

# Searching Deeply for Data, Results and Tools *What is Stopping Us?*

Philip E. Bourne  
University of California San Diego  
[pbourne@ucsd.edu](mailto:pbourne@ucsd.edu)  
<http://www.sdsc.edu/pb>

Relevant Work from Us:  
<http://www.sdsc.edu/pb/SummaryScholarComm.pdf>  
<http://www.slideshare.net/pebourne/searching-deeply-for-data-results-and-tools-what-is-stopping-us>

Berlin-9 Nov. 9, 2011

# Meredith

A story that celebrates all that  
open access has done and shows  
the promise for the future



**PLOS** COMPUTATIONAL  
BIOLOGY

An official journal of the International Society for Computational Biology

That is the good news...

Now let me tell you what I believe  
we need to do to make Meredith  
the rule rather than the exception

*Disclaimer: My viewpoint is that of a computational biologist*

# The Research Enterprise



Methods

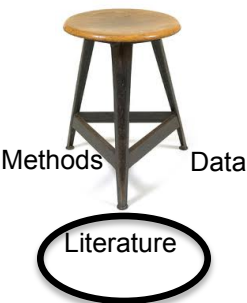
Data

Literature

# The Research Enterprise Without eScholarship

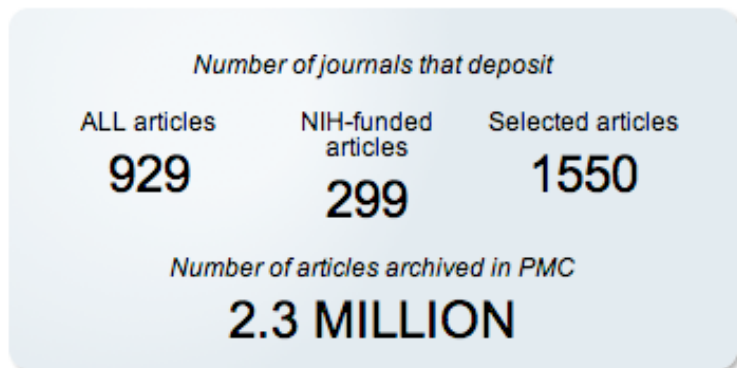


<http://www.flickr.com/photos/51282757@N05/5585299226/lightbox/>



# Literature

PubMed Central Contents Nov 8, 2011



- Meredith certainly benefited from the OA literature ...
- But we have a long way to go in several respects

PubMed Contents Nov 8, 2011

**PubMed.gov**  
US National Library of Medicine  
National Institutes of Health

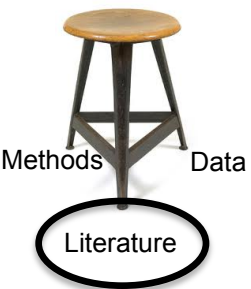
PubMed 1800:2100[dp]

RSS [Save search](#) [Limits](#) [Advanced](#)

[Display Settings:](#) ☒ Summary, 20 per page, Sorted by Recently Added

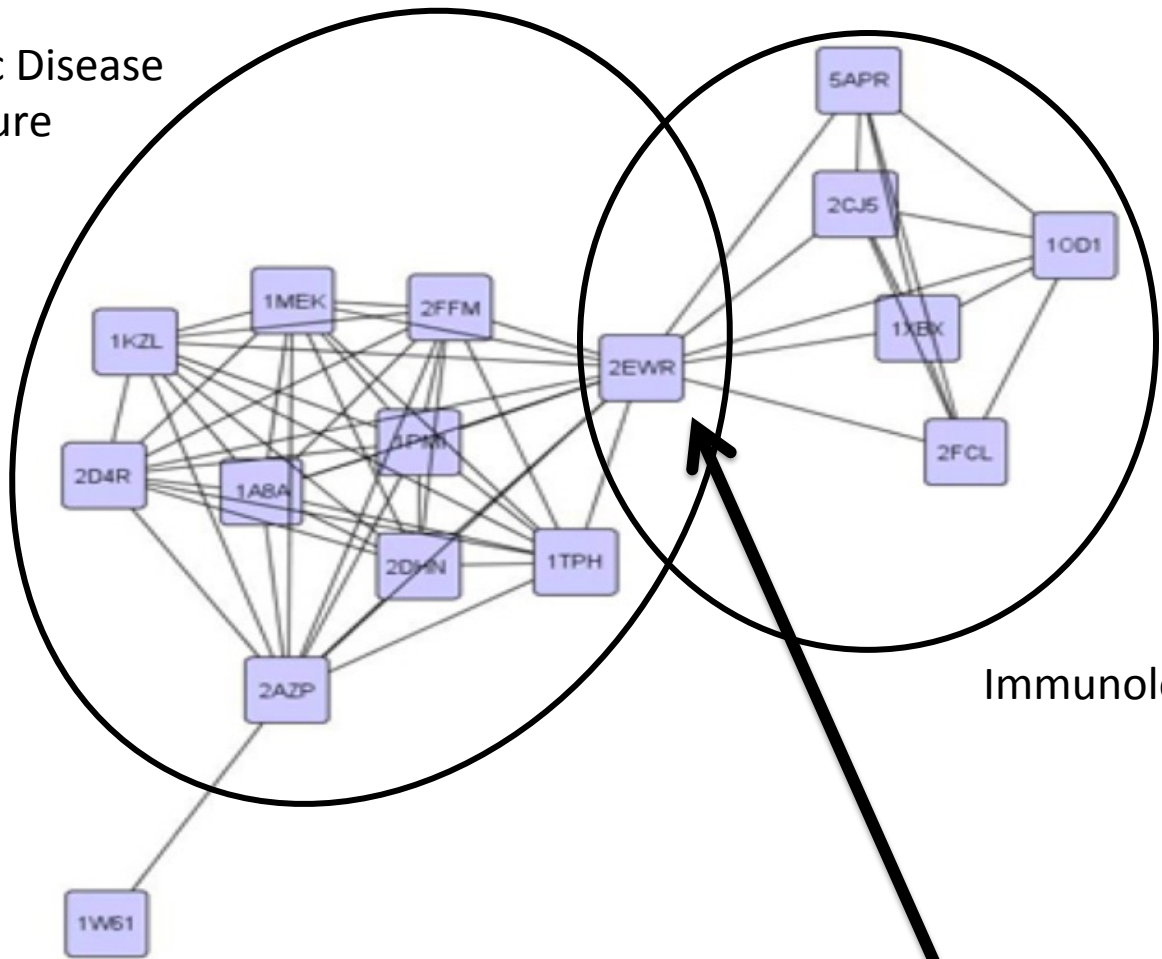
**Results: 1 to 20 of 21298403**

<< First < Prev Page **1** of 1064921



# Literature – But First What is the Promise

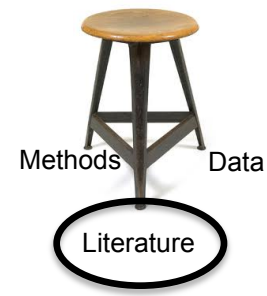
Cardiac Disease  
Literature



Immunology Literature

Shared Function

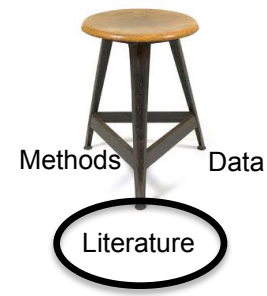
# Solution: We Need SciVerse™ like Developments Leveraging OA Content



- A very clever idea – The App model
- Leverage content
- Provide an open API
- Get the community to do all the work
- Drive folks to buy content

- Problem:
  - OA does not have the resources
  - OA + CA is more compelling
- Solution:
  - More developers
  - Federal and Foundation support to leverage OA content





# Literature - PMC

- Problems

- Producers do not care enough
- OA publishers are not aware enough – I can't even reliably parse the PMC license records to know what I can access

- Solutions

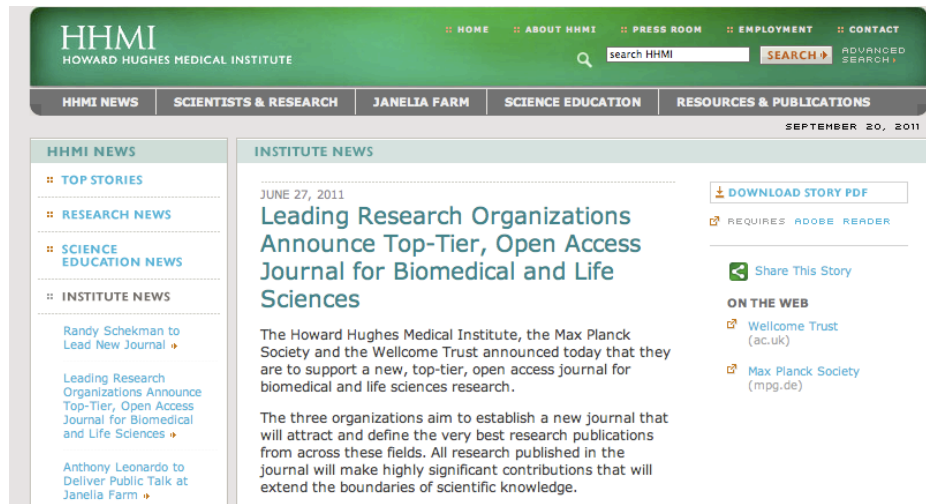
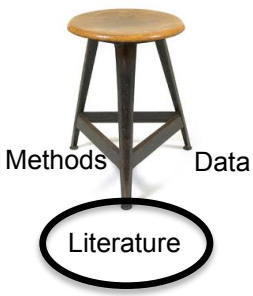
- Continue to advocate
- Raise awareness and establish consistency so that history does not repeat itself

## Advocacy

- Follow-on to Beyond the PDF  
<https://sites.google.com/site/beyondthepdf/>
- Force11 [www.force11.org](http://www.force11.org)

# Solutions:

## Scientists As Advocates



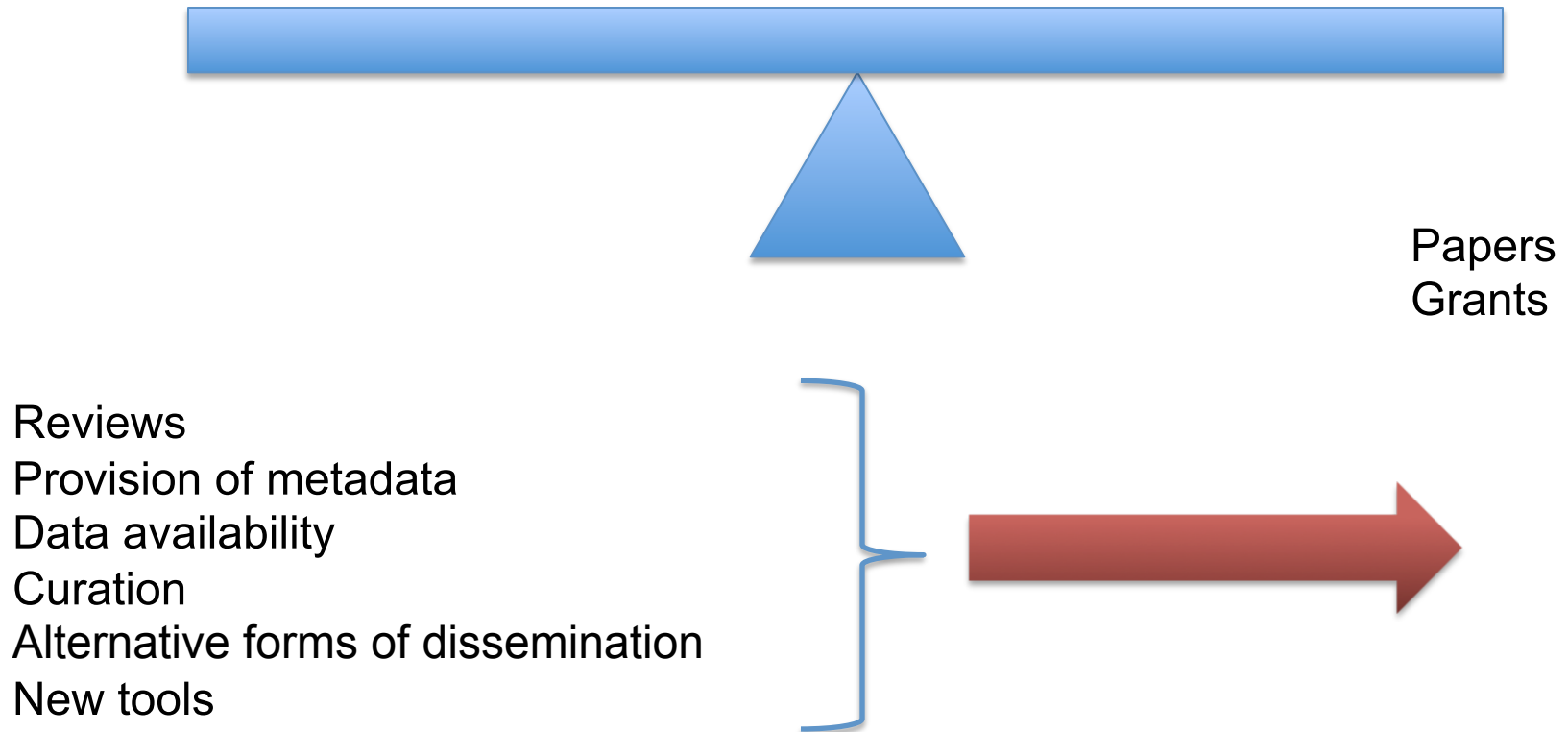
- My knee jerk reaction – is this the best a bunch of great minds can come up with!
- My more thoughtful reaction – every little bit helps – it will broaden awareness of the value of OA like nothing else

Solution:

- a) Play Upon a Scientist's Guilt re The Reward System
- b) Educate Evaluators

**The Right Thing To Do**

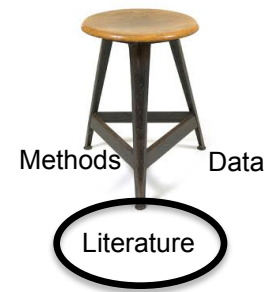
**Reward**



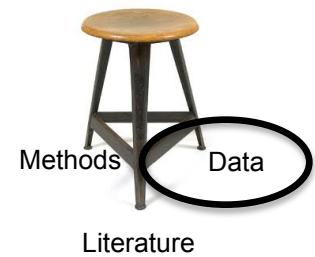
## Solution:

# Use the Traditional Reward System in New Ways

## The Wikipedia Experiment – Topic Pages



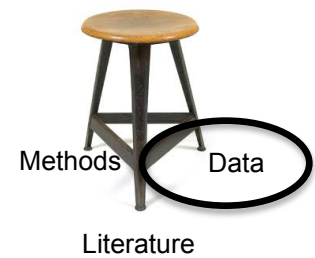
- Identify areas of Wikipedia that relate to the journal that are missing or stubs
- Develop a Wikipedia page in the sandbox
- Have a Topic Page Editor review the page
- Publish the copy of record with associated rewards
- Release the living version into Wikipedia



# Problems Regarding Data

- Meredith got the data she needed in part by bugging authors – It should be easier
- There is a long tail of data which is lost
- Institutional repositories are too institutional
- Journals are passing the buck
- We are heading towards the same issues with data repositories as we have with publishers

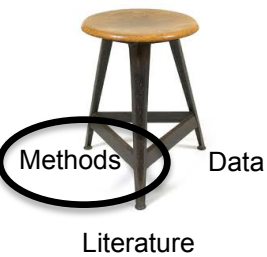
*Disclaimer: My viewpoint is that of a computational biologist*



# Solutions Regarding Data

- There are some wonderful resources out there (e.g., Entrez from the NLM) – copy them
- We need data repositories working together *now*
- We need more than a DOI – we need metadata catalogs for data so we can deep search and rank
- Data needs to be recognized as a publication

*Disclaimer: My viewpoint is that of a computational biologist*



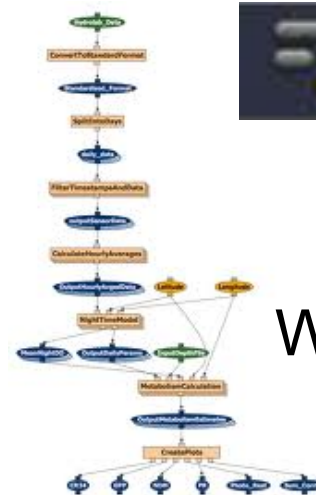
# Problems Regarding Methods

- Meredith had to use a “trial” license of Mathematica
- I cant easily reproduce my own research
- There is little reward for providing access to methods

*Disclaimer: My viewpoint is that of a computational biologist*



- 



# Wings



# Summary – Creating More Merediths

- OA groups to keep doing what they are doing and support leveraging the content as part of the research enterprise
- Advocacy
  - More initiatives like the Hargreaves data mining proposal
  - Funders now have data sharing policies, next is methods sharing
- Funding for collaboration, standards, OA killer apps

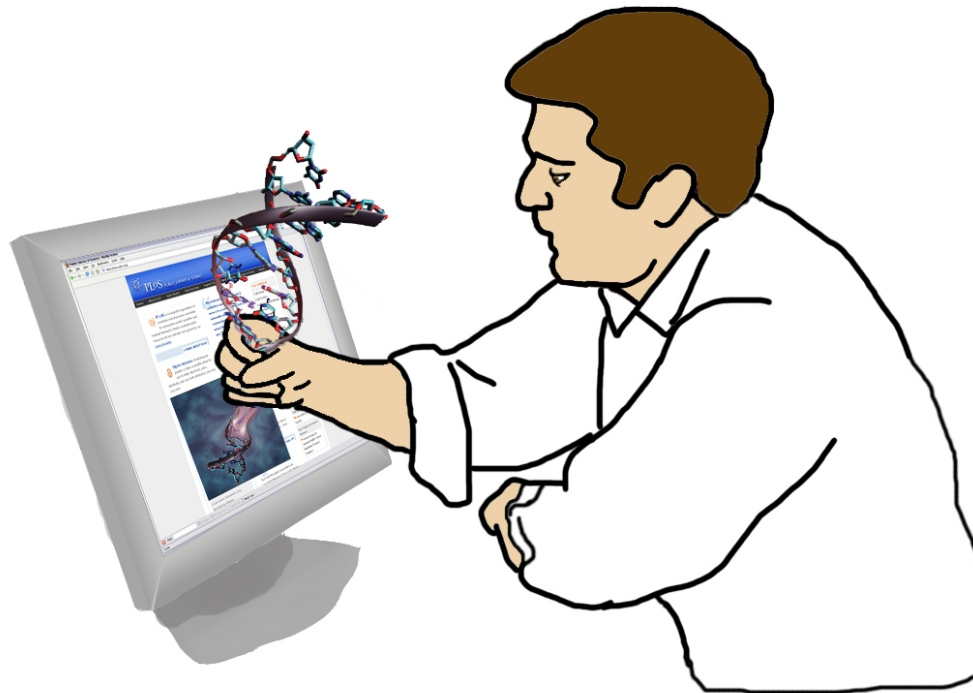
# General References

- *What Do I Want from the Publisher of the Future PLoS Comp Biol* 6(5): e1000787
- *Fourth Paradigm: Data Intensive Scientific Discovery* <http://research.microsoft.com/enus/collaboration/fourthparadigm/>

# References to Exemplars

- [Semantic Biochemical Journal - 2010: Using Utopia](#)
- Article of the Future, Cell, 2009:
- Prospect, Royal Society of Chemistry, 2009:
- [Adventures in Semantic Publishing, Oxford U, 2009:](#)
- **The Structured Digital Abstract, Seringhaus/Gerstein, 2008**
- **CWA Nanopublications – 2010**
- <https://sites.google.com/site/beyondthepdf/>
- <https://sites.google.com/site/futureofresearchcommunications/>
- <http://www.force11.org>

# Thank You!



[pbourne@ucsd.edu](mailto:pbourne@ucsd.edu)