

E-Science to Open Science

Collaboration around Data

Publishing Better Science Through Better Data 2015

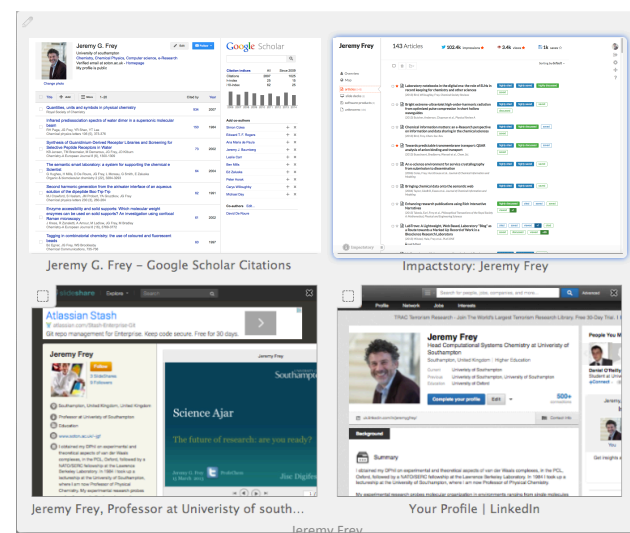
Jeremy Frey

08/11/15

Jeremy Frey

1

A Chemist's Digital Aura

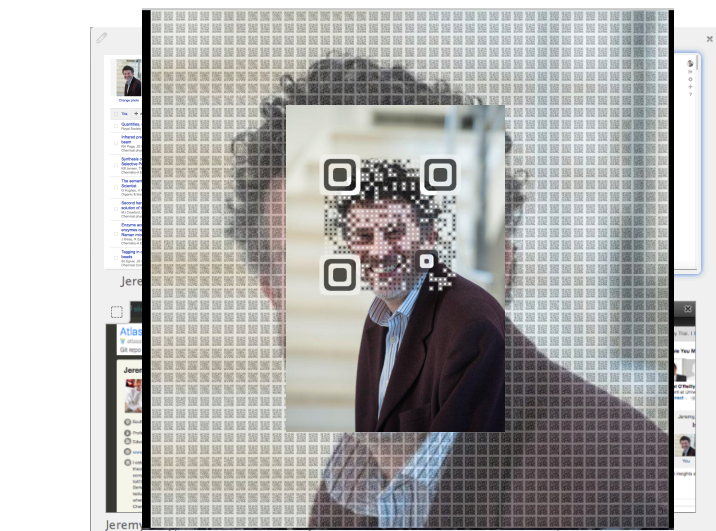


08/11/15

Jeremy Frey

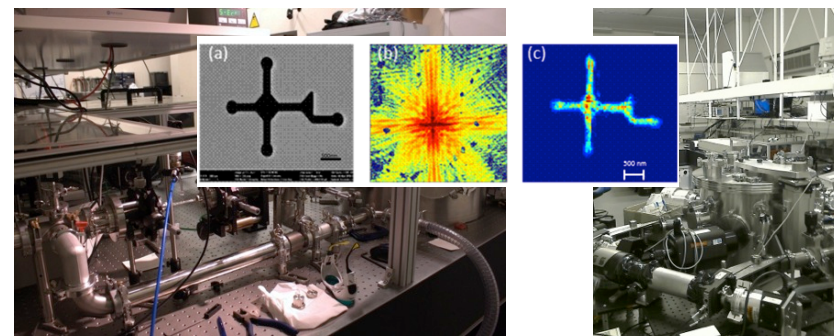
2

A Chemist's Digital Aura



<http://mosaically.com/PhotoMosaic/3289771f-0bf3-49bf-bc8a-1810167096ff>

3



08/11/15

Jeremy Frey

4

http://commons.wikimedia.org/wiki/File:ISS-30_Ice_floes_along_the_Kamchatka_coastline.jpg



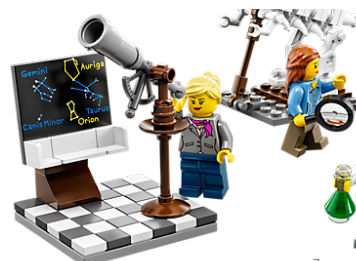
08/11/15

Jeremy Frey

5

Reducing Uncertainty

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



08/11/15

Jeremy Frey

7

7342.pdf (6 pages)

Water surface is acidic

Victoria Buch*, Anne Milet[†], Robert Váchá[‡], Pavel Jungeth[§], and J. Paul Devlin[¶]

ScienceDaily
Your source for the latest research news

December 19, 2006

used. First, constant volume and molecular dynamics simulations of water with a pair of H₃O⁺ and OH⁻, the simulation was carried out for boundaries and a slab with two open boundaries. This "on-the-fly" technique modeling of proton transfer systems on transfer and transitions between different forms are automatically included. However, the high computational cost of the model system, the duration

Featured Research
from universities, journals, and other organizations

Rewrite the textbooks on water's surface tension: Air-water interface is negatively charged by the adsorption of hydroxide ions

Date: March 19, 2014
Source: University of Melbourne
Summary: Researchers in Australia make significant differences in the next generation of biophysicists and engineers. The physical Chemistry of charged by the adsorption of hydroxide ions.

Related Topics
Matter & Energy
Nature of Water
Chemistry
Inorganic Chemistry
Thermodynamics
Biochemistry
Nuclear Energy

PAPER
www.rsc.org/faraday_d | Faraday Discussion

The surface of neat water is basic

James K. Beattie,^{*} Alex M. Djerdjef and Gregory G. Warr

Received 28th March 2008, Accepted 28th May 2008
First published as an Advance Article on the web 18th September 2008
DOI: 10.1039/b805266b

08/11/15

6

Intelligent (Open) Access to Data

THE ROYAL SOCIETY

Policy | Projects | Reports | Publications | PolicyLab meetings | Blog | more | search

Science as an open enterprise Final report

21 June 2012

The Science as an open enterprise report highlights the need to grapple with the huge deluge of data created by modern technologies in order to preserve the principle of openness and to exploit data in ways that have the potential to create a second open science revolution. Exploring massive amounts of data using modern digital technologies has enormous potential for science and to application in public policy and business. The report maps out the changes that are required by scientists, their institutions and those that fund and support science if this potential is to be realized.

Areas for action

- Six key areas for action are highlighted in the report:
- Scientists need to be more open among themselves and with the public and media
- Greater recognition needs to be given to the value of data gathering, analysis and communication
- Common standards for sharing information are required to make it widely usable

Download
Final report
PDF, 8.4 MB
Additional downloads
Summary
PDF, 4.7 MB
EPUB and Kindle version
ZIP, 4.7 MB
References (B6Tab)
ZIP, 1.1 MB

Science as an open enterprise
Final report: case studies of data use and data repositories and the launch event
published June 2012
Public meeting and seminar
held in November 2011
Call for evidence
closed August 2011

<https://royalsociety.org/policy/projects/science-public-enterprise/Report/>



08/11/15

Jeremy Frey

8

Confusion over open-data rules

The Public Library of Science's pioneering open-data mandate has prompted scientists to share more data online, but not everyone is complying with the regulation.

BY RICHARD VAN NOORDEN

The mantra of the nascent open-data movement — that scientists should share online all data underlying their findings — sounds simple. But it can be tough to achieve in practice. An informal audit of one of the movement's biggest proponents, the Public Library of Science (PLOS), shows that not everyone is complying with the publisher's pioneering open-data mandate, and hints at the challenges that journals can face in enforcing open-data goals.

The idea that the progress of research will be accelerated if others can easily and freely build on data sets is gaining currency. Last week the Bill & Melinda Gates Foundation in Seattle, Washington, announced that it would demand open data of the researchers it funds.

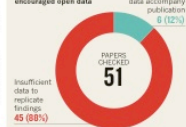
But whereas some research communities, such as geneticists and crystallographers, have long-established norms of open data, most funders and publishers (including *Nature*), mindful of researcher autonomy, merely exhort scientists to make their data open. Many surveys have found that scientists are worried about being scooped on future projects, or argue that they have signed agreements not to share their data.

So it was a step-change when in March PLOS made it a requirement that authors who publish in its journals share online all the data necessary to reproduce their studies. It was not

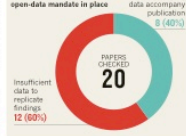
FREE THE DATA

In 2014, open-access publisher PLOS introduced a requirement that authors who publish in its journals make their underlying data freely available online. An informal audit of one type of population genetics study in one journal, PLOS ONE, shows that not everyone is complying — but the mandate is still a boon for the open-data movement.

2011-12, when PLOS encouraged open data



Since March 2014, after open-data mandate in place



job of external peer reviewers to check whether appropriate underlying data were available, he says. And although PLOS ONE was grateful for their help, "there is a learning curve here for all involved to understand what the data sharing standards are for all disciplines and data types". He adds that once someone complains, the journal has a system for investigating papers that do not comply with its open-data requirement.

The research teams in question, nine of whom responded to the *Nature* news team's request for an explanation, provide an insight into why data sharing does not always happen in the first place (full data are available at go.nature.com/8ggob6). Some had forgotten to upload their data and promptly rectified the fault. Four teams said that the journal's editors and referees had never asked them to share the underlying genetics data, suggesting confusion over what the policy means.

Others aired wider — and common — objections to online data sharing. Even though they had chosen to publish with PLOS, some authors said that they wanted to hold back their data for future studies, or did not want to share the raw data unless they knew future users' intent.

Steve Simpson at the University of Exeter, UK, who reported work on Omani clovefish, said he was happy to share raw data privately with potential collaborators, but not to upload results from 400 individual fish that had taken great effort to collect. "The study is described so that it could be replicated by another exper-

<http://www.nature.com/news/confusion-over-publisher-s-pioneering-open-data-rules-1.16409>



08/11/15

Jeremy Frey

9



Data is the new Spice



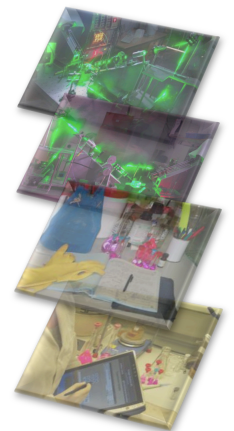
08/11/15

Jeremy Frey

11

"The Laboratory, The Internet and the Web"

- A laboratory is an Integration of people and equipment
- A complex socio-technical system
- Disrupted or integrated by digital infrastructure?



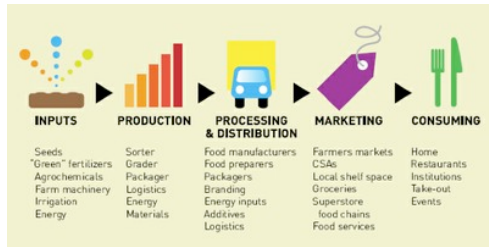
08/11/15

Jeremy Frey

12

Scientific Data Value Chain

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



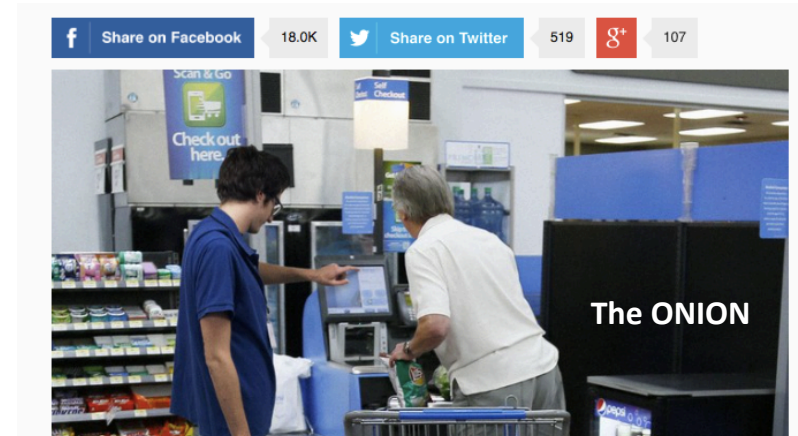
<http://www.dreamingnewmexico.org/food/ff-local-foodshed>

Jeremy Frey

13

08/11/15

IT (Information Transformation)



08/11/15

Jeremy Frey

14



<http://997waystobeagreatspeaker.com/2010/12/how-to-tell-a-story/>

Jeremy Frey

15

08/11/15

*Skills we need are
story telling skills
not data mining
skills*

Clive Humby
DE/2014 Keynote

© Cartoonbank.com



"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

08/11/15

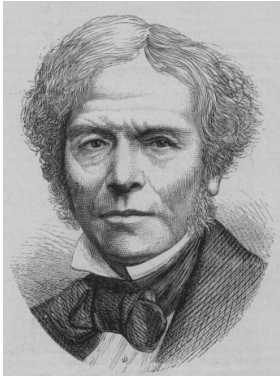
Jeremy Frey

16

ON BEING A SCIENTIST

A GUIDE TO RESPONSIBLE CONDUCT IN RESEARCH

T H I R D E D I T I O N



08/11/15

Faraday's laboratory notebooks are also remarkable in the amount of detail that they give about the design and setting up of experiments, interspersed with comments about their outcome and thoughts of a more philosophical kind. All are couched in plain language, with many vivid phrases of delightful spontaneity....

Peter Day, 'The Philosopher's Tree: A Selection of Michael Faraday's Writings'

Jeremy Frey

17



08/11/15

Jeremy Frey

18

Methods are as important as the data



BBC News - 'Show Your Working': What 'ClimateGate' means

NEWS

Page last updated at 14:56 GMT, Tuesday, 1 December 2009

E-mail this to a friend

Printable version

'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.

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

Reviving Rio
The second Earth Summit holds the promise to cure the world's environmental ills

» Your comments

RECENT ARTICLES

- » EU must increase climate ambition
- » Counting the cost of alien invasions
- » Has CITES had its day?
- » Curious case of Kiwi hedgehogs
- » Restoring natural capital
- » It's still real and it's still a problem
- » Hailing the arrival of alien predators

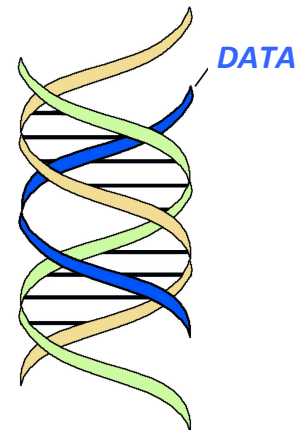
08/11/15

Jeremy Frey

<http://news.bbc.co.uk/1/hi/technology/8388485.stm>

19

Theory & Experiment



08/11/15

Jeremy Frey

20

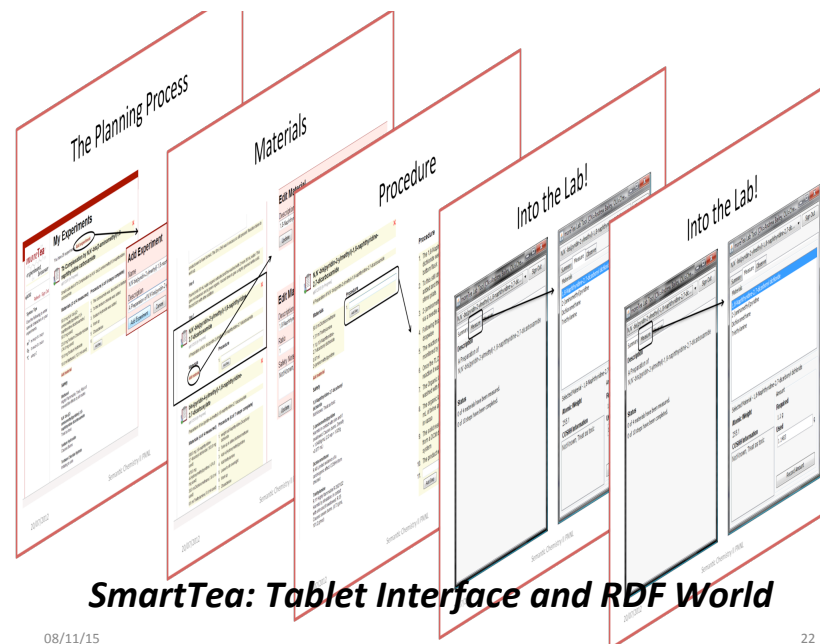
Digital Research Notebooks

Note taking

08/11/15

Jeremy Frey

21



SmartTea: Tablet Interface and RDF World

08/11/15

22

The screenshot displays the LabTrove interface with a detailed report on the transformation of plasmid p042 via Taq and Genomorphi. The report includes a table of results and a section for 'Test digests to check the activity of two batches of EcoRI and NcoI'.

Images and sketch commenting

LabTrove: Easy Communication
www.labtrove.org

Highly inter-linked report

The screenshot shows the Luke Marazzi's Y3 Chemistry Project website. It features a video titled 'TLT660/4b Hot Stage Video' and a sidebar with navigation links.

Digital and Physical



Links to other notebook pages

CLOUDS

Comments
: Symmetry properties o
n August 2015 @ 15:22

OneNote Online

Research Journal 2015

FILE HOME INSERT VIEW PRINT Tell me what you want to do OPEN IN ONENOTE GIVE FEEDBACK TO MICROSOFT

Find on this Page (Ctrl F)

Journal

reviews

theory exploits

uppl

ICAN 2015

CLI

Quick Notes

EXTATIC

satellites

Nobel lecture

papers

talks

Testing

RPLN phase matching along dZ

More generally for ordinary and extraordinary index for MgO doped LNBO3:

$$n^2 = a_1 + b_1 f + \frac{a_2 + b_2 f}{\lambda^2 - (a_3 + b_3 f)^2} + \lambda^2 - a_5 - a_6 \lambda^2$$

with:

Parameters	5% MgO doped CLN	1% MgO doped SLN
a_1	5.756	5.653
a_2	0.0983	0.1185
a_3	0.2020	0.2091
a_4	189.32	89.61
a_5	12.52	10.85
a_6	1.32×10^{-4}	1.57×10^{-4}
b_1	2.860×10^{-4}	7.941×10^{-7}
b_2	4.700×10^{-4}	3.134×10^{-4}
b_3	6.113×10^{-4}	-4.941×10^{-4}
b_4	1.516×10^{-4}	-2.388×10^{-4}

for congruent LNBO3 (CLN) and stoichiometric LNBO3 (SLN)

So what's the phase mismatch?

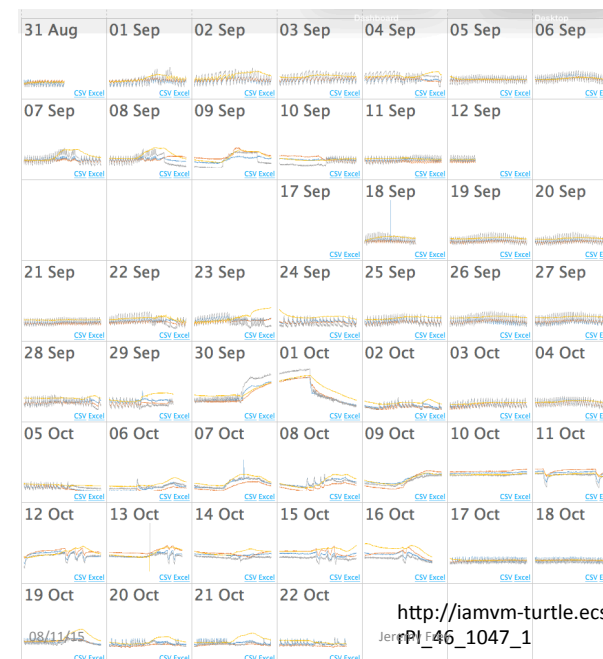
$$\Delta k = \frac{2\pi n}{\lambda} \Rightarrow \Delta k = \frac{2\pi n_{\text{exp}}}{\lambda_{\text{exp}}} - \frac{2\pi n_{\text{theo}}}{\lambda_{\text{theo}}}$$

$$\Delta k = \frac{2\pi n_{\text{exp}}}{\lambda_{\text{exp}}} - \frac{2\pi n_{\text{theo}}}{\lambda_{\text{theo}}}$$

08/11/15

Jeremy Frey

29



http://iamvm-turtle.ecs.soton.ac.uk/devices/rpi_46_1047_1

30



sharing

08/11/15

Jeremy Frey

31

Chemistry Central Journal

IMPACT FACTOR 2.19

Search [this journal] for [Go]

Advanced search

Login

All Notebooks | Help | Support | About

LabTrove

Synthesis of 4-substituted methyldene oxindole

On cooling the resultant solid was filtered, dried under reduced pressure and then purified by sublimation to give the product. Full details of experimental processes and observations of the synthesis of 3a-e are available at http://pubs.labtrove.soton.ac.uk/synth_methyl_oxindole/Experimental%20Procedure [website].

Additional file 1. DeepZoom: 4-methyldene oxindole synthesis and characterization. Microsoft Silverlight plug-in is required to view this. Use + and - buttons or scroll wheel to zoom in and out of image to view individual spectra. Use home button to reset view to full image. Click table cells to follow link to relevant section of electronic lab notebook for full supporting data (spectral assignments etc). Format: ZIP Size: 37.1MB Download file

	IR	¹ H NMR	¹³ C NMR (inside zoom)	HPLC	MS	SXRD	PXRD
3a							
3b							
3c							
3d							
3e							

Amount of Chemical solvent used Oxindole (g) 4-chlorobenzonitrile

08/11/15

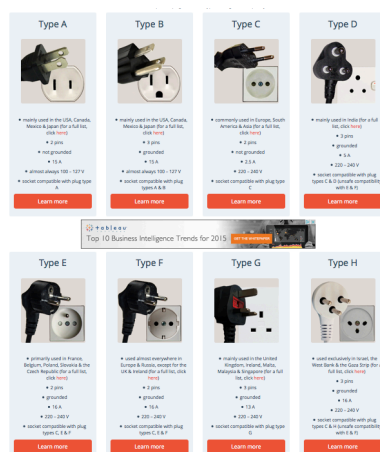
Jeremy Frey

32

The Grid.....Standards!



08/11/15



Jeremy Frey

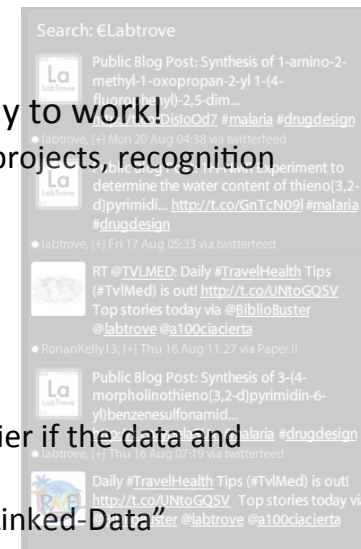
33

Open Notebook Science

- Certainly not always the way to work!
 - IPR, Commercial, long term projects, recognition issues, etc

*It is not necessary to change.
Survival is not mandatory.
W. Edwards Deming*

- But
 - Makes connection much easier if the data and processes are “Open”
 - Easy to export & access of “Linked-Data”



08/11/15

Jeremy Frey

34



LabTrove Open Notebooks Mat Todd's Malaria Project

Praziquantel - Open Science

Pictet-Spengler route to Praziquantel

Racemic Resolution of Praziquantel and Praziquanamine

Hydrolysis of PZQ

Monitoring the Reaction Progress by TLC

Racemic Resolution of Praziquantel and Praziquanamine

Hydrolysis of PZQ

Monitoring the Reaction Progress by TLC

08/11/15

Jeremy Frey

35



08/11/15

Jeremy Frey

36

Libraries

Effects on the Institution, Building and Personnel

- The replacement of traditional services by IT utilities;
- Opportunities to exploit IT to provide new value-add services;
- The potential for changing and new roles for librarians, such as information management, curation, and training.



http://en.wikipedia.org/wiki/George_Peabody_Library#mediaviewer/File:George-peabody-library.jpg

08/11/15

Jeremy Frey

37

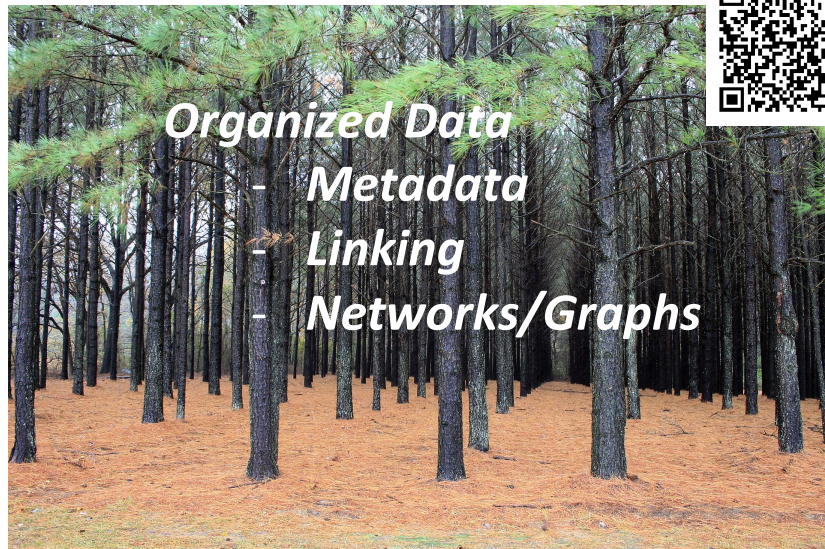


http://upload.wikimedia.org/wikipedia/commons/4/41/Looking_through_the_Trees_in_Spearywell_Wood_-_geograph.org.uk_-_424274.jpg

08/11/15

Jeremy Frey

38



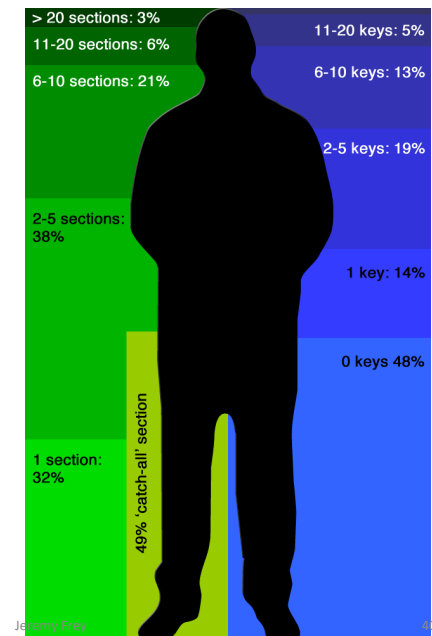
http://en.wikipedia.org/wiki/Plantation#mediaviewer/File:Pinus_taeda_plantation.jpg

08/11/15

Jeremy Frey

39

A big proportion of our users are not really adding metadata!



Jeremy Frey

08/11/15

40

Chemistry blogs results



50 chemistry-related blogs – 10,436 ‘tags’

08/11/15

Jeremy Frev

41

NASA results



08/11/15

Jeremy Frev

53 NASA blogs - 1226 'tags'



Plans and Provenance

Jeremy Frey

Smart Tea Project



08/11/15

Jeremy Frey

Chemistry and Tea: An Interesting History

A Nice cup of Tea
- George Orwell

Chemists seek perfect cup of tea – for Orwell
– London (Reuters)

[How to make the perfect cup of tea](#) (pdf download)
- The Royal Society of Chemistry provide the "definitive recipe for a cup of tea"

Trouble is brewing
- Institute of Physicists
dispute RSC findings for
making tea

[British Standards Institute](#),
enter standard no.:
BS6008:1980 (6 pages,
£22.00)
[Guardian story \(and
abbreviated version\)](#) of this
BSI prize winning document.

Counterfeit tea (or smouch)
- The BBC (and Michael Quinion for the explanation of smouch)

44



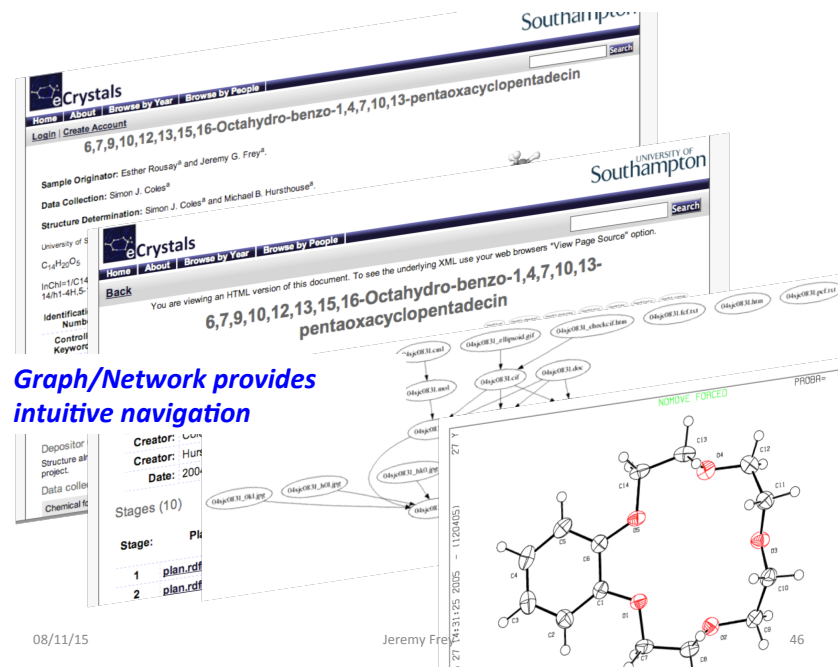
"In theory, there is no difference between theory and practice. But, in practice, there is."

Unknown (possibly Yogi Berra or Manfred Eigen)

08/11/15

Jeremy Frey

45



Graph/Network provides intuitive navigation

08/11/15

Jeremy Frey

46

Dissemination is part of the research
But beware of "Engrooved* behaviour

- * concept highlighted in teaching by Paul Trowler
 - <http://www.brad.ac.uk/sustainable-universities/media/sustainableuniversities/Keynote-Prof-Paul-Trowler.pdf>



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

08/11/15

Jeremy Frey

47

The System

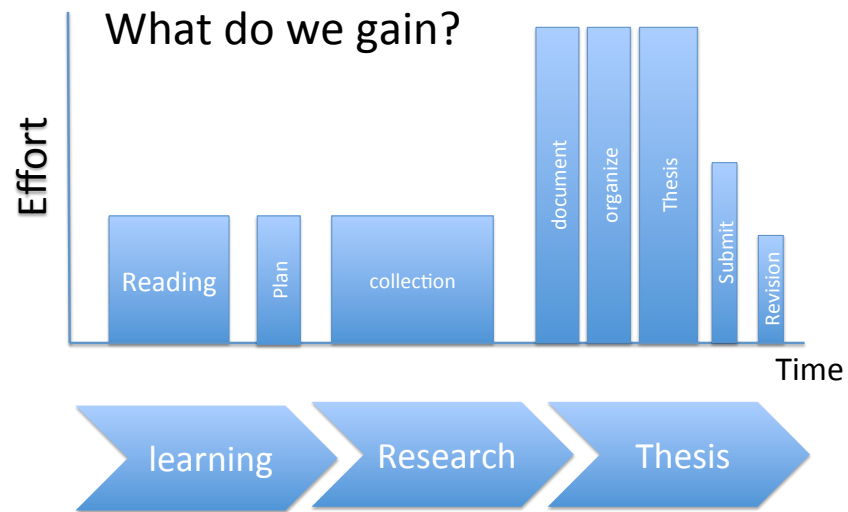
adapted from Atul Gawande 2014 Reith Lectures

- "Primitive idea"
 - You ought to do x.
- Medieval approach
 - You **must** do x! You must follow the standards
 - Will take away your grant....!
- Modern world(?)
 - Becomes the norm, make it easy to do
 - Checklists
 - **Feedback loops**

08/11/15

Jeremy Frey

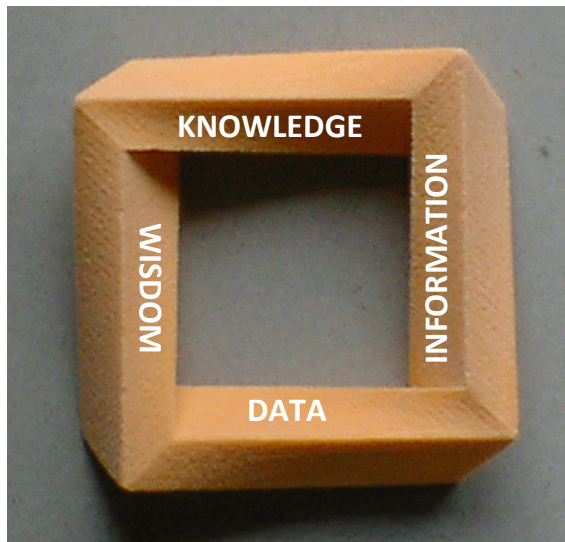
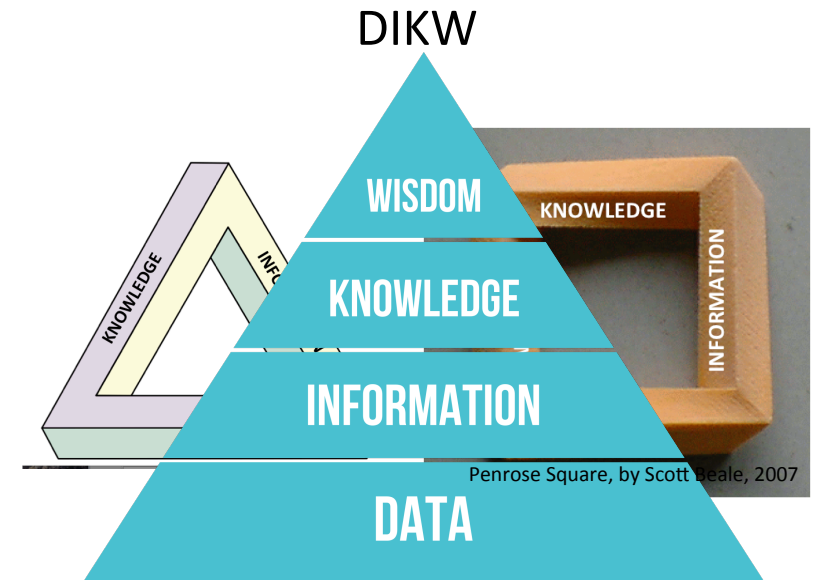
48



08/11/15

Jeremy Frey

49



<http://usuaris.tinet.org/klunn/polyhedra.html>



08/11/15

Jeremy Frey

52

Acknowledgements

- My research group
 - Chemists & Computer Scientists
 - UG project students, PG, and Post-Docs
 - Visitors
- My colleagues
 - Chemistry, Physics, ORC, ECS, Engineering, Statistics, Mathematics
 - Southampton, Oxford, QMUL, Ulster, Penn State,
- Funding
 - University of Southampton, Chemistry
 - EPSRC, RCUK, JISC/Jisc, IBM, Microsoft, Innovate UK, Lonza
 - Horizon 2020
- Haiku for help with images

08/11/15

Jeremy Frey

53

The Lego Movie



*Everything is Awesome,
Everything is cool when you're part of a team*



08/11/15

Jeremy Frey

54

Thank you for listening



Trust me Mort - no electronic communications
superhighway, no matter how vast and sophisticated,
will ever replace the art of the schmooze

08/11/15

Jeremy Frey

55