

# e-Research and the Demise of the Scholarly Article

David De Roure



digital.humanities  
@ oxford

PHILOSOPHICAL  
TRANSACTIONS:  
GIVING SOME  
ACCOMPT  
OF THE PRESENT  
Undertakings, Studies, and Labours  
OF THE  
INGENIOUS  
IN MANY  
CONSIDERABLE PARTS  
OF THE  
WORLD.

Vol I.

For Anno 1665, and 1666.

In the SAVOY,  
Printed by T. N. for John Martyn at the Bell, a little with-  
out Temple-Bar, and James Allestry in Duck-Lane,  
Printers to the Royal Society.

PHILOSOPHICAL  
TRANSACTIONS  
OF

THE ROYAL  
SOCIETY



MATHEMATICAL,  
PHYSICAL  
& ENGINEERING  
SCIENCES

## An e-Research approach to Web-scale music analysis

David De Roure, Kevin R. Page, Benjamin Fields, Tim Crawford, J. Stephen Downie and Ichiro Fujinaga

*Phil. Trans. R. Soc. A* 2011 **369**, 3300-3317  
doi: 10.1098/rsta.2011.0171

### References

[This article cites 5 articles](#)

<http://rsta.royalsocietypublishing.org/content/369/1949/3300.full.html#ref-list-1>

[Article cited in:](#)

<http://rsta.royalsocietypublishing.org/content/369/1949/3300.full.html#related-urls>

### Subject collections

Articles on similar topics can be found in the following collections

[e-science](#) (43 articles)

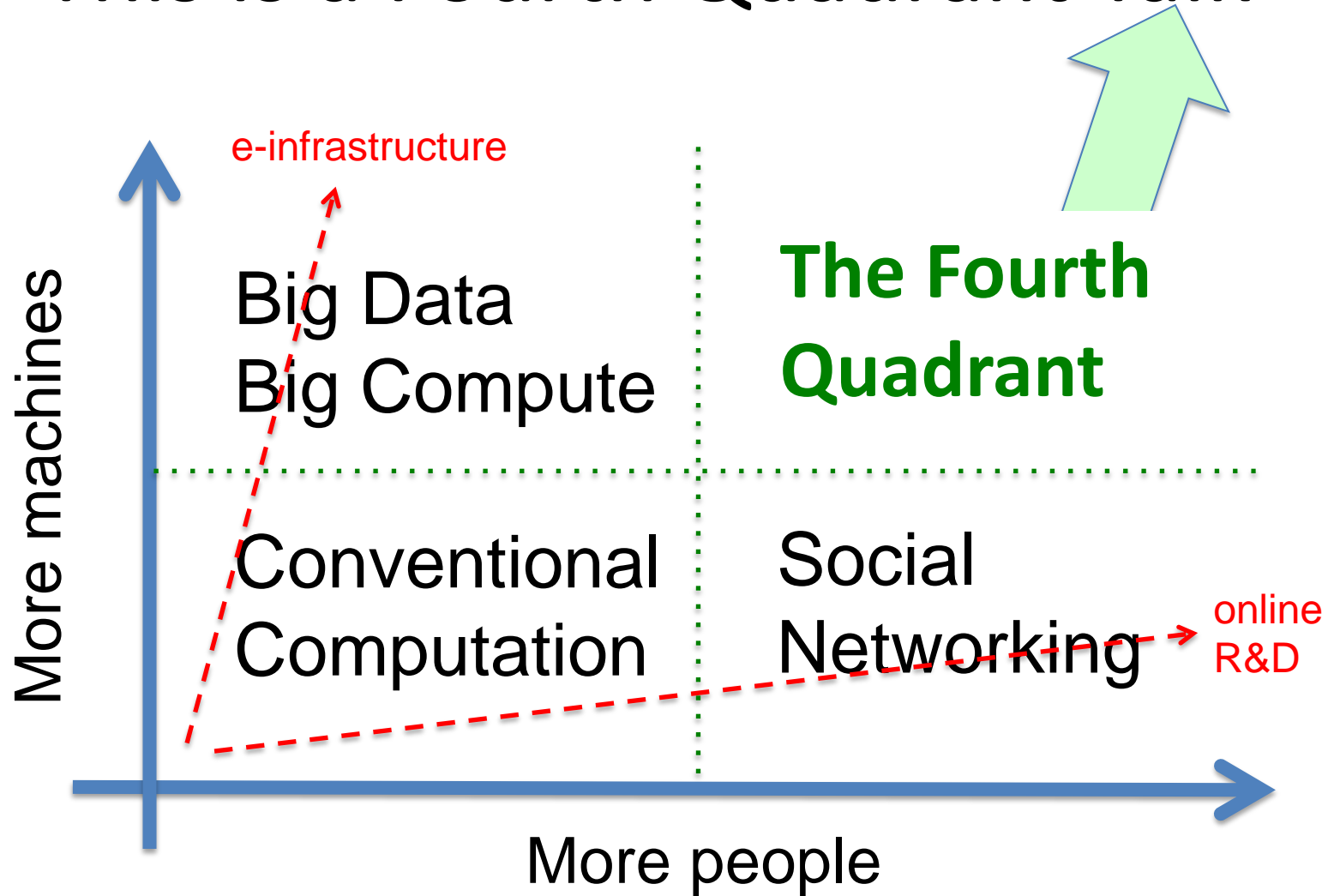
### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

A revolutionary idea...  
Open Science!

To subscribe to *Phil. Trans. R. Soc. A* go to:  
<http://rsta.royalsocietypublishing.org/subscriptions>

# This is a Fourth Quadrant Talk



Deluge!!!

Data!!

Scientists

Social Scientists

Humanists

Funding agencies

Policy makers

Librarians

Publishers

Internet architects

Christine Borgman



Here is the evidence, now  
what is the hypothesis?  
The complementary roles of  
inductive and hypothesis-driven  
science in the post-genomic

Douglas B. Kell<sup>1</sup>

**Summary**

It is considered in some quarters that the only viable means of scientific advancement in scientific knowledge is the development of technology (beyond the 'hypothesis-led'—might be of value)—must be the hypothetico-deductive here that data- and technology alternatives to hypothesis-driven knowledge discovery but collaborative partners with them. hypothesis-poor. Here, collaborative generating novel hypothesis in the genomic era. *BioEssays*  
© 2003 Wiley Periodicals, Inc.

**WIRED** SUBSCRIBE >>

WIRED MAGAZINE: 16.

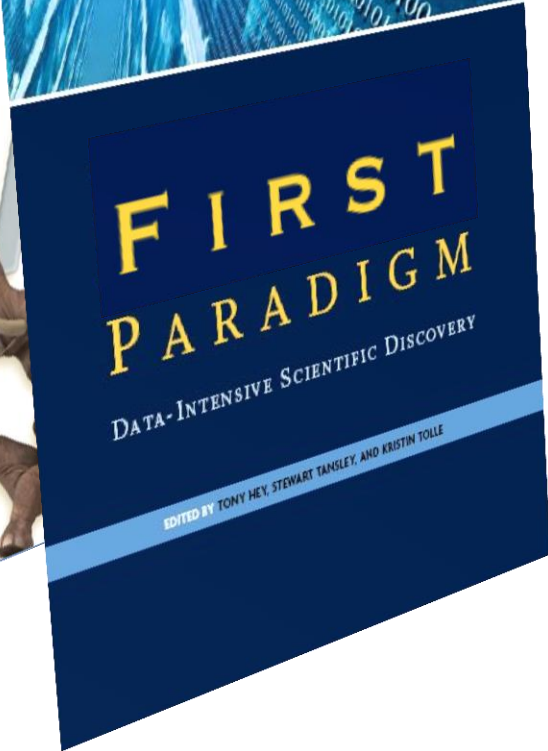
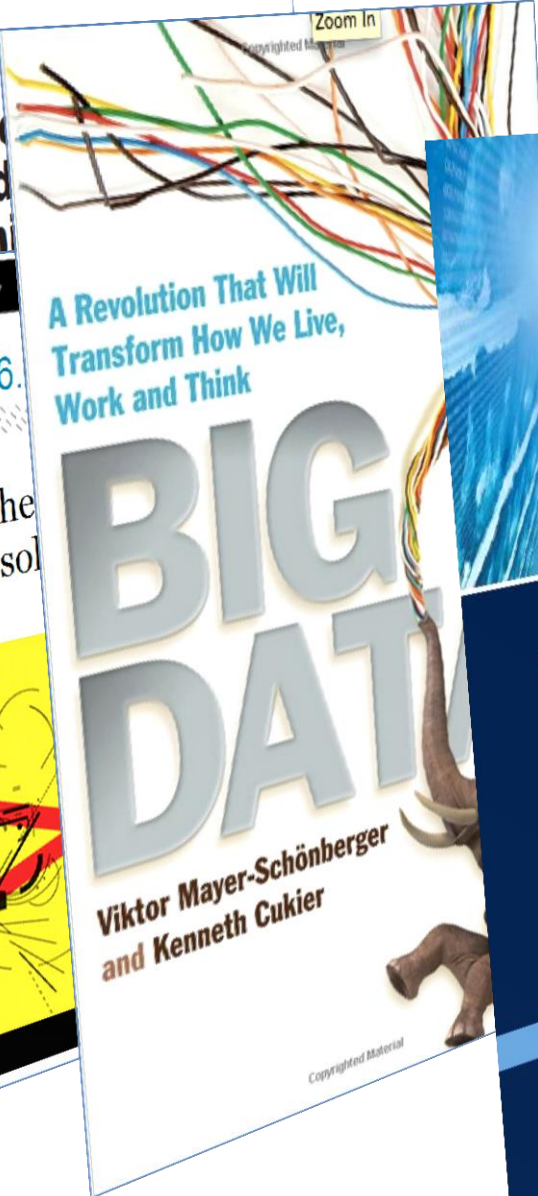
SCIENCE : DISCOVERIES

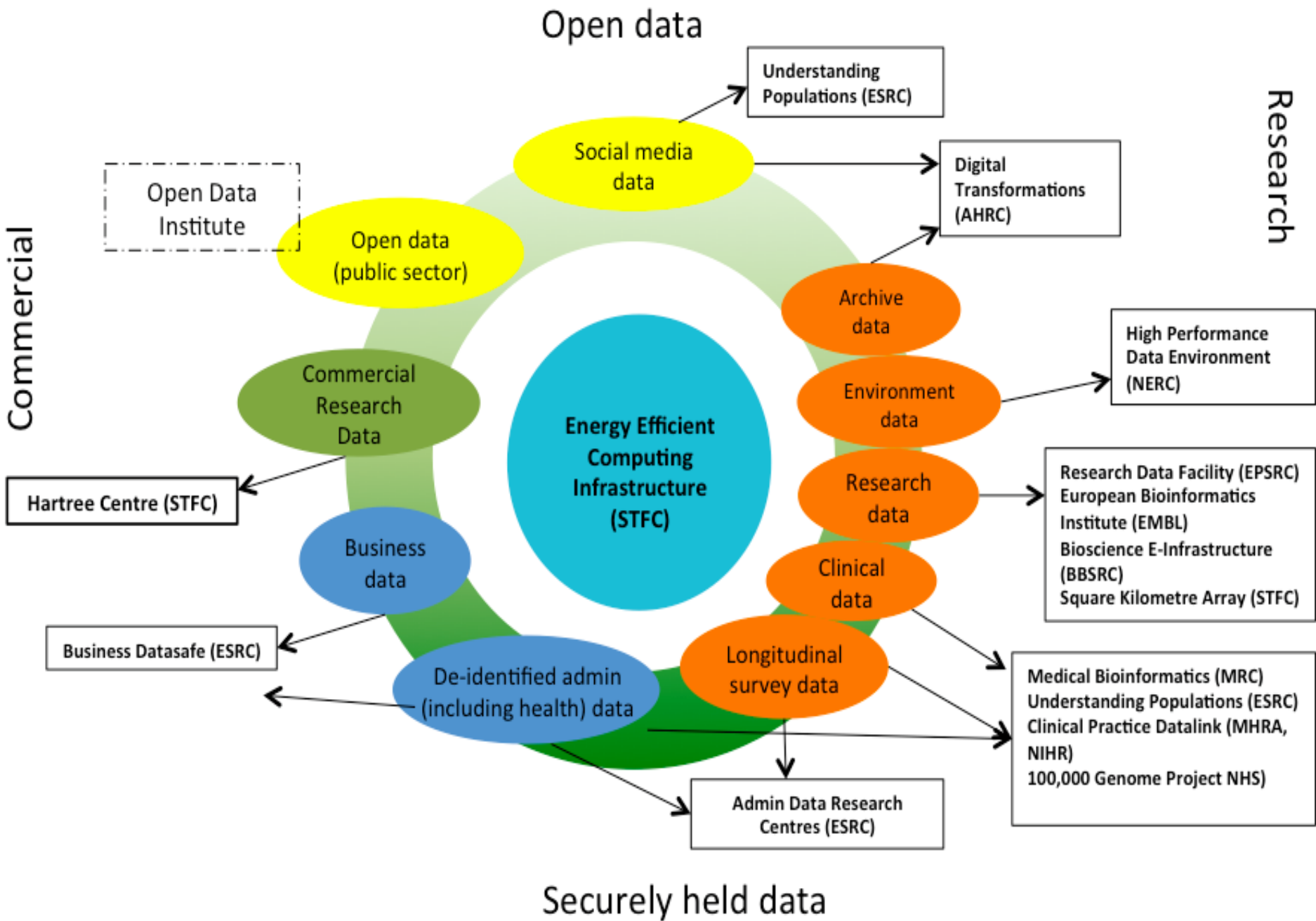
The End of Theory: The  
Scientific Method Obsol

By Chris Anderson 06.23.08



Illustration: Marian Bantjes





Notifications and automatic re-runs

Autonomic

Self-repair

Curation

New research?

*Machines are users too*



# ZOONIVERSE

REAL SCIENCE ONLINE

## Study explosions on the Sun

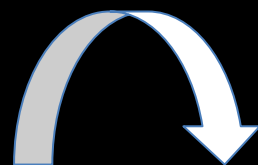
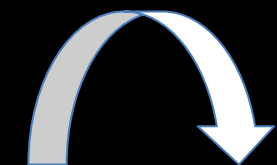
Explore interactive diagrams to learn about the Sun and the spacecraft monitoring it.

## Find planets around stars

Lightcurve changes from the Kepler spacecraft can indicate transiting planets.

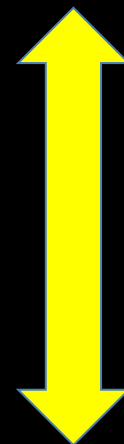
Image  
Classification

Talk  
Forum



**Citizen Scientists**

*data reduction*



**Scientists**



Mode  
using

Help scientists recover Arctic

Scientists at NOAA's National

The data gathered by Ancient

You can help scientists from the







## Beyond the PDF2 Conference

Tuesday, March 19, 2013 to Wednesday, March 20, 2013

Pakhuis de Zwijger, Amsterdam, NL

Twitter: #btpdf2



**The demise of the paper around 2030 can be attributed to several factors:**

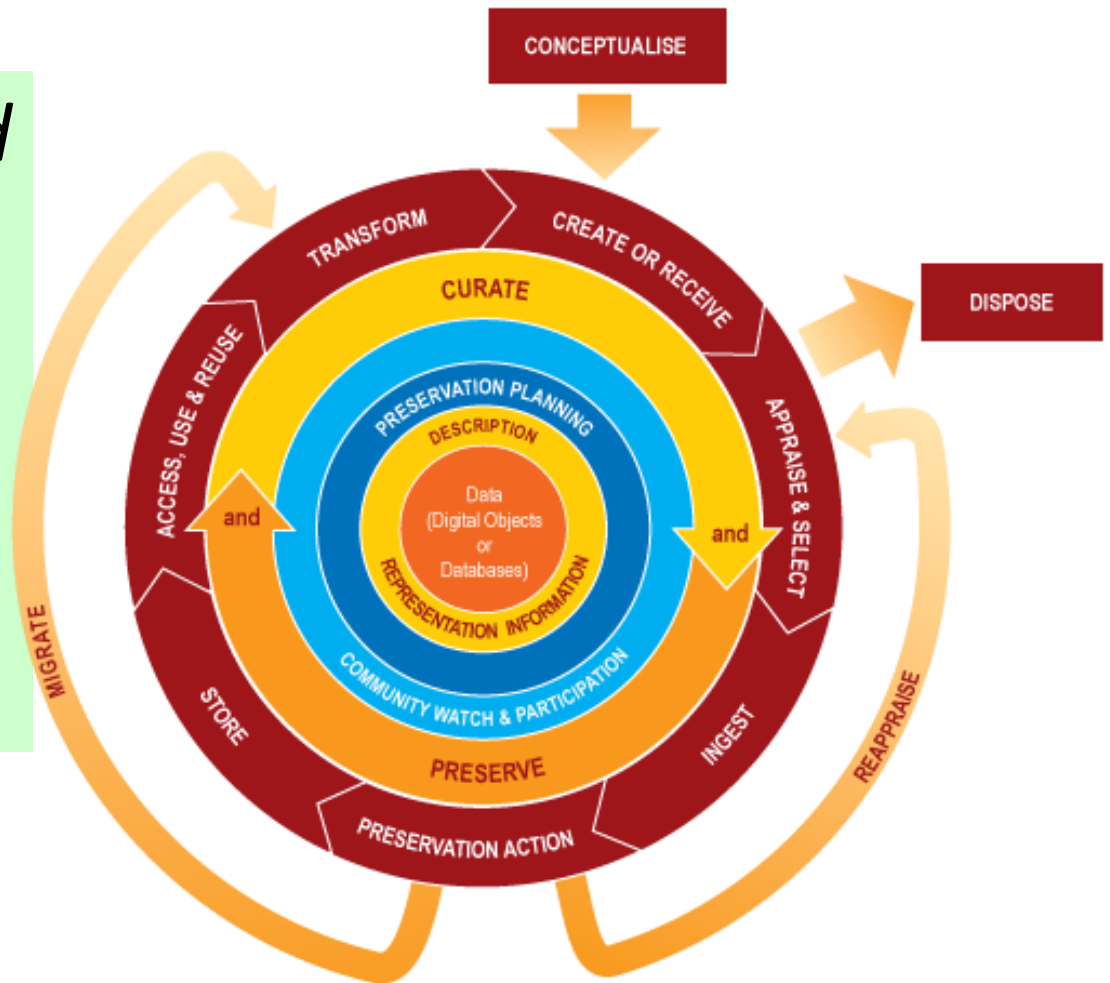
- 1. It was no longer possible to include the evidence in the paper.**
- 2. It was no longer possible to reconstruct a scientific experiment based on a paper alone.**
- 3. Writing for increasingly specialist audiences restricted essential multidisciplinary re-use.**
- 4. Research records needed to be readable by computer to support automation and curation.**
- 5. Single authorship gave way to casts of thousands of collaborators and citizen scientists, leading to failure of the authorship and incentive model.**
- 6. Quality control models scaled poorly with the increasing volume and "open access" movement, obscuring innovation.**
- 7. Alternative reporting was found necessary for compliance with increasingly stringent scientific and industrial regulations.**
- 8. Frustrated by inefficiencies in scholarly communication that stifled progress, research funders demanded change.**

**The demise of the paper around 2030 can be attributed to several factors:**

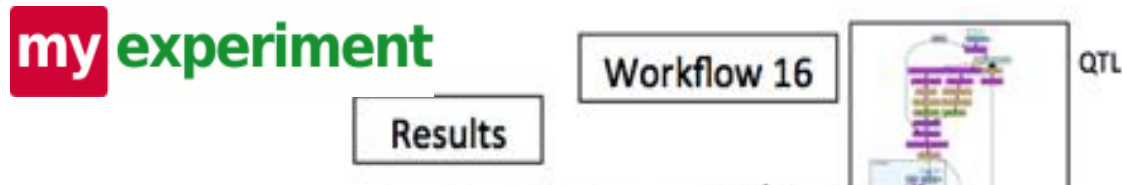
- 1. It was no longer possible to include the evidence in the paper.**
- 2. It was no longer possible to reconstruct a scientific experiment based on a paper alone.**
- 3. Writing for increasingly specialist audiences restricted essential multidisciplinary re-use.**
- 4. Research records needed to be readable by computer to support automation and curation.**
- 5. Single authorship gave way to casts of thousands of collaborators and citizen scientists, leading to failure of the authorship and incentive model.**
- 6. Quality control models scaled poorly with the increasing volume and “open access” movement, obscuring innovation.**
- 7. Alternative reporting was found necessary for compliance with increasingly stringent scientific and industrial regulations.**
- 8. Frustrated by inefficiencies in scholarly communication that stifled progress, research funders demanded change.**

1. It was no longer possible to include the evidence in the paper – container failure!

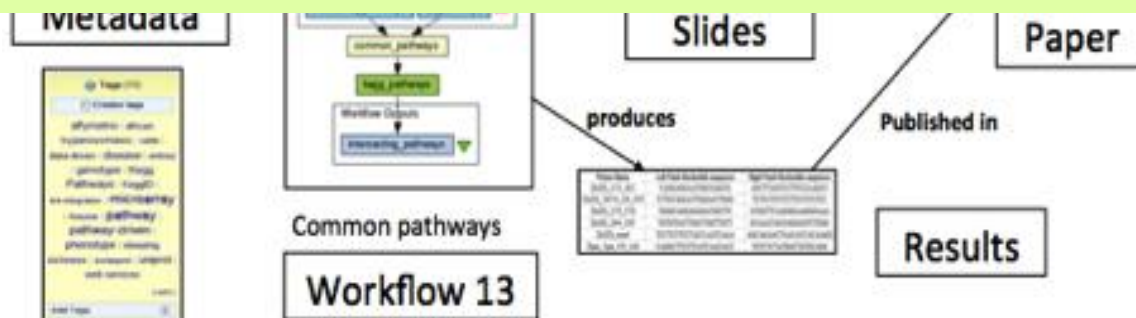
*“A PDF exploded today when a scientist tried to paste in the twitter firehose...”*



- It was no longer possible to reconstruct a scientific experiment based on a paper alone



**Meanwhile considerable confusion arose in the crisis of reproducibility. This tenet of the scientific method, based on independent reconstruction of experiments, suffered when the sharing of digital artefacts seriously interfered with the notion of independence: sharing was at once beneficial and self-defeating.**

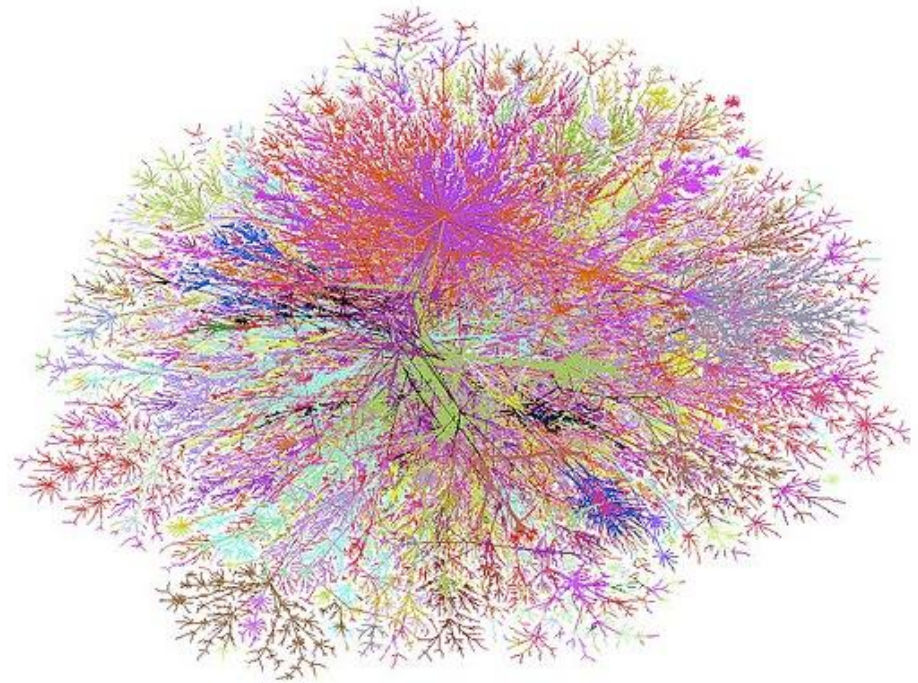
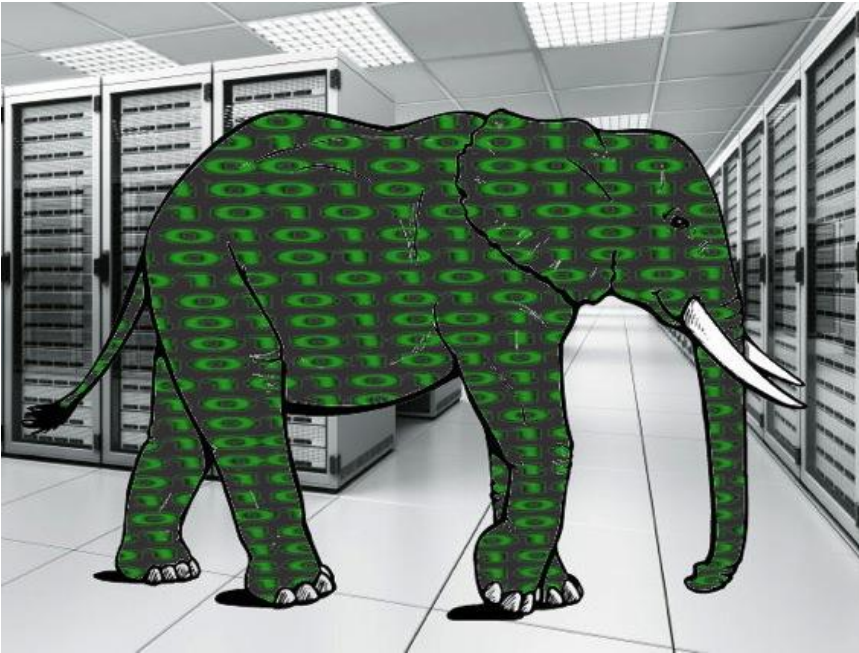


### 3. Writing for increasingly specialist audiences restricted essential multidisciplinary re-use

*Grand Challenge Areas:*

- *Energy*
- *Living with Environmental Change*
- *Global Uncertainties*
- *Lifelong Health and Wellbeing*
- *Digital Economy*
- *Nanoscience*
- *Food Security*
- *Connected Communities*
- *Resilient Economy*

#### 4. Research records needed to be readable by computer to support automation and curation



A computationally-enabled sense-making network of expertise, data, models and narratives.

# 5. Single authorship gave way to casts of thousands

## Climate



**Model Earth's climate using wartime ship logs**  
Help scientists recover worldwide weather observations made by Royal Navy ships.  
aaWeather



**Classify over 30 years of tropical cyclone data.**  
Scientists at NOAA's National Climatic Data Center need your help.  
CycloneCenter

## Humanities



**Study the lives of ancient Greeks**  
The data gathered by Ancient Lives helps scholars study the Oxyrhynchus collection.  
ANCIENT LIVES

## Nature



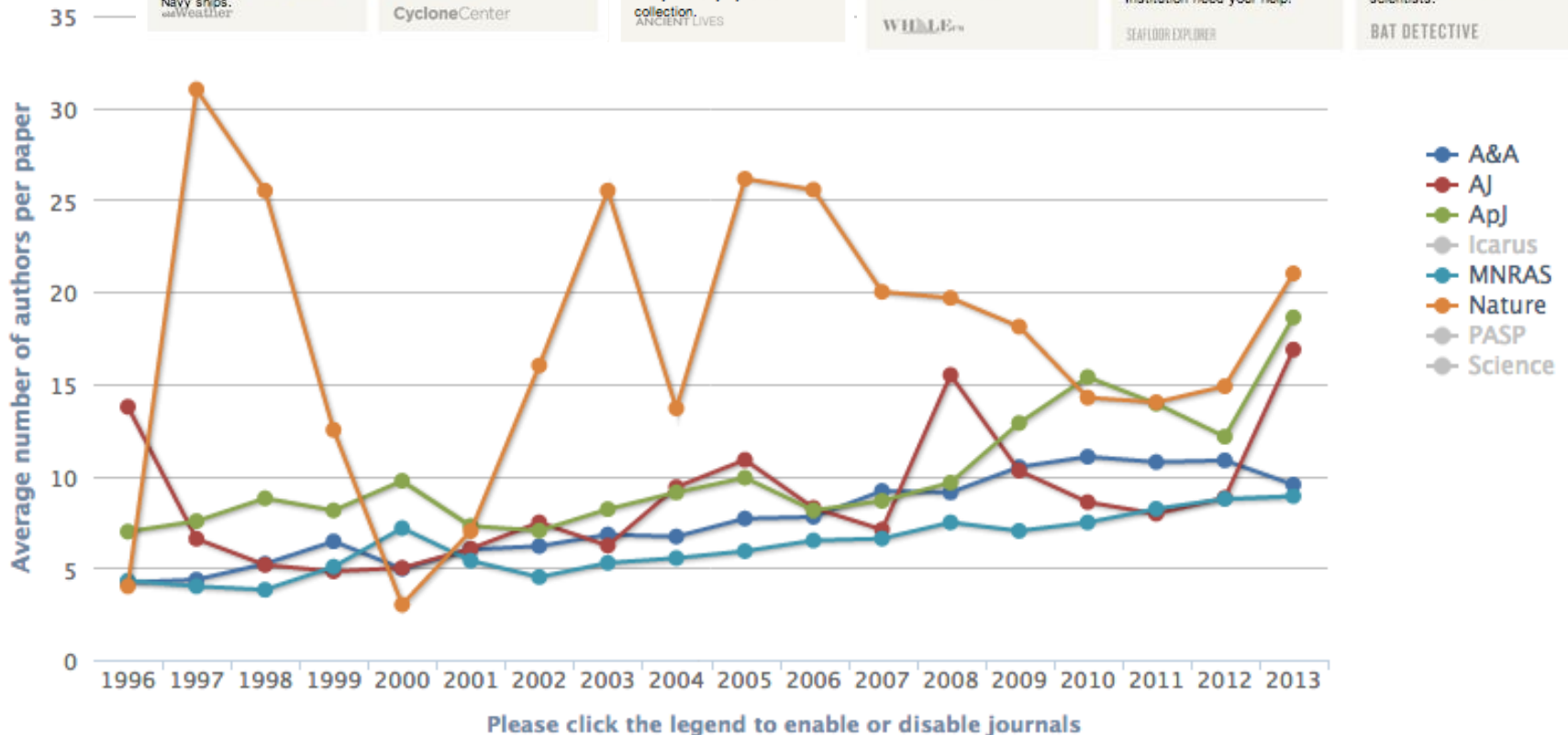
**Hear Whales communicate**  
You can help marine researchers understand what whales are saying.  
WHALES



**Help explore the ocean floor**  
The HabCam team and the Woods Hole Oceanographic Institution need your help!  
SEAFLOOR EXPLORER

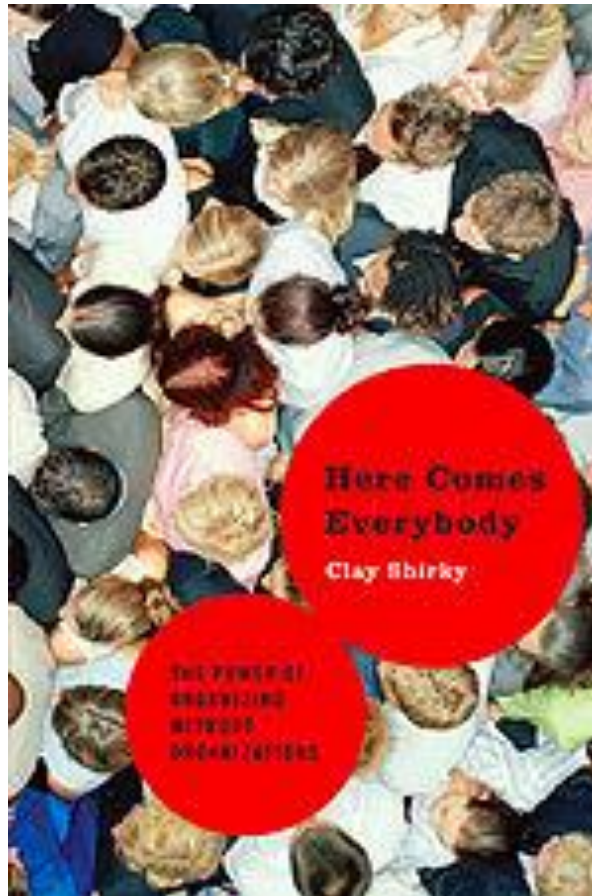


**You're hot on the trail of bats!**  
Help scientists characterise bat calls recorded by citizen scientists.  
BAT DETECTIVE





## 6. Quality control models scaled poorly with the increasing volume



# 7. Alternative reporting necessary for compliance with regulations

## 'Show Your Working': What 'ClimateGate' means



### VIEWPOINT

Mike Hulme and Jerome Ravetz

**The "ClimateGate" affair - the publication of e-mails and documents hacked or leaked from one of the world's leading climate research institutions - is being intensely debated on the web. But what does it imply for climate science? Here, Mike Hulme and Jerome Ravetz say it shows that we need a more concerted effort to explain and engage the public in understanding the processes and practices of science and scientists.**

As the repercussions of [ClimateGate](#) reverberate around the virtual community of global citizens, we believe it is both important and urgent to reflect on what this moment is telling us about the practice of science in the 21st Century.

In particular, what is it telling us about the social status and perceived authority of scientific claims about climate change?

We argue that the evolving practice of science in the contemporary world must be different from the classic view of disinterested - almost robotic - humans establishing objective claims to universal truth.



“ Practising scientists know that they do not simply follow a rulebook to do their science, otherwise it could be done by a robot ”

### ▶ THE GREEN ROOM

A weekly series of thought-provoking opinion pieces on environmental topics



### Pinch of salt

Idea that the world's food production must double "is wrong"

▶ Your comments

### RECENT ARTICLES

- ▶ Time to close global energy gap
- ▶ Mangroves offer win-win opportunity
- ▶ Is wildlife being eaten to extinction?
- ▶ Humanity needs to take 'giant leap'
- ▶ Protecting wildlife in conflict zones
- ▶ Could things go from bad to worse?
- ▶ Fussy eaters – what's wrong with GM?
- ▶ Calls for 'old-fashioned' revolution
- ▶ Payback time for home generation
- ▶ IPCC's 'vital role' in climate policy
- ▶ Decision time for whaling
- ▶ Hooking the 'pirates'

### LINKS

- ▶ Copenhagen summit
- ▶ Richard Black's Earth Watch
- ▶ Earth News

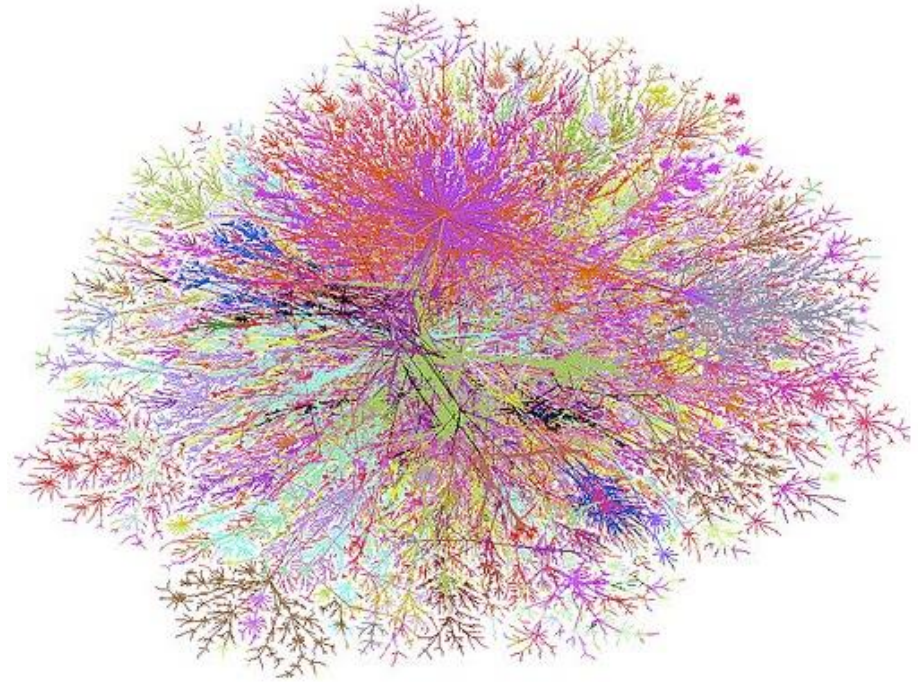
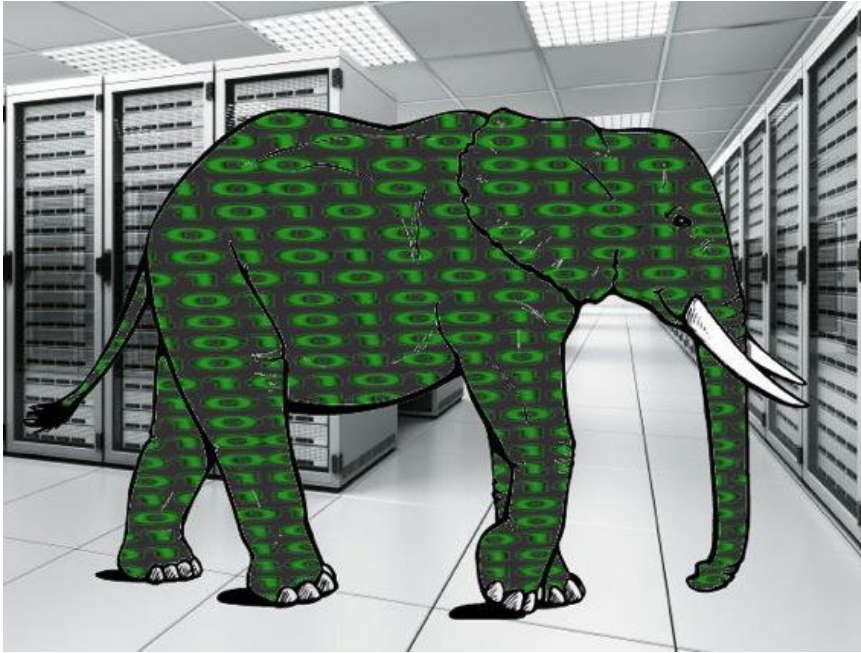
## 8. Research funders frustrated by inefficiencies in scholarly communication

*An investment is only worthwhile if*

- *Outputs are discoverable*
- *Outputs are reusable*
- *Outputs accrue value*

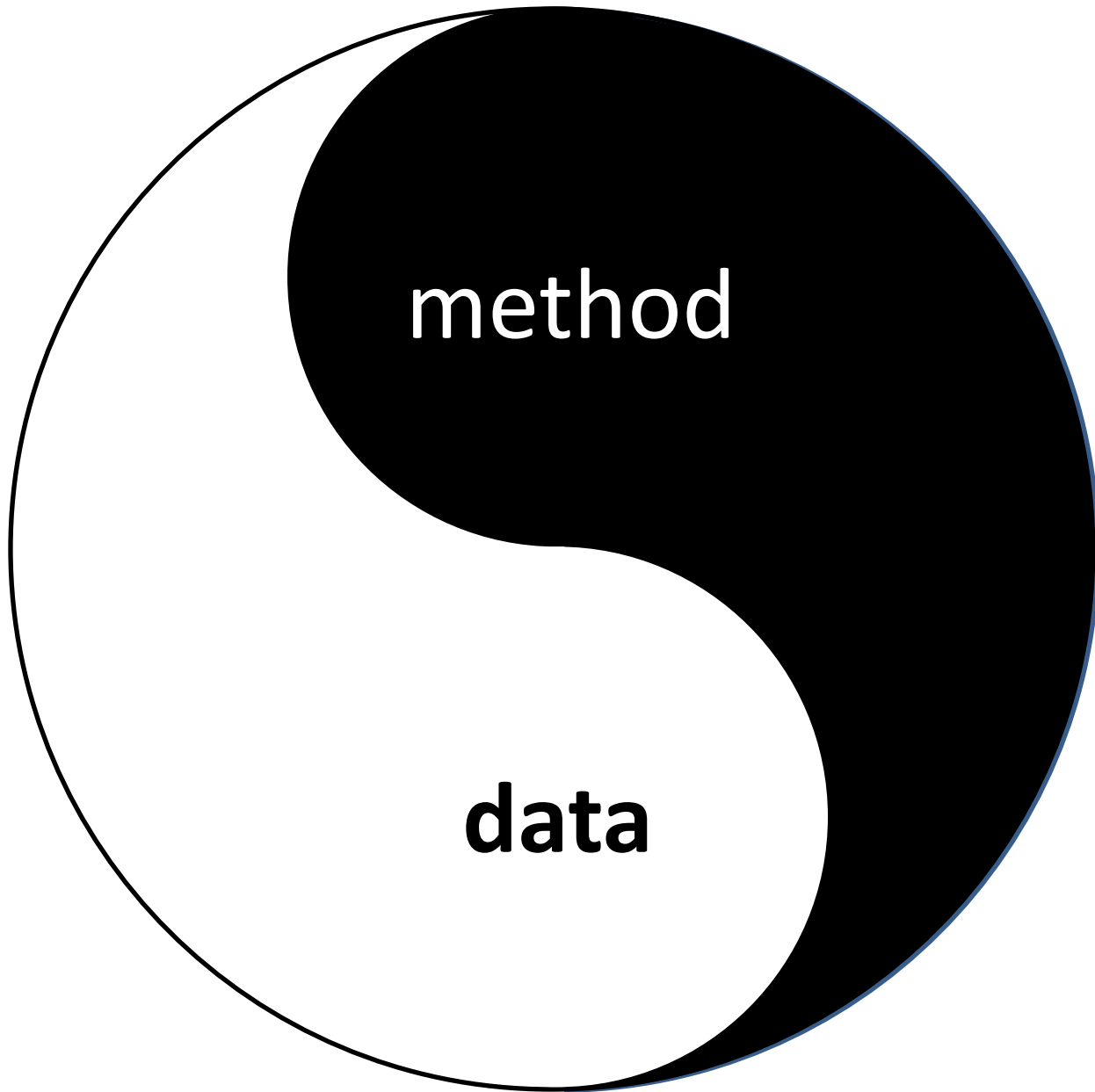
**3. Full value from investment.** Investments are less worthwhile unless all the outputs are discoverable and re-useable. Is one model better for encouraging re-use of research data and of software? Does it help with dissemination and resource discovery? Does it help with new forms of data and new digital artefacts? Is it better for others adding value to the outcomes, so that I can better fund innovation over existing resources which adds to their value? (Metapoint: I have a duty to maximise return on investment.)

# Big data elephant versus sense-making network?



The challenge is to foster the co-constituted socio-technical system on the right i.e. a computationally-enabled sense-making network of expertise, data, models and narratives.

This requires a “social machines” perspective from the outset as well as humanistic input. The Web, and with it Web Science, are an important exemplar.

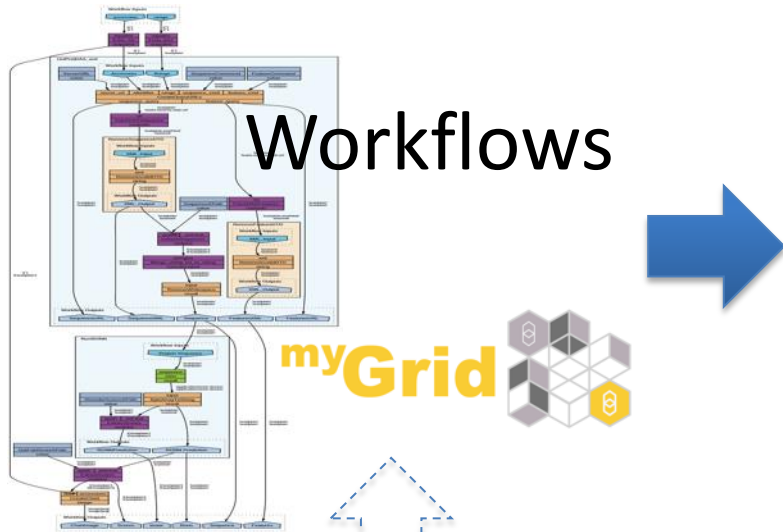


method

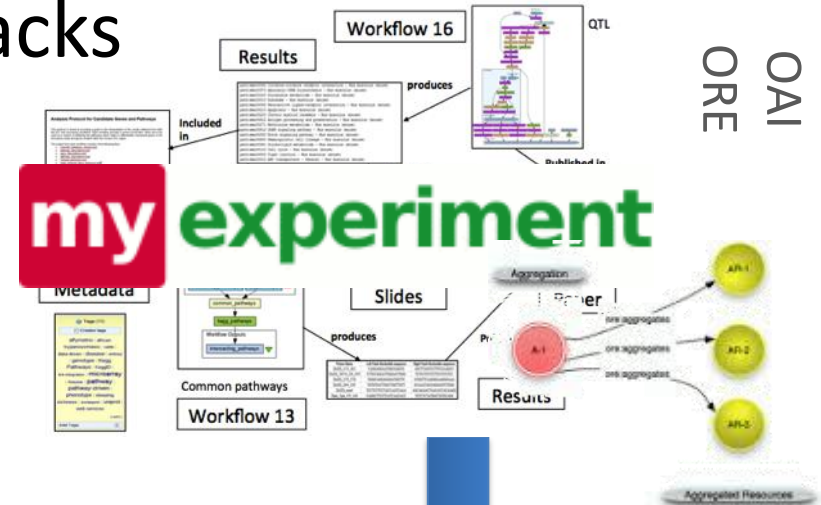
data



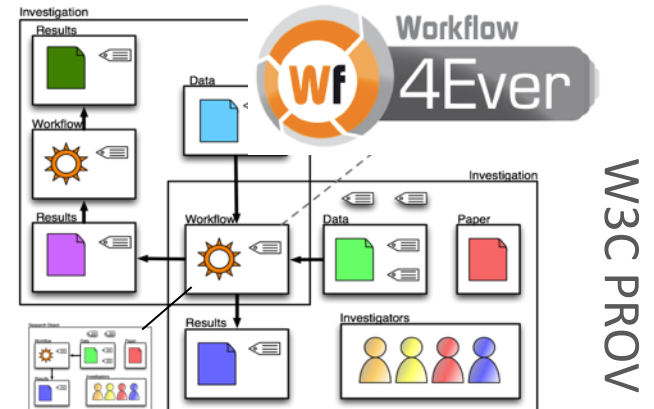
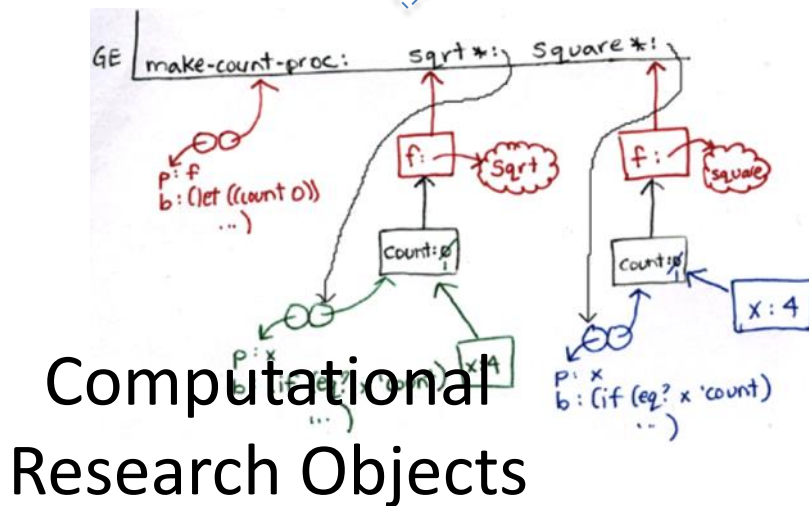
# Evolving the myExperiment Social Machine



Packs



Research Objects



# The R dimensions



**Reusable.** The key tenet of Research Objects is to support the sharing and reuse of data, methods and processes.

**Repurposeable.** Reuse may also involve the reuse of constituent parts of the Research Object.

**Repeatable.** There should be sufficient information in a Research Object to be able to repeat the study, perhaps years later.

**Reproducible.** A third party can start with the same inputs and methods and see if a prior result can be confirmed.

**Replayable.** Studies might involve single investigations that happen in milliseconds or protracted processes that take years.

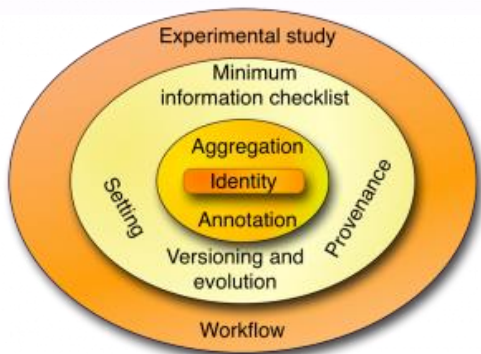
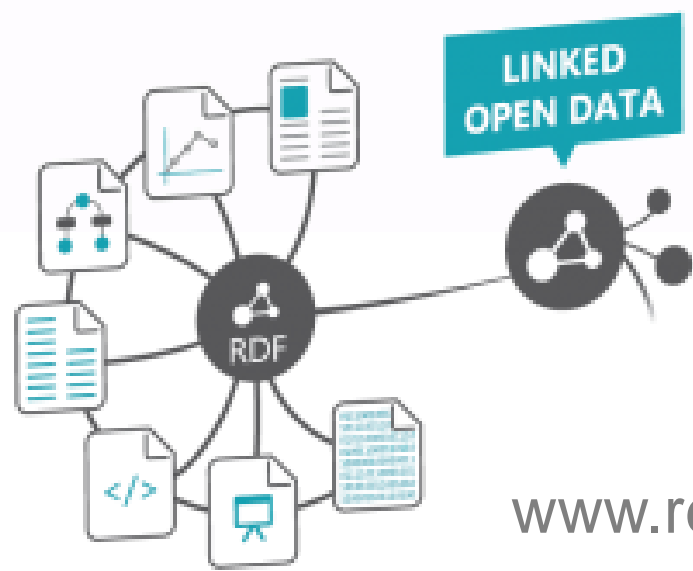
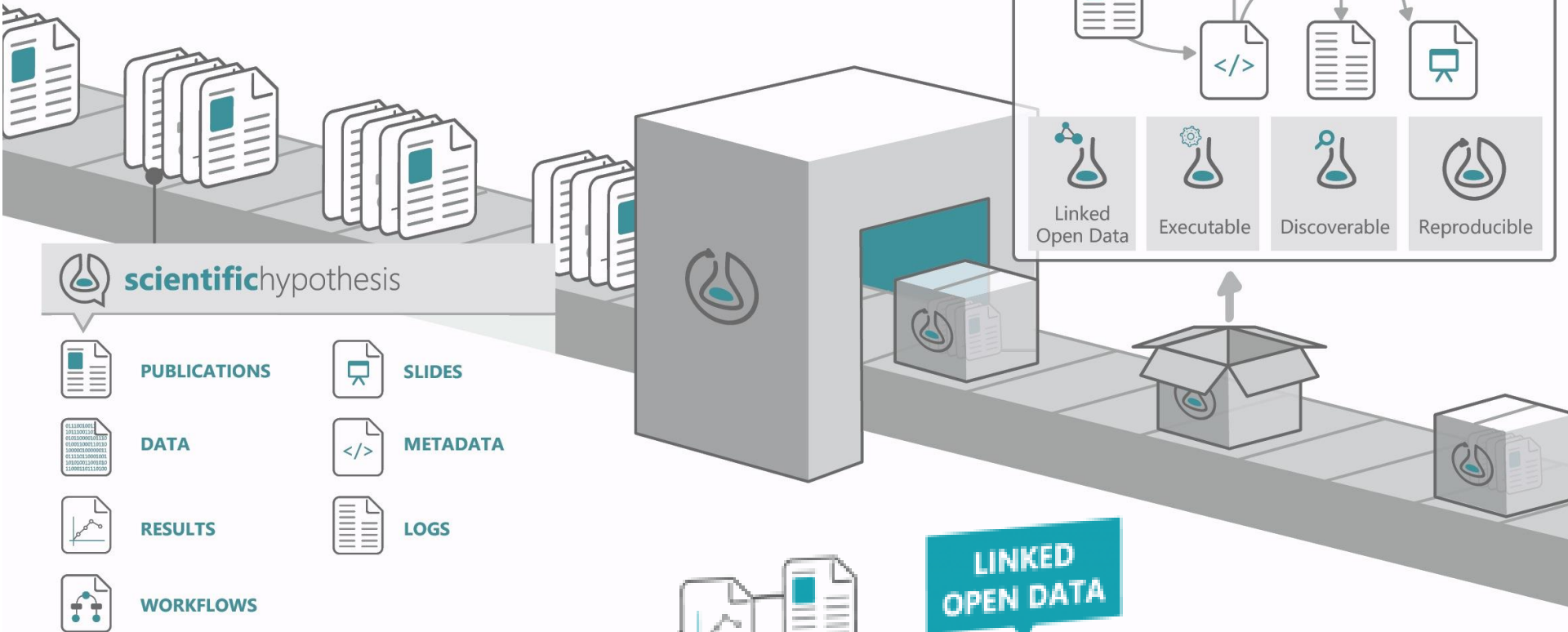
**Referenceable.** If research objects are to augment or replace traditional publication methods, then they must be referenceable or citeable.

**Revealable.** Third parties must be able to audit the steps performed in the research in order to be convinced of the validity of results.

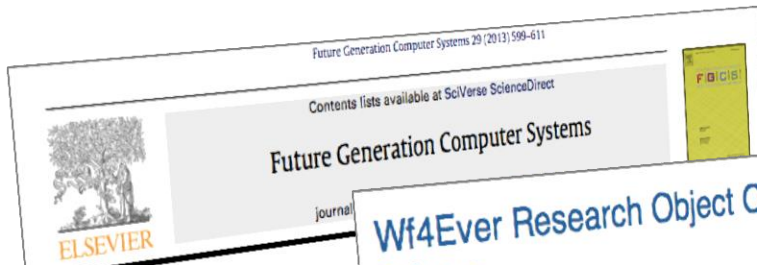
**Respectful.** Explicit representations of the provenance, lineage and flow of intellectual property.



Enabling **reproducible**, transparent research.



[www.researchobject.org](http://www.researchobject.org)



## Why linked data is not enough

Sean Bechhofer<sup>a,\*</sup>, Iain Buchan<sup>b</sup>, David Philip Couch<sup>b</sup>, Don Cruickshank<sup>c</sup>, Mark Danius Michaelides<sup>c</sup>, Stuart Owen<sup>a</sup>, Da

<sup>a</sup> School of Computer Science, University of Manchester, UK  
<sup>b</sup> School of Community Based Medicine, University of Manchester, UK  
<sup>c</sup> School of Electronics and Computer Science, University of Southampton  
<sup>d</sup> Oxford e-Research Centre, University of Oxford, UK

### ARTICLE INFO

Article history:  
Received 9 March 2011  
Received in revised form  
18 July 2011  
Accepted 5 August 2011  
Available online 19 August 2011

Keywords:  
Research object  
Linked data  
Reproducibility  
Reuse  
Sharing  
Publishing

### 1. Introduction

Changes are occurring in the ways in which scientific workflows, research protocols, standard procedures and algorithms for analysis or simulation are used to manipulate and produce data. Experimental or data and scientific models are typically “born digit

## Wf4Ever Research Object Ontologies and Vocabularies Primer

Unofficial Draft 13 April 2012

Editor:  
[Jun Zhao](#), University of Oxford

Authors:  
[Khalid Belhajjame](#), University of Manchester  
[Daniel Garjo](#), UPM  
[Oscar Corcho](#), UPM  
Esteban Garcia Cuesta, ISOCO  
[Stian Soiland-Reyes](#), University of M

This document is licensed under a [Creative Commons](#)

### Abstract

This document is to provide an accessible can understand “what” the RO Model project can be used to describe an aggregation structured format.

### Status of This Document

This document is merely a public working standing of any kind and does not represent an organisation.



W3C Community and Business Groups

Search blogs

CURRENT GROUPS    REPORTS    ABOUT

Research Object for...

## Research Object for Scholarly Communication Community Group

Research investigations are increasingly collaborative and require “borrowing strength” from the outputs of other research. Conventional digital publications are becoming less sufficient for the scientists to access, share, communicate, and enable the reuse of scientific outputs. The need to have a community-wide container data model to encapsulate the actual research data and methods with all the contextual information essential for interpreting and reusing them is becoming more and more imperative, for the science, publisher, as well as digital library communities.

A number of different community groups and projects are now creating some form of container, bundling or aggregation mechanism (particularly using ORE OAI), partially driven by the above goal. There is a clear need and benefit to facilitate a consensus among these representations. In the ROSC community group we aim to provide an open platform for gathering and discussing current development of various container models and their implementations. These data models should be driven by the need of facilitating the reuse and exchange of the actual digital knowledge and the inspection of the reproducibility of scientific investigation results.

Participants

No Chairs currently choosing chairs.

- Mailing List
- Wiki
- Chat
- RSS
- Contact Group

[www.researchobject.org](http://www.researchobject.org)

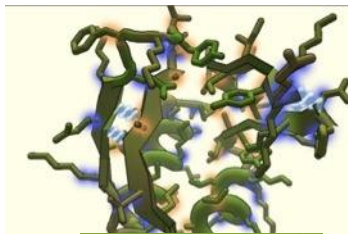
Join the W3C Community Group [www.w3.org/community/rosc](http://www.w3.org/community/rosc)

# The Order of Social Machines

Real life is and must be full of all kinds of social constraint – the very processes from which society arises. Computers can help if we use them to create abstract social machines on the Web: processes in which the people do the creative work and the machine does the administration... The stage is set for an evolutionary growth of new social engines.

Berners-Lee, *Weaving the Web*, 1999

# Some Social Machines



WIKIPEDIA

amazon mechanical turk  
Artificial Artificial Intelligence

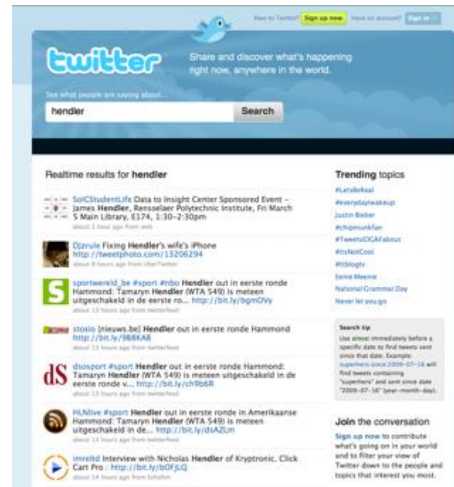
Find an interesting task

Work

Earn money



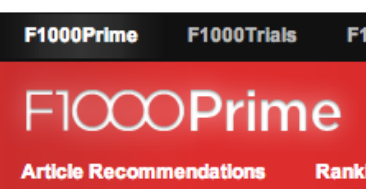
What's the score at the Bodleian?



DAY 1: STATE of the ART TO  
**BEYOND**



**FORCE11**



Beyond the PDF<sup>2</sup>  
**Scholarly  
Machines  
Ecosystem**

# Discussion points

1. Citation in tomorrow's sense-making network of humans and machines:
  - What are the artefacts / social objects?
  - How and why are they cited?
2. Think about an ecosystem of interacting Scholarly Social Machines
3. Science as Social Computation?

*Thanks to Jun Zhao, Kevin Page, my{Experiment,Grid}, wf4ever*

**david.deroure@oerc.ox.ac.uk**

[www.oerc.ox.ac.uk/people/dder](http://www.oerc.ox.ac.uk/people/dder)

<http://www.scilogs.com/eresearch>

@dder



[www.digitalsocialresearch.net](http://www.digitalsocialresearch.net)

[digital.humanities.ox.ac.uk](http://digital.humanities.ox.ac.uk)

[www.researchobject.org](http://www.researchobject.org)

[www.software.ac.uk](http://www.software.ac.uk)

[webscience.org](http://webscience.org)