed a striking

Curated with  
all metadata to  
be exported

Identified as a  
Biological entity  
and editable

|               |                        |
|---------------|------------------------|
| Title         | Homeobox protein NANOG |
| Protein Name  | Homeobox protein NANOG |
| Gene Symbol   | Nanog                  |
| UniProt ID    | Q80764                 |
| EntrezGene ID | 71950                  |
| Organism      | Mus musculus           |

| Name  | UniProt ID | EntrezGene ID | Gene Symbol |
|-------|------------|---------------|-------------|
| Nanog |            |               |             |

Search: Mus musculus | Wikipedia | Google | DBGET | SGD

Results for name: Nanog

|              |         |                |         |         |
|--------------|---------|----------------|---------|---------|
| UniProt (20) | See all | EntrezGene (5) | See all | Compare |
|--------------|---------|----------------|---------|---------|

**Protein Name** Homeobox protein NANOG  
**UniProt ID** Q80764  
**Reviewed** Yes  
**Mnemonic** NANOG\_MOUSE  
**Organism** Mus musculus  
**Genes** Nanog, Ecata4, Enk  
**Synonyms** Homeobox protein NANOG, ES cell-associated protein 4, Early embryo (4 synonyms)...

Choose Entity for selected

**Protein Name** Homeobox protein NANOG  
**UniProt ID** Q5TMB3  
**Reviewed** Yes  
**Mnemonic** NANOG\_MOUSE  
**Organism** Mus musculus molossinus  
**Genes** Nanog, Stm1  
**Synonyms** Homeobox protein NANOG, Homeobox transcription factor Nanog (2 synonyms)...

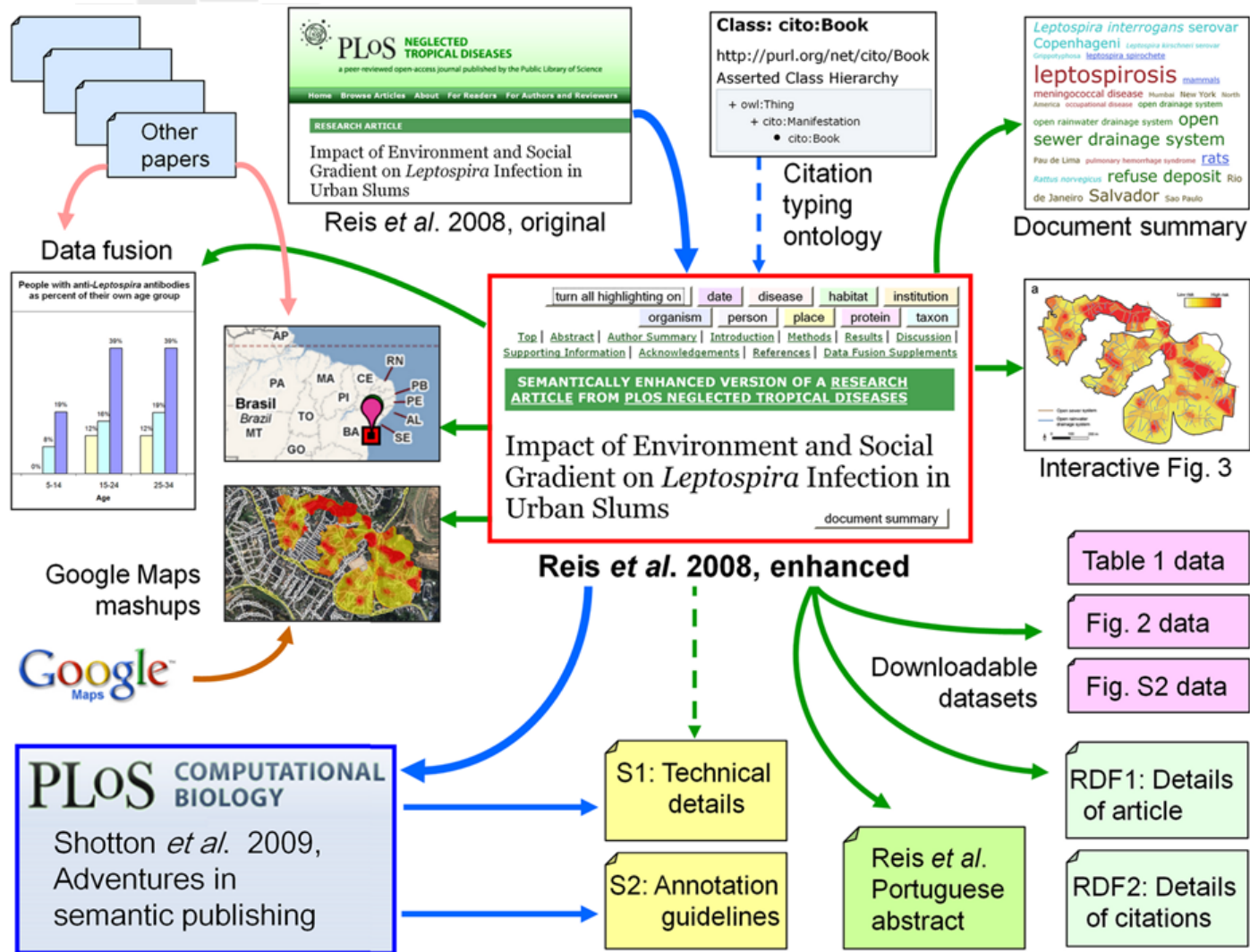
Choose Entity for selected

**Protein Name** Nanog protein  
**UniProt ID** E7ZN41  
**Reviewed** No  
**Mnemonic** B7ZN41\_MOUSE  
**Organism** Mus musculus  
**Genes** Nanog  
**Synonyms** Nanog protein (1 synonym)...

Choose Entity for selected

**Protein Name** MCG132219  
**UniProt ID** A2RS90  
**Reviewed** No  
**Mnemonic** A2RS90\_MOUSE  
**Organism** Mus musculus  
**Genes** Nanog, MCG\_132219  
**Synonyms** MCG132219, Nanog homeobox (2 synonyms)...

Choose Entity for selected



<http://iphylo.blogspot.fr/2009/04/semantic-publishing-towards-real.html>

# Negative results

Log on  Journals Gateways

 JOURNAL OF NEGATIVE RESULTS IN BIOMEDICINE

Search  for   [Advanced search](#)

Home Articles Authors Reviewers About this journal My JNRBM

**General information**  
Editorial Board  
FAQ  
Article processing charge FAQ  
Article metrics FAQ  
Contact us  
Privacy policy  
Promote

## About *Journal of Negative Results in Biomedicine*

[Aims and scope](#) | [Open access](#) | [Article-processing charges](#) | [Indexing services](#) | [Publication and peer review process](#) | [Editorial policies](#) | [Citing articles in \*Journal of Negative Results in Biomedicine\*](#) | [Why publish your article in \*Journal of Negative Results in Biomedicine\*?](#)

This page includes information about the aims and scope of *Journal of Negative Results in Biomedicine*, editorial policies, open access and article-processing charges, the peer review process and other information. For details of how to prepare and submit a manuscript through the [online submission system](#), please see the [instructions for authors](#).

### Aims & scope

*Journal of Negative Results in BioMedicine* is an open access, peer-reviewed, online journal that provides a platform for the publication and discussion of unexpected, controversial, provocative and/or negative results in the context of current tenets.

*Journal of Negative Results in BioMedicine* aims to encourage scientists and physicians of all fields to publish results that challenge current models, tenets or dogmas. The journal invites scientists and physicians to submit work that illustrates how commonly used methods and techniques are unsuitable for studying a particular phenomenon. *Journal of Negative Results in BioMedicine* strongly promotes and invites the publication of clinical trials that fall short of demonstrating an improvement over current treatments. The aim of the journal is to provide scientists and physicians with responsible and balanced information in order to improve experimental designs and clinical decisions.

Articles published in traditional journals frequently provide insufficient evidence regarding negative data. They hardly allow a rigorous evaluation of the quality of these results. In addition, controversial results that refute a current model or simply negative results within a current dogma, frequently meet considerable resistance before they are acknowledged. This is particularly the case if current techniques and technologies are too crude to shed further light on the findings. As more sophisticated techniques become available such findings may turn out to have been groundbreaking only decades later.



Search information pages

ADVERTISEMENT

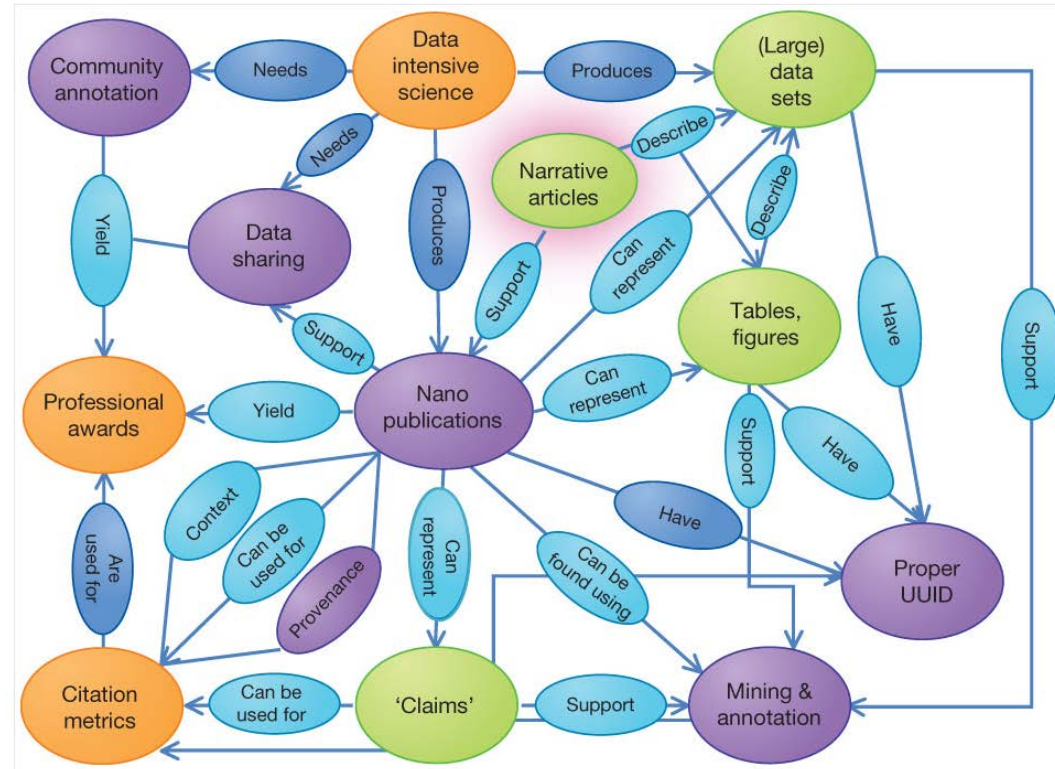


# Why do we need literature ? Nanopublication



Figure 2: A proposal for the future of scholarly communication.

From  
The value of data  
Barend Mons, Herman van Haagen, Christine Chichester, Peter-Bram 't Hoen, Johan T den Dunnen, Gertjan van Ommen, Erik van Mulligen, Bharat Singh, Rob Hoof, Marco Roos, Joel Hammond, Bruce Kiesel, Belinda Giardine, Jan Velterop, Paul Groth & Erik Schultes  
*Nature Genetics* 43, 281–283 (2011) | doi:10.1038/ng0411-281  
Published online 29 March 2011



nature  
genetics

Home | Current issue | Comment | Research | Archive | Authors & referees | About the journal

home > archive > issue > commentary > full text

NATURE GENETICS | COMMENTARY

The value of data

Barend Mons, Herman van Haagen, Christine Chichester, Peter-Bram 't Hoen, Johan T den Dunnen, Gertjan van Ommen, Erik van Mulligen, Bharat Singh, Rob Hoof, Marco Roos, Joel Hammond, Bruce Kiesel, Belinda Giardine, Jan Velterop, Paul Groth & Erik Schultes

Affiliations | Contributions | Corresponding author

*Nature Genetics* 43, 281–283 (2011) | doi:10.1038/ng0411-281  
Published online 29 March 2011

<http://dx.doi.org/10.1038/ng0411-281>



# What can we measure ?

## Metrics

# What does DORA say?

DORA makes one general and 17 specific recommendations.

## General recommendation:

Do not use journal-based metrics, such as Journal Impact Factors (JIFs), as surrogate measures of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions.

### For Organizations That Supply Metrics

- Be transparent
- Provide access to data
- Discourage data manipulation
- Provide different metrics for primary literature and reviews

### For Publishers

- Cease to promote journals by Impact Factor; provide an array of metrics
- Focus on article-level metrics
- Identify different author contributions
- Open the bibliographic citation data
- Encourage primary literature citations

### For Research Institutions

- When hiring and promoting, state that scientific content of a paper, not the JIF of the journal where it was published, is what matters
- Consider value from all outputs and outcomes generated by research

### For Funding Agencies

- State that scientific content of a paper, not the JIF of the journal where it was published, is what matters
- Consider value from all outputs and outcomes generated by research

### For Researchers

- Focus on content
- Cite primary literature
- Use a range of metrics to show the impact of your work
- Change the culture!

San Francisco  
**DORA**  
Declaration on Research Assessment



# Evaluation

What kind of new metrics ?



## Métriques

Journal : Ecography (2013)

Facteur d'impact à 2 ans : 4,207

Facteur d'impact à 5 ans : 5,776

Notoriété à 2 ans : Excellente (biodivers.conserv. ; ecology)

### Article

Nb de consultations de cette notice : 80

Web of Science® Times Cited : 181

Altmetric 6

Tweeted by 11  
On 1 Facebook pages

See more details

429 readers on Mendeley  
4 readers on CiteULike

# Next step : over the audience

## ❖ Need to qualify the citation

| Citation Function          | Description                                     |
|----------------------------|---|
| Based_on <sup>+</sup>      | A work is based on the cited work               |
| Corroboration <sup>+</sup> | Two works corroborate each other                |
| Discover <sup>+</sup>      | Acknowledge the invention of a technique        |
| Positive <sup>+</sup>      | The cited work is successful                    |
| Practical <sup>+</sup>     | The cited work has a practical use              |
| Significant <sup>+</sup>   | The cited work is important                     |
| Standard <sup>+</sup>      | The cited work is a standard                    |
| Supply <sup>+</sup>        | Acknowledge the supplier of a material          |
| Contrast <sup>=</sup>      | Compares two works in a neutral way             |
| Co-citation <sup>=</sup>   | Citations that appear closely                   |
| Neutral <sup>=</sup>       | The cited work not belonging to other functions |
| Negative <sup>-</sup>      | The weakness of the cited work is discussed     |

Table 1: Annotation Scheme for Citation Function: <sup>+</sup> represents POSITIVE sentiment, <sup>=</sup> represents NEUTRAL sentiment, and <sup>-</sup> represents negative sentiment

<http://www.aclweb.org/anthology/R13-1052>





# How to change peer reviewing ?

# Peer reviewing

Debats on research integrity,  
transparence

- ❖ Review Post publication
- ❖ The end of anonymous PR?
- ❖ Review citable
- ❖ New Players

PubPeer > Genet. Mol. Res.

## "Screening relevant genes of tolerance to low phosphorus in maize using cDNA-amplified fragment length polymorphism"

H.Y. Jiang, Z. Li, J. Zhao, Q. Ma, B.J. Cheng, S.W. Zhu, Genet. Mol. Res. (2015)

Comments (0):

Enter new comment below (Please read the [How To](#))

Post New

→ Go to Article PDF

★ Get alerts for new activity

✉ Invite others to the conversation

## Get Credit for Peer Review

Publons allows you to record, verify, and showcase your peer review contributions in a format you can include in job and funding applications (without breaking reviewer anonymity).

- Get recognition for all the peer review work you do for journals.
- See how you compare with our review stats and graphs.
- Help your university climb the university leaderboard.
- Simply forward your 'thank you for reviewing' emails to reviews@publons.com to add reviews.

40,019

Reviewers

128,382

Reviews


8,690


Journals


Read a recent feature of Publons on **nature**:


“ *The scientists who get credit for peer review* ”

PUBLONS IN THE PRESS

 Sign in using Google

 Sign in using Mendeley

 Sign in using ORCID

 Sign in using LinkedIn

Or create a new account

Publishing Name

Email

Password

CREATE ACCOUNT

By registering you are agreeing to our terms of service and privacy policy.



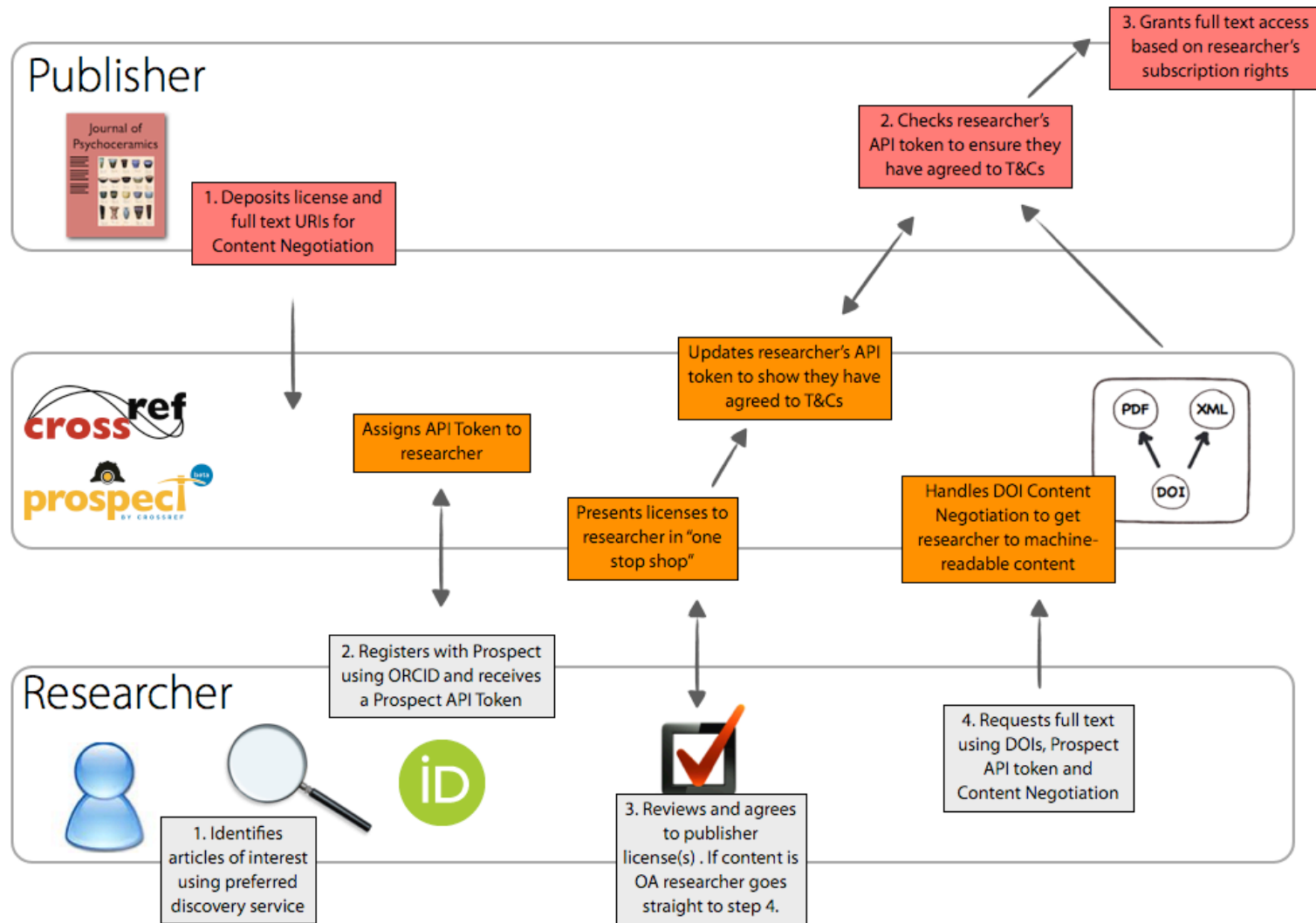
# Text mining

## How to control it

# Challenges

- ❖ The need : Textual big data : we can't read everything
- ❖ To extract information from unstructured text
- ❖ Without having nothing more to pay to the publishers
  - ✓ a « natural right » to do it
  - ✓ download without restriction

# But ... They organize the control ...





# How to link articles and data ?

# Data and journals

❖ Data linked to articles in « old journals »

❖ Data journals

The screenshot shows a journal article page from *Chemical Geology*, Volume 273, Issues 3-4, 15 May 2010, Pages 151-167. The article title is "An ancient metasomatic source for the Walvis Ridge basalts" by Vincent J.M. Salters and A.S. Sach-Kocher. The abstract discusses the proposed origins for the Enriched Mantle 1 (EM1) component, suggesting it is a pure sediment or enriched melt from the subcontinental lithosphere. It mentions that the data is consistent with the EM1 composition being formed by the addition of a melt to a mantle with bulk Earth-like composition, followed by melt extraction of a low degree melt. The article also notes that the Walvis Ridge data shows two distinct two-component mixing trends. A map of the Walvis Ridge region is included, showing the location of sites S27, S28, and S29. The article is available in PDF format and is part of the Elsevier collection.

## Data journals, data articles

A growing list:

<http://proj.bank.ac.uk/prepare/blog/DataJournalsList>

A data paper describes a dataset, giving details of its collection, processing, software, file formats etc, without the requirement of novel analyses or ground breaking conclusions. It allows the reader to understand the when, how and why data was collected and what the data-product is

The infographic highlights the benefits of data journals. It features a central image of a book titled "Publishing Data Journals". The text states: "Resolving the publishing bottleneck for biodiversity" and quotes Peter Norvig: "Science is a combination of gathering facts and making theories; neither can progress on its own. In the history of science, the laborious accumulation of facts is the dominant mode, not a novelty." The infographic lists three key benefits:

- Making "small" data big:** No lower/upper limit of manuscript size, Publish all kinds of biodiversity-related data, Reduced page charges affordable to all.
- More than just data journals!** Integrated text and data publishing, Complete online reviews and editing, Community ownership of data, Free of charge in launch phase.
- Community peer-review:** 7 weeks from submission to decision, 3 days from acceptance to publication, Public peer-review on author's choice.

The infographic is associated with INRA (Science & Impact) and includes a navigation bar with links for "Home", "About", "Policies", "Less", "Discover", "Data", "Microsystems", "Data Publications", "Society", and "November 2014".

❖ Data must be in data repositories (avoid supplementary materials)



# But which one ?

- ❖ Only one (linked with the publication workflow)
  - ✓ linked to the journal (ex : Giga Science)
  - ✓ Outside the journal
- ❖ Many ?
  - ✓ Is there any quality standards ?

# How to assess a data repository ?

SCIENTIFIC DATA

To assist our authors, *Scientific Data* maintains a list of repositories that are compatible with our [data deposition policies](#)<sup>1</sup>. Please provide the following information to help us evaluate the suitability of your repository for inclusion in this list<sup>2</sup>:

## Basic details

1. Repository Name:
2. Repository URL:
3. If indexed by BioSharing<sup>3</sup> provide the ID or the URL for your database record:
4. If indexed by re3data<sup>4</sup> provide the ID or the URL for your database record:
5. How large is your current user base?
6. How many datasets are currently hosted by the repository?
7. How long has your resource been available to the community?
8. Do you accept data submissions from the entire scientific community? If data submission is restricted to certain groups, please describe who may deposit data.

## File limitations

9. What type of experimental data can be hosted by the repository? If the repository only accepts specific file formats please state what these are.
10. What is the maximum file size that can be handled by the repository?
11. Are there any limitations to the amount of data that an individual is able to upload?

## Licensing

12. Are there any terms of use or registration processes that data users must agree to, prior to gaining access to the data? Under what licence is data access provided? If you do not provide free and open access to the data hosted, please state the reason for this.
13. Can data depositors choose which licence they wish to use for their particular dataset?

## Data persistency and stability of access

14. What type of identifier is assigned to hosted datasets? N.B. We strongly encourage new repositories to mint DataCite DOIs.
15. Are researchers able to modify or remove datasets after publication?
16. Is there a versioning system in place?
17. Do you guarantee persistent access to datasets, and for how long?

## Costs

18. What is the cost to researchers wishing to host their research data?
19. What are the current and long-term cost recovery (sustainability) plans?

Questionnaire to assist with Scientific Data repository evaluation  
Document Last Updated: April 2015

Page 1

This work is licensed under the Creative Commons Attribution 4.0 International License.  
To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

SCIENTIFIC DATA

20. If any proportion of the running costs are currently met by an academic grant, what is the plan for maintaining access to hosted data should funding be withdrawn in the future?

## Peer review

21. Are you able to facilitate confidential peer review of hosted datasets by the scientific peer reviewers invited to review the data by *Scientific Data*? N.B. *Scientific Data* peer reviewers must be able to access datasets in a manner in which they remain anonymous to repository managers, as well as repository users.
22. If possible please provide examples of currently hosted research datasets, with associated peer-reviewed publications

## Curation and metadata

23. Is there any curation support for researchers uploading their datasets?
24. Do you capture any metadata about hosted datasets in a standardised form?

## Notes

<sup>1</sup>This list is not intended to be a comprehensive list of research data resources, and should not be considered as a substitute for other data repository accreditation efforts.

<sup>2</sup>The aim of this evaluation process is to identify those repositories which are able to serve the wider community, and so are appropriate for *Scientific Data* to recommend to authors. We are glad to consider descriptions of data stored in project-specific or community-specific repositories on a case-by-case basis, even if the host repository is not selected for listing on our recommended repository list. Therefore please be aware that a decision not to list a repository as recommended should not be interpreted as a comment on the quality and utility of a repository for the immediate community it serves.

<sup>3</sup>BioSharing (<http://www.biosharing.org>) is a repository indexing service for life science, environmental and biomedical databases, whilst re3data (<http://www.re3data.org>) indexes data repositories in all other fields.

Questionnaire to assist with Scientific Data repository evaluation  
Document Last Updated: April 2015

Page 2

This work is licensed under the Creative Commons Attribution 4.0 International License.  
To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

<http://www.nature.com/sdata/data-policies#repo-suggest>

# The list approved by Scientific data

The screenshot shows a Figshare article page. At the top, there is a search bar and navigation links for 'Browse' and 'Upload'. The article title is 'Scientific Data recommended repositories\_June 2015'. The main content is a table with 9 rows, listing various data repositories. To the right of the table, there are statistics: 464 views, 3 shares, and a citation count. Below the statistics, there is a 'Published on' date and time, and a file size of 10.58 KB. The 'Categories' section includes 'Science Policy'. The 'Authors' section lists 'Scientific Data'. The 'Tags' section includes 'data repositories' and 'data sharing'. The 'License' section is 'CC-BY'. The 'Export' section includes 'Export to RefWorks' and 'Export to BibTeX'. At the bottom, there is a 'Share this:' section with social media icons for Facebook, Twitter, and Google+, and an 'Embed\*' button. Below that, there is a 'Cite this:' section with the citation text and a DOI link.

|   | A                                    |
|---|--------------------------------------|
| 1 | Repository Name                      |
| 2 | DNA DataBank of Japan                |
| 3 | EMBL Nucleotide Sequence Database    |
| 4 | GenBank                              |
| 5 | dbSNP                                |
| 6 | European Variation Archive           |
| 7 | dbVar                                |
| 8 | Database of Genomic Variants Archive |
| 9 | FBI Metagenomics                     |

SciData repos list\_June 2015

Enlarge Download

Share this: Share 0 Tweet 3 G+1 0 Embed\*

Cite this: Data, Scientific (2015): Scientific Data recommended repositories\_June 2015. figshare. <http://dx.doi.org/10.6084/m9.figshare.1434640> Retrieved 16:27, Jul 01, 2015 (GMT)

\*The embed functionality can only be used for non commercial purposes... more

<http://www.nature.com/sdata/data-policies/repositories>

[http://figshare.com/articles/Scientific Data recommended repositories June 2015/1434640](http://figshare.com/articles/Scientific_Data_recommended_repositories_June_2015/1434640)



# Conclusion

Or some perspectives

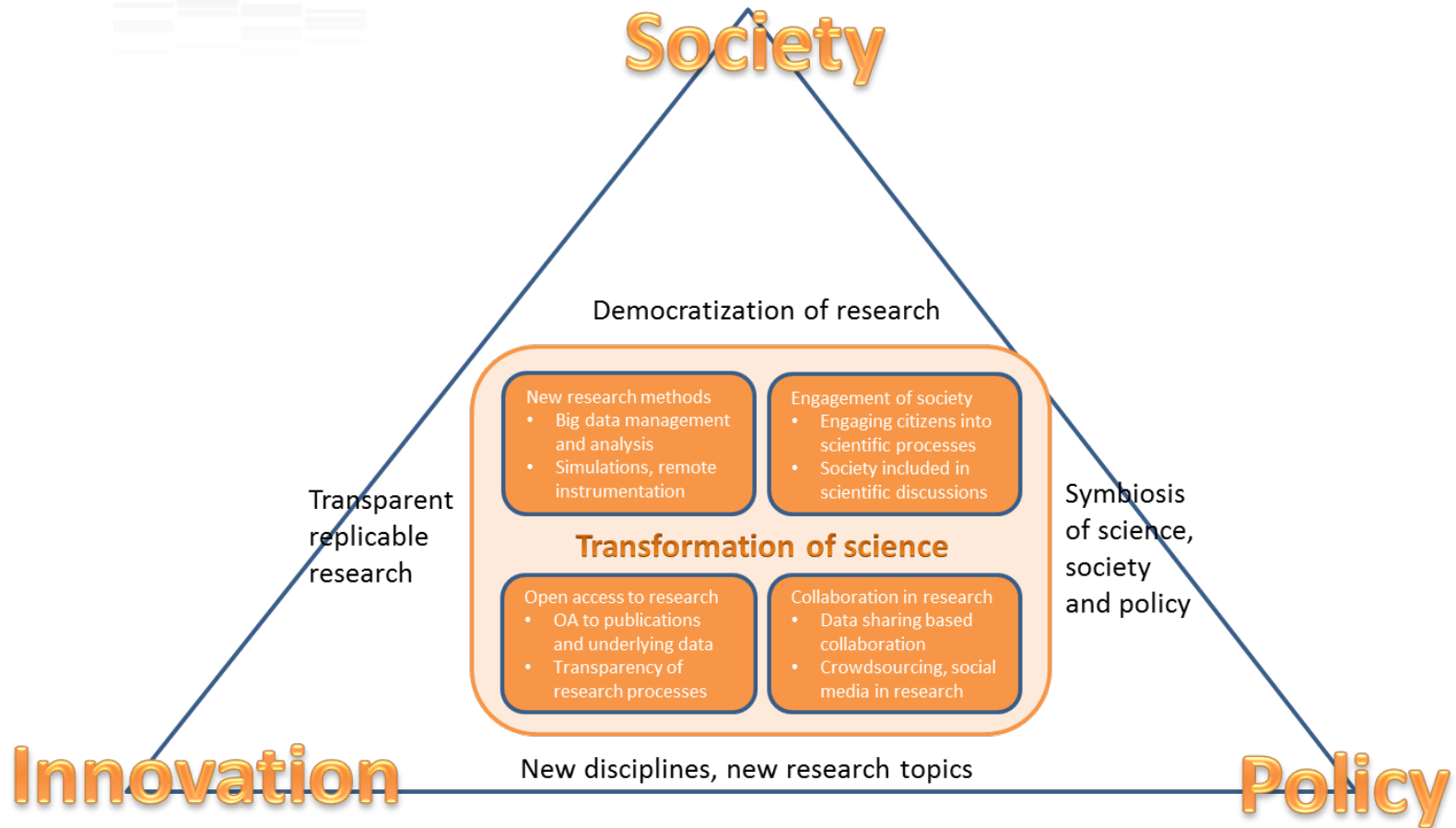
# New tools for the scientific process

## Typical workflow examples



<https://101innovations.wordpress.com/>

# How journals can tackle this challenges ?



<https://ec.europa.eu/digital-agenda/en/open-science>



- ❖ Rethink the scientific publishing process :
  - ✓ less and less frontiers between e-labnotebooks and results dissemination and treatment
- ❖ What are you waiting from your publisher :
  - ✓ only to put PDF on line
  - ✓ what else ?
- ❖ What is the publisher of the future ?