

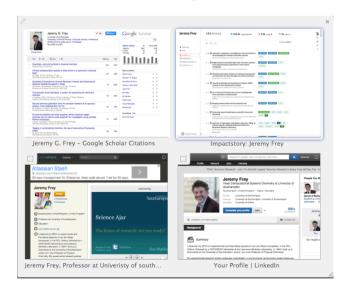
## **Digital Chemistry**

Reducing and Managing Uncertainty

Chemistry 2.0 (or 3.0)



#### A Chemist's Digital Aura





"We have lots of information technology. We just don't have any information."

## Google "Digital Chemistry"



But also note that my search results will probably not be the same as yours!

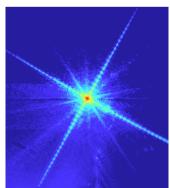
#### Outline

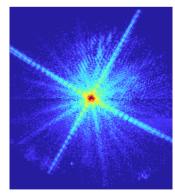
- Introduction
- Computational Chemistry and Computational Support for Chemistry Researchers
- Open Access vs Intelligent Access to Data
- Digital Notebooks & Data Citation
- The Future



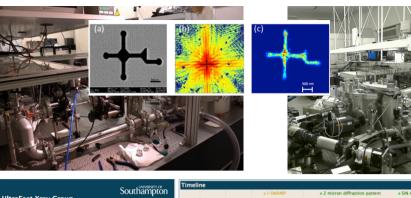


## Coherent diffraction Imaging for lensless soft x-ray microscopy





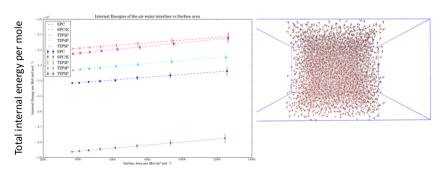
Key is computational phase reconstruction, which using GPU, can be done in time comparable to the experimental acquisition time. GPU cluster will enable ensemble solution and thus estimate the uncertainty in the reconstructing, but would need on demand.





/11/14 The future of laboratory data

## Air/Water slab MD simulation

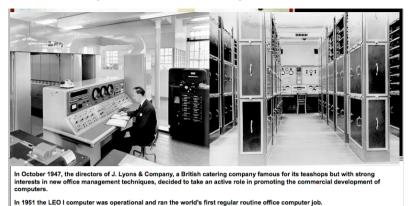


Surface area per molecule

Computational power now available means that sufficient accuracy and low uncertainty can be obtained for a whole series of small and large water slabs allowing the surface contribution to the total energy to be reliably "observed" and entropy contributions to surface tension derived.

#### Lyons Electronic Office (LEO) 1951

#### *Information Transformation*



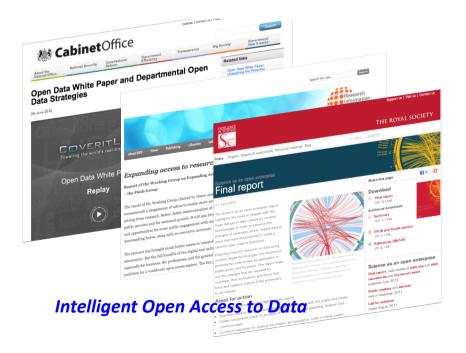
## Collaboration

- Research is increasingly inter-disciplinary
- Need help in information exchange
- If is it not available to a computer then data is essentially lost

## **Reducing Uncertainty**

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











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....

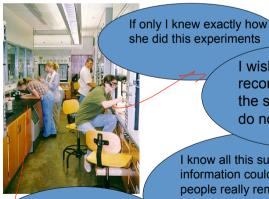
Peter Day, "The Philosopher's Tree: A Selection of Michael Faraday's Writings'



"We have lots of information technology. We just don't have any information."

What is the story? What is the why?

#### THE RESEARCH NARRATIVE



I wish I could get the numbers from this graph - the pdf is not much use. I wish I had recorded things at the start the way I do now....

I know all this supplementary information could be useful but will people really remember the format? Is it worth all the hassle?

Typical Laboratory

## Archiving of data

- Experiments are often repeated
  - Data stored locally on a computer and can't be found
  - Handwriting can't be read
  - Laboratory notebooks lost or damaged
  - Correct data not recorded first time round



http://www.ecs.soton.ac.uk/regenesis/picture

#### Transformation of plasmid JRH4712/66 into BW25141 by electroporation

Transformations were set up according to the following protocol: LB Ampicillin arabinose plates and SOC medium were warmed to 37  $^{\circ}\text{C}$  briefly before the arabinose plates were spread with X-glu (80  $\mu\text{L}, 1:1$  X-glu and LB) and allowed to continue warming.

BW25141 cells, plasmid JRH4712/66, p042, and electroporator cuvettes

-	1	+ve ctrl	-ve ctrl	
BW25141	40 µL	40 µL	40 µL	
plasmid 4712/66	4 µL	0 μL	0 μL	
p042	0 μL	4 μL	0 μL	

Cuvettes were electroporated at 1.75 kV, immediately had SOC medium (950  $\mu$ L) added and the transformant transferred to eppendorf. The transformants were incubated at 37  $^{\circ}$  C for one hour with shaking. The transformants were diluted 1 in 20 with LB and 100  $\mu$ L added to LB amp arabinose plates and incubated at 37  $^{\circ}$  Covernight.



Jennifer Hale | Beta-glucuronidase | Comments (3)

Archives
January 2007 (24)
December 2006 (11)
November 2006 (5)
Sections
beta-galactosidase preparation and
assays (18)
Beta-glucuronidase (18)
Data (Formatting) (1)
Software discussions (2)
Starting materials and reagents (1)
Lab Book Ref
JRH4712-63 (1)
JRH4712-64 (2)
JRH4712-66 (1)
et digastions to shock the

Test digestions to check the activity of two batches of EcoRI and Ncol

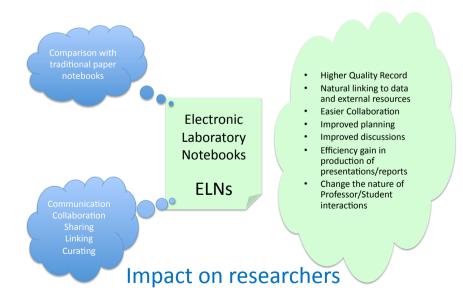
Digestions w		2 up a			5	6	7	8	9	10	11
4712/80 blue	8 µL	-	-	8	-	-	-	8 µL		-	-
4712/80 white	-	8 µL	-	-	8 µL	-	-	-	8 µL	-	-
p042	-	-	5 µL	-	-	5 µL	5 µL	-	-	5 µL	5 μ
water	7.5 µL	7.5 µL			7.5 µL	10.5 µL	10 µL	7.5 µL		10.5 µL	10 µL
EcoRI buffer	2 µL	2 µL	2 µL	-	-	-	2 µL	2 µL	2 µL	2 µL	2 μ
NEB buffer 4	-	-	-	2 µL	2 µL	2 μL	-	-	-	-	-
BSA	2 µL	2 µL	2 µL	2 µL	2 μL	2 µL	2 µL	2 µL	2 µL	2 µL	2 μ
EcoRI (a)	0.5 μL		0.5 µL	-	-	-	0.5 µL	-	-	-	-
Ned	-	-	-		0.5 µL	0.5 µL	0.5 µL	-	-	-	0.5 µL
EcoRI (b)	-	-	-	-	-	-	-			0.5 µL	0.5 µL

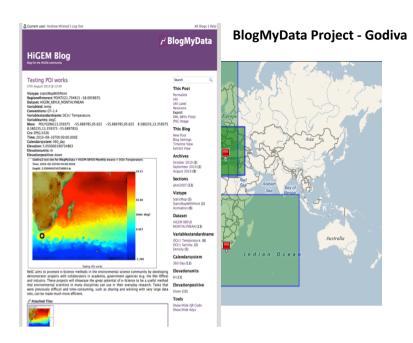
Digestions were incubated in a waterbath at 37 °C for 3 hours

## The Trove Software



www.labtrove.org





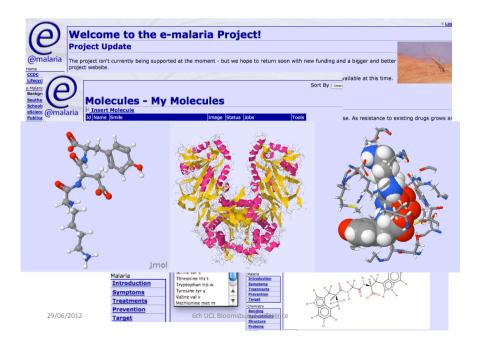


The internet has also changed how we communicate with one another — and what is known as 'Web 2.0' is here. Blogs, social-networking sites, wikis and podcasts were alien concepts not all that long ago, but they are here to stay and some of them are significantly impacting the communication of science, and chemistry is no exception. Nature Publishing Group joined the chemical blogosphere back in March 2006 when we launched 'The Sceptical Chymist' blog and 'ChemPod' — a chemistry-themed podcast if you hadn't guessed. To reflect the firm inroads that chemistry seems to have made into the blogging culture, *Nature Chemistry* will feature a monthly column called 'Blogroll' that will round up stories that have caught the attention of the chemistry bloggers. Our first topics include a citation controversy and a reminder that boron isn't boring!

#### **Nature Chemistry**

Web 2.0 & Semantic Web

#### **USER GENERATED CONTENT**



#### Citizen Science





Foldit Kaggle

#### 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

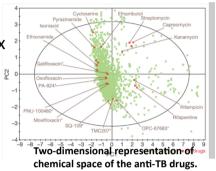
- But
  - Makes connection much easier if the data and about adjusted processes are "Open"

    Daily stravelHealth Tips (#TVIMed) is out
  - Easy to export & access of "Linked-Data" elebtrove @a100cacterta

## Big Data

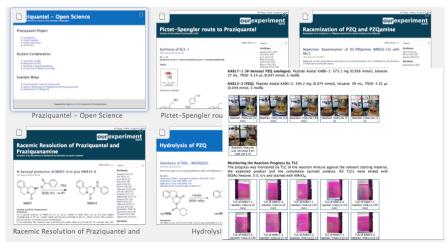
- Chemistry we have -
- Big, Broad, Bothersome, & Brocken,

High Dimensional,
Heterogeneous, Complex
Data on an uncertain
landscape as we still do
not have a really good
grasp of Chemical Space

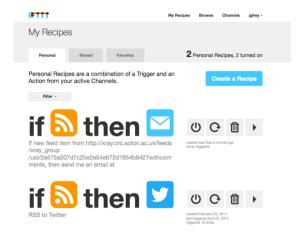




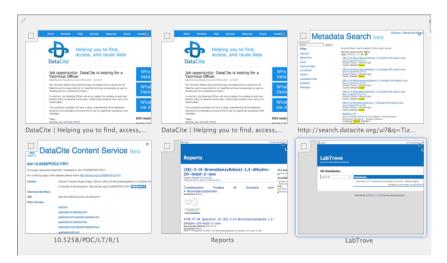
#### LabTrove Open Notebooks Mat Todd's Malaria Project

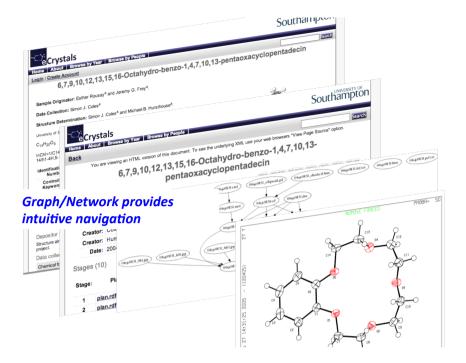


## LabTrove -> RSS -> Email / Twitter



#### DataCite DOI





#### We must speed up the knowledge discovery process



All I am saying is that now is the time to develop the technology to deflect an asteroid



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

In times of change it is even more important to bring dissemination in to the lab...

# DISSEMINATION IS PART OF THE RESEARCH

## Change in the whole way we design and build experiments



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



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