Collection, Curation, Citation at Source: Publication@Source 10 years on

Jeremy G. Frey, Simon J. Coles,
Colin L. Bird, and Cerys Willoughby

University of Southampton
A Chemist’s Digital Aura

Jeremy Frey, University of Southampton
"The like frailties are to be found in the Memory; we often let many things slip away from us, which deserve to be retain'd, and of those which we treasure up, a great part is either frivolous or false; and if good, and substantial, either in tract of time obliterated, or at best so overwhelmed and buried under more frothy notions, that when there is need of them, they are in vain sought for."

“The next remedies in this universal cure of the Mind are to be applied to the Memory, and they are to consist of such Directions as may inform us, what things are best to be stor'd up for our purpose, and which is the best way of so disposing them, that they may not only be kept in safety, but ready and convenient, to be at any time produc'd for use, as occasion shall require.

Robert Hooke, *Micrographia* 1665
10 years back

End-to-end linking
Knowledge cycle
Collaboration
Semantics

metadata

CombeChem
e-Science is about global collaboration in key areas of science, and the next generation of infrastructure that will enable it.

[Sir John Taylor, 1999]
Science as an open enterprise: open data for open science

Open Access to papers is perhaps an economic argument. Access to the data is fundamental to the progress of science.
Bringing dissemination in to the lab... use and re-use of data... the digital research notebook
Dissemination is my responsibility

But ... capture @ source remains an issue ...

Bench → Data/Information value pathway → Publication

Requires high quality, semantically rich, metadata
Finding the “Wood”
From the “Trees”

http://upload.wikimedia.org/wikipedia/commons/4/41/Looking_through_the_Trees_in_Spearywell_Wood_-_geograph.org.uk-_424274.jpg
Organized Data
- Metadata
- Linking
- Networks/Graphs

05/12/2014
Beyond Open Access
Metadata is the currency of collection, curation, and citation

- Definitions can, and do, depend on one’s perspective.
- Metadata is indispensable to the lifecycle of data, information, and knowledge.
- ‘The most important reason to invest time and energy in developing metadata is that human memory is short.’
  [Michener et al, 1997 and Hooke 1665]
- Semantic metadata can be exploited in all three phases of a typical research project: planning (discovery), enactment, and dissemination.
A large proportion of our users are not really adding metadata!
Metadata

- User Created

- Metadata Sentence

  - Attempt to get verbs (processes) as well as nouns (things)

http://en.wikipedia.org/wiki/Toki_Pona
Metadata – Templates – Sharing - Usability

How much progress have we made?
Scientific Data Value Chain

- Does the Web currently/have the potential to disrupt or support the data value chain?

http://www.dreamingnewmexico.org/food/ff-local-foodshed
Responsibilities

Dissemination is my responsibility

Continuing need for more and better tools
“Engrooved*”

* concept highlighted in teaching by Paul Trowler

– http://www.brad.ac.uk/sustainable-universities/media/sustainableuniversities/Key-note-Prof-Paul-Trowler.pdf

http://www.todayandtomorrow.net/2010/02/22/quarter-mile-groove/

05/12/2014

Beyound Open Access
Reducing and Managing Uncertainty
Reducing Uncertainty

- What is already known?
- Who is doing what?
- How was a result obtained?
- Access to Data
- Integration of Data.

- Provenance & Planning
Methods are as important as the data
LabTrove: Easy Communication

Highly inter-linked report
http://997waystobegreatspeaker.com/2010/12/how-to-tell-a-story/
Skills we need are story telling skills not data mining skills

Clive Humby
DE/2014 Keynote

“We have lots of information technology. We just don’t have any information.”

What is the story? What is the how and the why?

THE NARRATIVE
Future science depends on preserving, maintaining and adding value to the research record.
Change in the whole way we design and build...

Digital fabrication is so much more than 3D printing

3D Printers: A radical change to the experimental section of a paper!
THE AUTOMATED LAB

Start-up firms say robotics and software that autonomously record every detail of an experiment can transform the efficiency and reliability of research.
All I am saying is that now is the time to develop the technology to deflect an asteroid.
Acknowledgements

- **People**
  - Colleagues and students from Chemistry, ECS, Maths
  - Oxford, UKOLN, RAL
  - IBM, Microsoft, ustwo,

- **Funding**
  - This work was funded by the RCUK e-Science programme
  - the EPSRC National Crystallography Service,
  - the HEFCE and JISC Data Management Programme and the University Modernisation (UMF),
  - and the RCUK Digital Economy Theme, IT as a Utility Network+ (EPSRC EP/K003569).
Trust me Mort - no electronic communications superhighway, no matter how vast and sophisticated, will ever replace the art of the schmooze.