

The CombeChem Project

From Smart Laboratories to Smart Papers

Jeremy G. Frey
School of Chemistry
University of Southampton, UK



Jeremy G. Frey
University of Southampton



Topics

- Introduction to the “CombeChem” Project
- The concept of Publication @ Source
- The Laboratory
 - Smart Tea, Smart Labs & Tablets
 - Instruments on the Grid
- The ‘Blog’ as the Lab Notebook
- Dissemination
 - Repositories
 - Live Documents, Smart Papers, Meetings



27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia



Chemists and programming

Many Chemists
think that they can
program



You still use FORTRAN!!

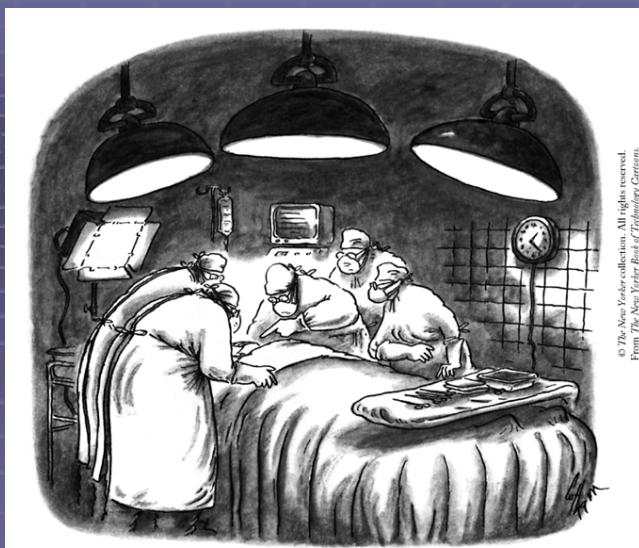
27 June 2007

Jeremy G. Frey



e-Workflow

Some
Chemists
can

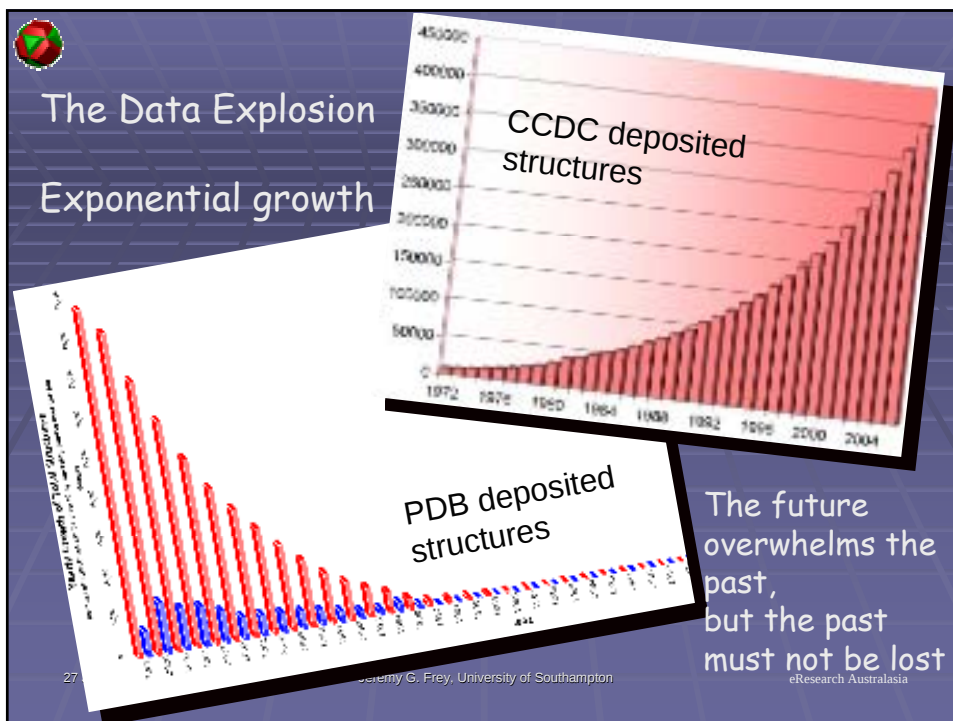


What about that! His brain still uses perl scripts

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia



The CombeChem Project

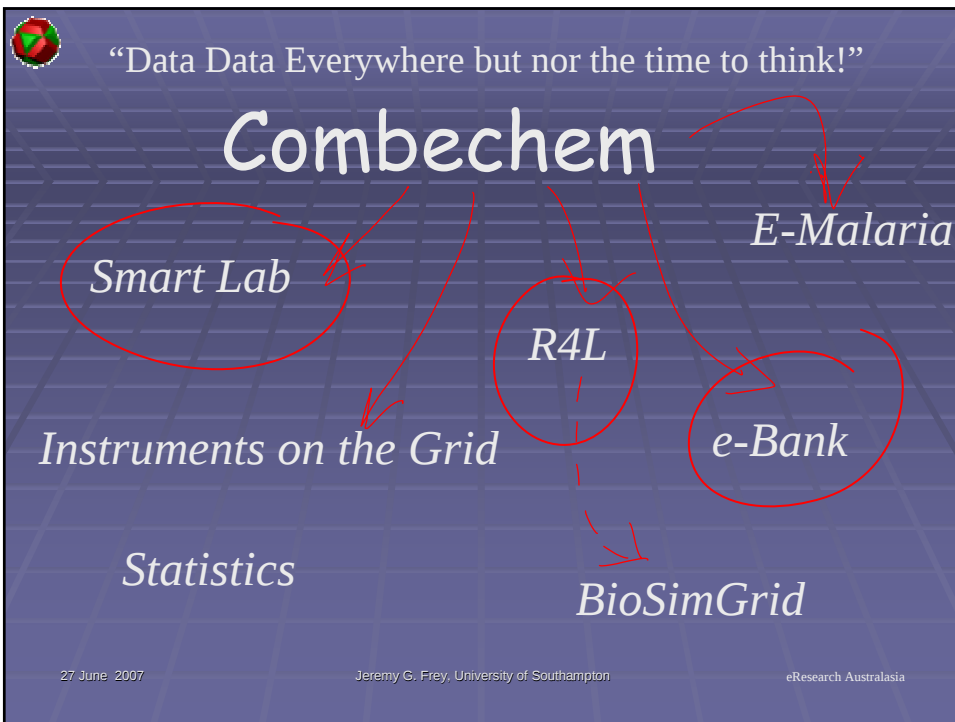
- 'End to End' linking
 - Data (life-)cycle
- Do things 'right' at the start
 - Make sure the metadata is of high quality
 - Record properly at source in Digital Form
- Extensive provenance
 - **Publication@Source**
- The Chemistry Lab
 - People & Machines working together

QuickTime™ and a TIFF (Uncompressed) decompressor are needed to see this picture.

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

27 June 2007

Jeremy G. Frey, University of Southampton



If only I knew exactly how she did this experiments

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?

I wish I could get the numbers from this graph - the pdf is not much use.

Typical laboratory conversations

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia



- Need to make the data available

- Need to be able to find it

- But how to expose it?



© The New Yorker collection. All rights reserved.
From The New Yorker Book of Technology Cartoons.

First, they do an online search

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia



Supervisors and Managers

I am sure we collected that information a few years ago...

The details should be in her thesis.....

Can you read what he says here....?

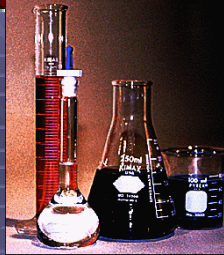
Can you find the file of data that were used to make the plot?

Some problems are due to the lack of information recorded at the time, others are due to loss of information over time.

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia



ChemLab
The Chemistry 3/5 & 6 Laboratories

- General Information
- Instruments & Techniques
- Chemistry 3/5 Experiments
- Chemistry 6 Experiments

DARTMOUTH COLLEGE

[Sample Write-up](#)

Safety

- [General Rules](#)
- [Safety Equipment](#)
- [Safety Hazards](#)
- [Emergency Procedures](#)
- [Emergency Treatments](#)

Resources

- [Applets](#)
- [General FAQ](#)
- [Uncertainty](#)

[ChemLab Home](#)

Electronic Laboratory Notebooks

Info
Techniques
Chem 3/5
Chem 6

How to Keep a Notebook

One of the most useful skills you will acquire in the laboratory is the proper use of a laboratory notebook. Notebooks, or other formally kept records, are an essential tool in many careers, ranging from that of the research scientist to that of the practicing physician. The effort invested in developing good habits of notebook use will be amply repaid for students who pursue a future in the basic or applied sciences. Experience has indicated that skillful notebook use is developed by most students only through continued special effort—it does not come naturally. Some of the main principles of sound notebook use are outlined below.

The laboratory notebook is a permanent, documented, and primary record of laboratory observations. Therefore, your notebook will be a bound journal with pages that should be numbered in advance and never torn out. A notebook will be supplied to you before the first laboratory period. Write your name, the name of your TA, and your lab section on the cover of your notebook. All notebook entries must be in ink and clearly dated. No entry is ever erased or obliterated by pen or "white out". Changes are made by drawing a single line through an entry in such a way that it can still be read and placing the new entry nearby. If it is a primary datum that is changed, a brief explanation of the change should be entered (e.g. "balance drifted" or "reading error"). No explanation is necessary if a calculation or discussion is changed; the section to be deleted is simply removed by drawing a neat "x" through it.

Permanent, documented and primary record of laboratory observations

27 June 2007



ChemLab
The Chemistry 3/5 & 6 Laboratories

- General Information
- Instruments & Techniques
- Chemistry 3/5 Experiments
- Chemistry 6 Experiments

DARTMOUTH COLLEGE

necessary if a calculation or discussion is changed; the section to be deleted is simply removed by drawing a neat "x" through it.

In view of the fact that a notebook is a primary record, data are not copied into it from other sources (such as this manual or a lab partner's notebook, in a joint experiment) without clear acknowledgment of the source. Observations are never collected on note pads, filter paper, or other temporary paper for later transfer into a notebook. If you are caught using the "scrap of paper" technique, your improperly recorded data may be confiscated by your TA or instructor at any time. It is important to develop a standard approach to using a notebook routinely as the primary receptacle of observations.

Each week at the beginning of lab lecture, you will turn in your prelab problems from the manual for grading. Problems not turned in at the beginning of lab lecture will be

Observations are never collected on note pads, filter paper or other temporary paper for later transfer into a notebook

If you are caught using the "scrap of paper" technique, your improperly recorded data may be confiscated by your TA

G. Frey, University of Southampton
eResearch Australasia



COSHH

Leverage off things we already have to do



“We have a cunning plan”

27 June 2007

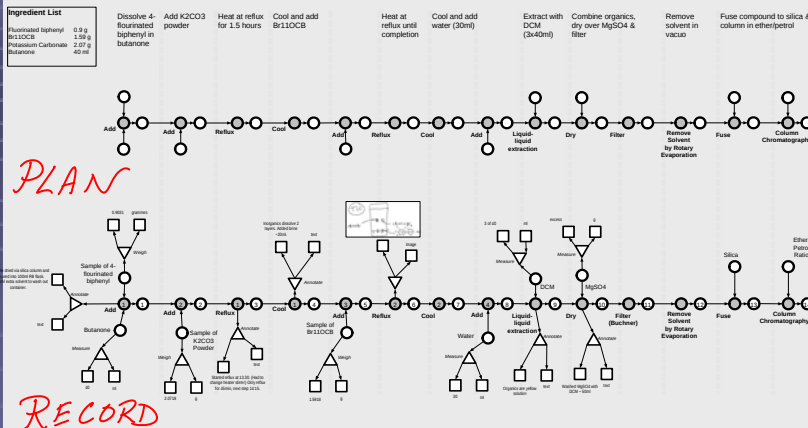
Jeremy G. Frey, University of Southampton



eResearch Australasia



TO DO LIST



27 June



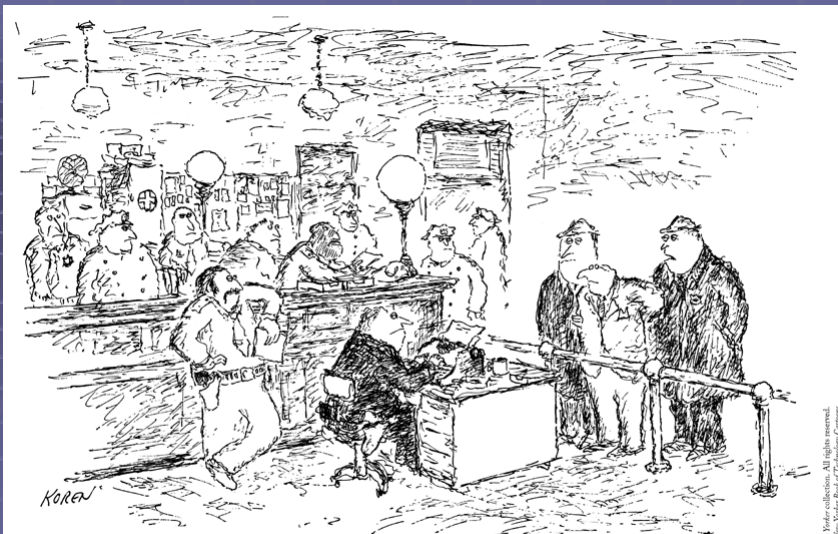
Metadata Lifecycle

- Creation and maintenance of metadata
- Need a metadata infrastructure as well as a data infrastructure
- Capture process as well as results
- Automatic metadata generation when possible
- Human annotation will always be needed

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia

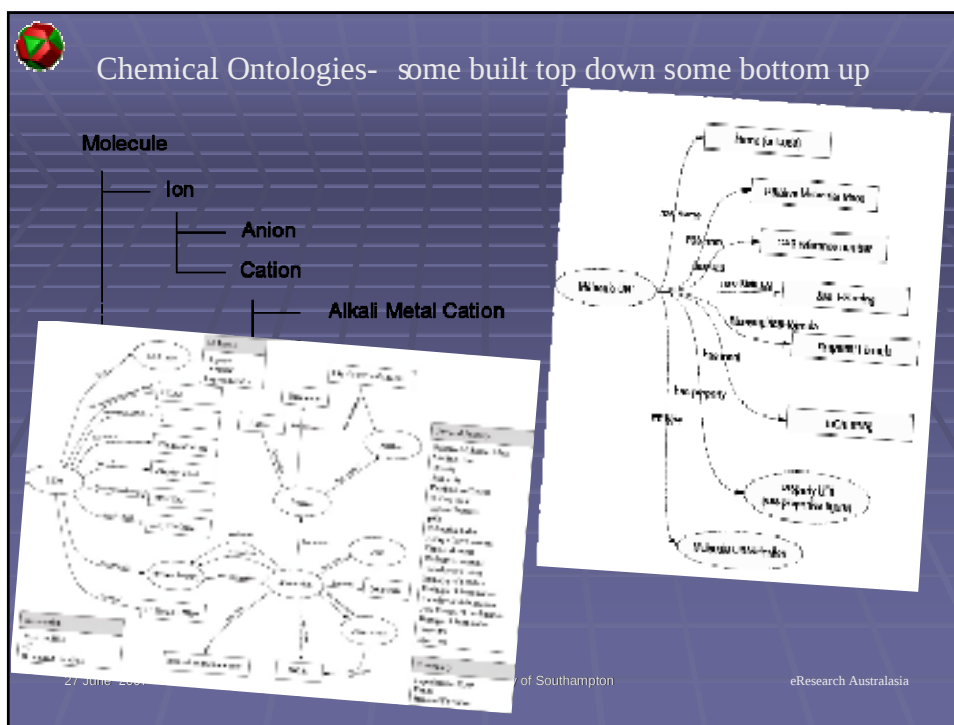
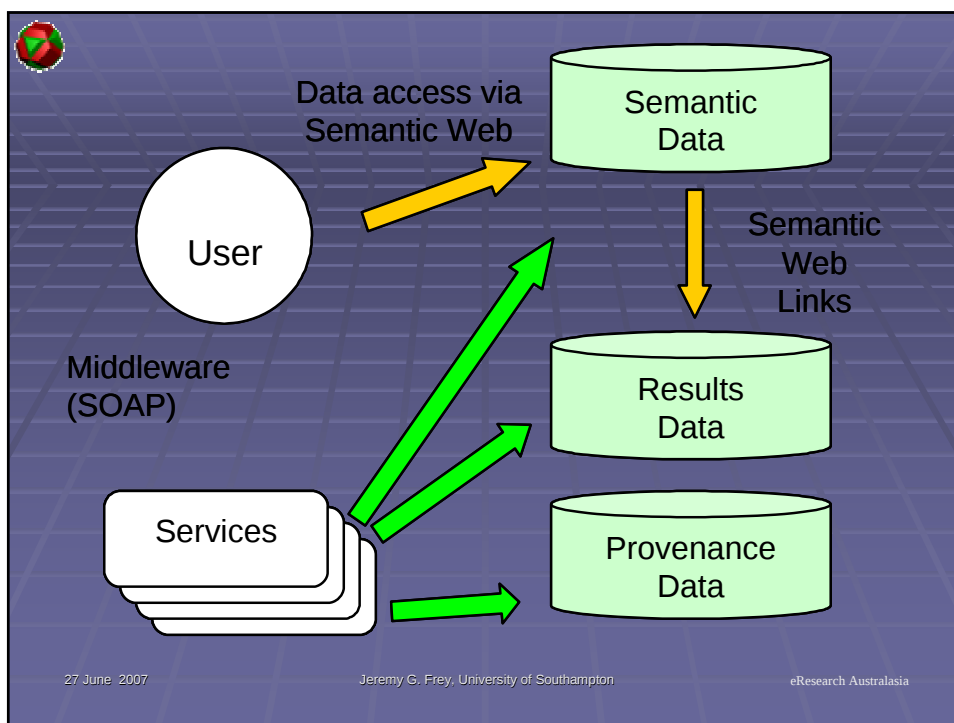


He is charged with expressing contempt for meta-data

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia





Properties in RDF

```
<c:OrganicMolecule
  rdf:about="file:///storage/ba8efc2ce0edada69d63b02d1b86
  30c6.rdf">
<c:has-inchi>1.12Beta/C12H13NO2/c1-2-15-8-9-5-6-11(14)12-
  10(9)4-3-7-13-12/h1H3,2H2,3-7H,8H2,14H</c:has-inchi>
<c:has-cas>22049-19-0</c:has-cas>
<c:has-empirical-formula>C12H13NO2</c:has-empirical-
  formula>
<c:has-stereocentres>0</c:has-stereocentres>
<c:has-property>
<c:MeltingPoint>
<c:has-information>
<c:Information>
<c:has-value>150</c:has-value>
<c:has-uncertainty>
<c:Range>
<c:has-value>16</c:has-value>
</c:Range>
</c:has-uncertainty>
</c:Information>
</c:has-information>
</c:MeltingPoint>
</c:has-property>
</c:OrganicMolecule>
```

Schema

```
<rdfs:Class rdf:about="&c;OrganicMolecule">
  <rdfs:label>Organic Molecule</rdfs:label>
  <rdfs:subClassOf rdf:resource="&c;Molecule" />
</rdfs:Class>

<rdfs:Class rdf:about="&c;PhysicalProperty">
  <rdfs:label>Property</rdfs:label>
</rdfs:Class>

<rdfs:Class rdf:about="&c;PartitionCoefficient">
  <rdfs:label>Partition Coefficient</rdfs:label>
  <rdfs:subClassOf
    rdf:resource="&c;PhysicalProperty" />
  <rdfs:description>Ratio of substance dissolved
    in octan-1-ol and water
  </rdfs:description>
</rdfs:Class>
```

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia



Quantities and Units using RDF

```
<ch:Quantity>
<ch:has-value>100</ch:has-value>
<unit:has-unit>
<unit:Unit>
  <rdf:type rdf:resource="unit:Joule"/>
  <unit:prefix rdf:resource="unit:Milli"/>
  <unit:power-of>1</unit:power-of>
</unit:Unit>
</unit:has-unit>
<unit:has-unit>
<unit:Unit>
  <rdf:type rdf:resource="unit:Kelvin"/>
  <unit:power-of>-1</unit:power-of>
</unit:Unit>
</unit:has-unit>
</ch:Quantity>
```

Try to avoid space craft crashing into Mars, or
planes running out of fuel

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia

TRIANGLE

* CAS * INCHI * NAME * SMILES * EMPIR

1202-32-0

Search the DB. Canned Example

query number = [10]

• MOLECULE

type	Resource
type	Molecule
has-SMILES	c1ccc(cc1C=CC(C)N(=O)=O)-O
has-name	B-ETHYL-B-NITROSTYRENE
has-mw	177.2
has-steric-centres	0
has-empirical-formula	C10H11NO2
has-cas	1202-32-0

48 29587

- type: OrganicMolecule
- has-property: 48 29587
 - type: PhysicalProperty
 - type: BoilingPoint
 - type: Resource
- has-information: 48 29587
 - type: Resource
 - type: Information
 - has-date: 2005-2-28
 - has-value: 107.3
- has-uncertainty: 48 29587
 - has-unit: 48 29587
 - type: Source
 - type: PhysProp
 - type: Resource
 - has-method: 54 29587
 - type: Resource
 - type: Laboratory
 - of-quality: 48 29587
 - has-condition: 16 29587
 - type: PhysicalProperty
 - type: Solubility
 - type: Resource

© University of Southampton 2005
ACCESS RESTRICTED
Version 1.0.4

Need instruments of the Grid- X Ray Crystallography The NCS

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia



The Smart Laboratory: Pub-Sub

- Pub Sub & Data Brokers
 - Important way to include laboratory environmental & people data in a flexible manner
 - Backed by archive database service
 - Correlate with experiments via time & place
 - Facilitate automation

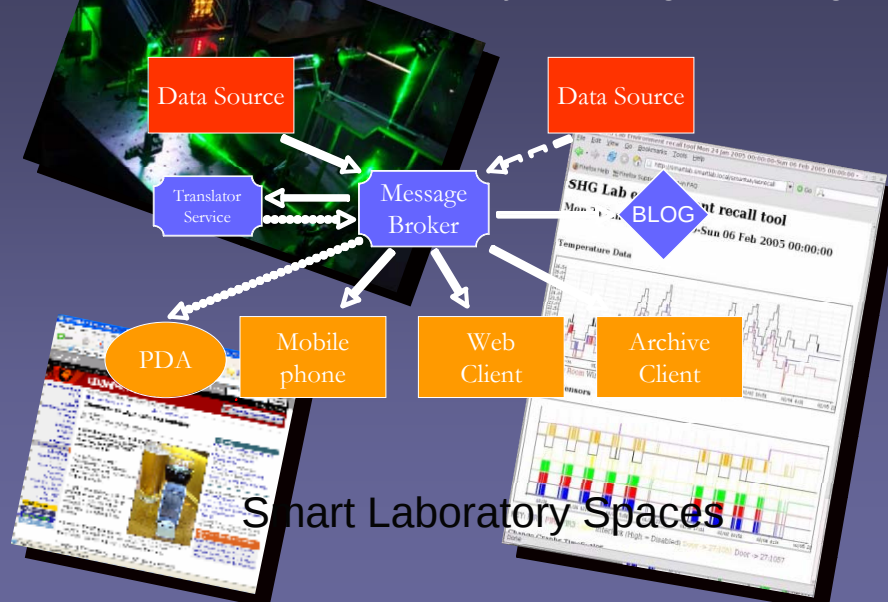


27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia

Pub-Sub systems provide the flexible & extensible approach to distribution of real time laboratory monitoring & archiving





27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia



Much more automation in modern chemistry





Experiments on the Grid



National Crystallography Grid Service

No, the computers are up, We're down

eResearch Australasia



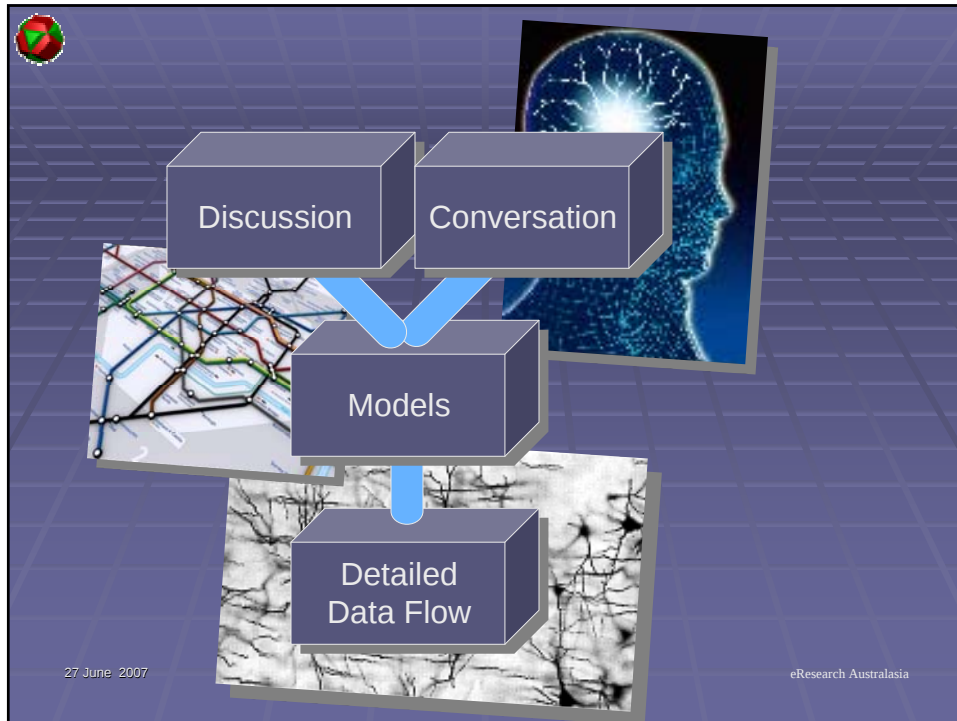
Plans

- Plans in advance are useful
- This is the way things are supposed to be done
- The Plan provides a digital context so increases the value of planning
- Key to our 'Smart Lab' approach....
- But is it the best way?

27 June 2007

Jeremy G. Frey, University of Southampton

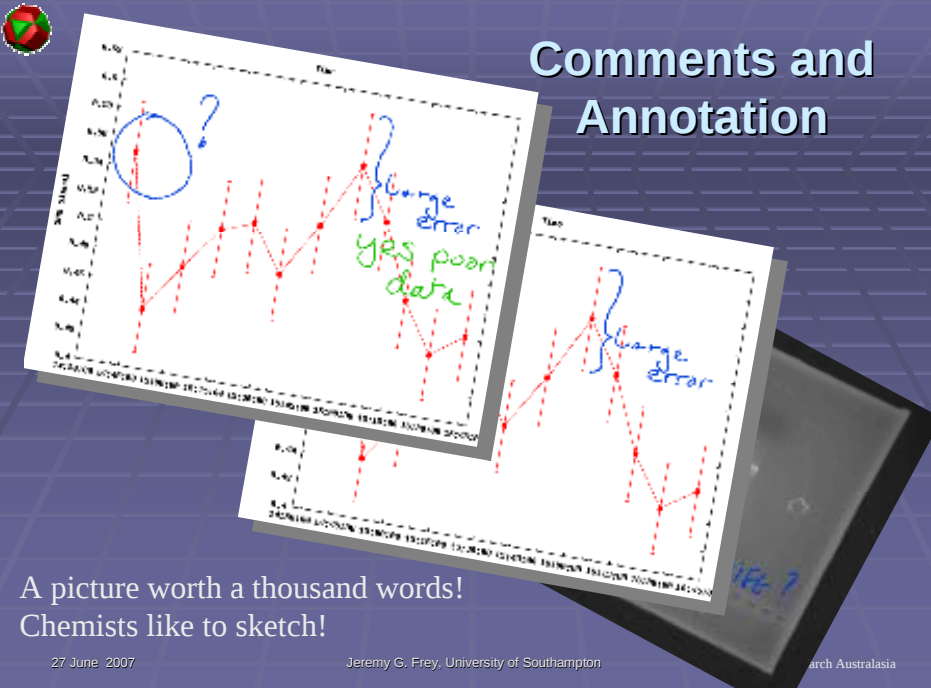
eResearch Australasia



Laboratory “Blogs”

- Laboratory notebook is a Blog
- Encourage and facilitate collaboration
- Need a data repository behind the Blog
 - R4L
 - E-Bank
- Flexible
 - Service oriented approach being developed
- A VRE

27 June 2007 Jeremy G. Frey, University of Southampton eResearch Australasia



Comments and Annotation

A picture worth a thousand words!
Chemists like to sketch!

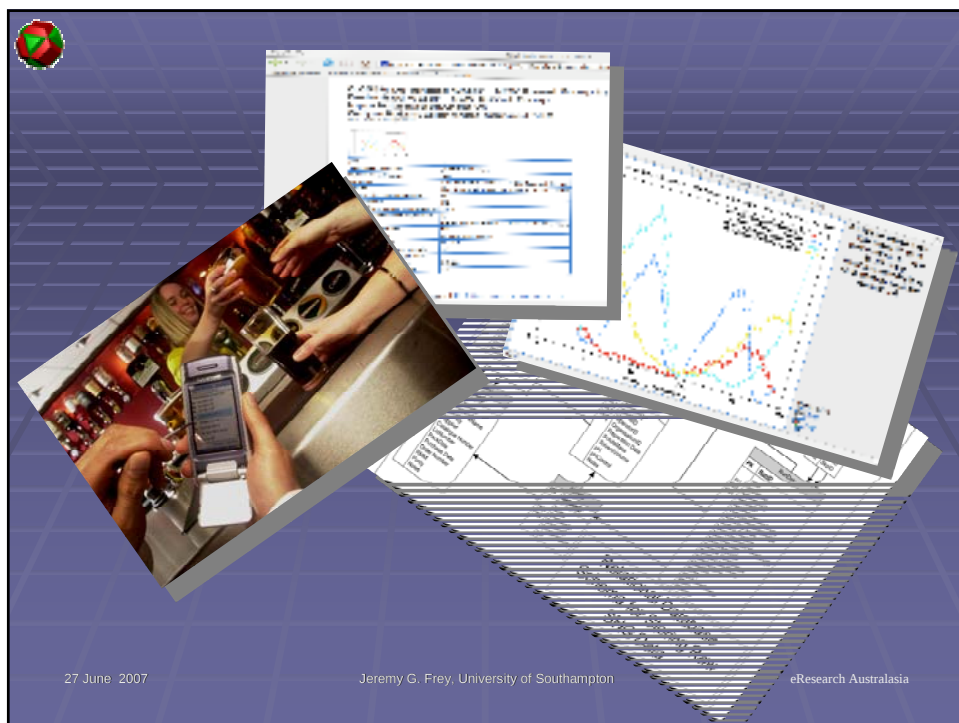
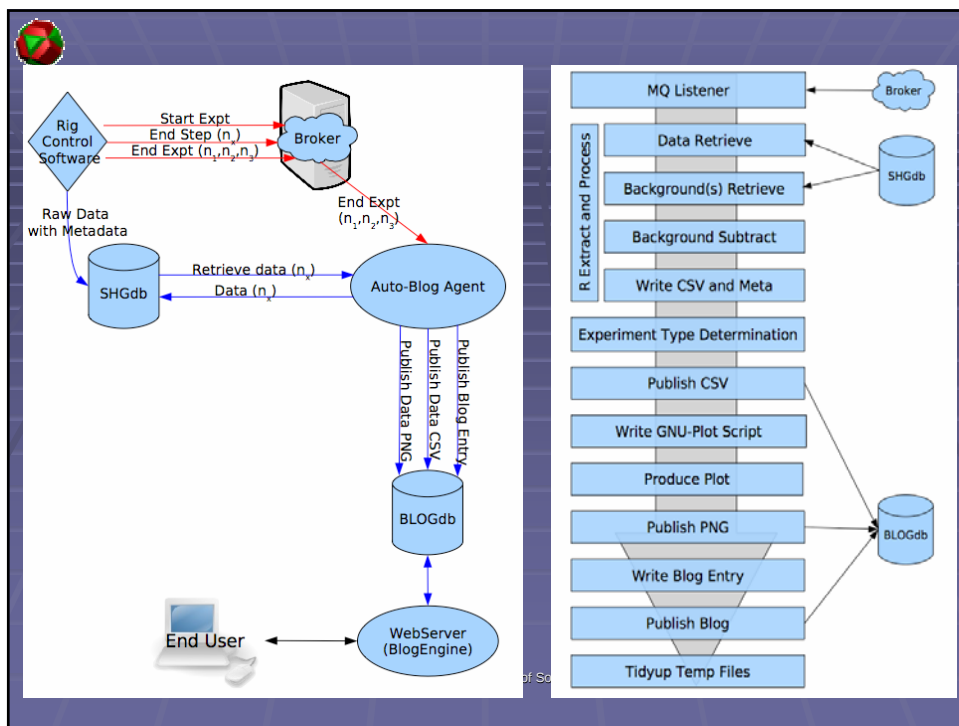
27 June 2007 Jeremy G. Frey, University of Southampton eResearch Australasia



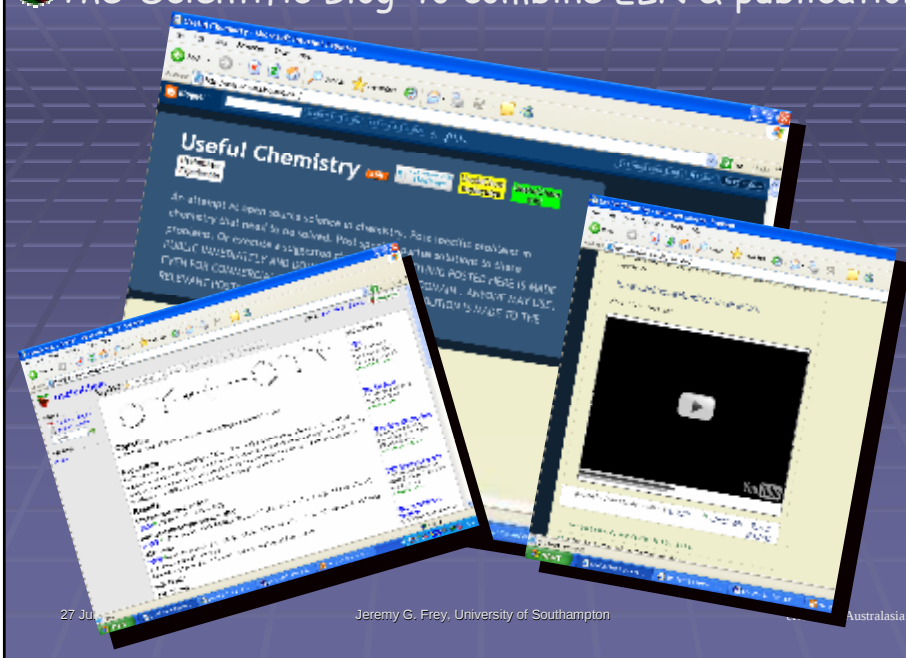
Blog-jects

- Equipment become first class members of the web
- Interacts well with Pub-Sub as items are attached to topics, topics relate the Bog items
- With automation this evolves to a two-way communication
- Live Copy essential

27 June 2007 Jeremy G. Frey, University of Southampton eResearch Australasia



The 'Scientific Blog' to combine ELN & publication



CoAKTing

Record the 'Scientific Conversation' -
this part of the record often exists only
in the 'grey literature'



People, Equipment &
Data are no different
online

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

“I can’t explain it - it’s just a
funny feeling that I’m being
Googled.”

But do you trust the information?

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia



Growing ne
equivalent

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

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

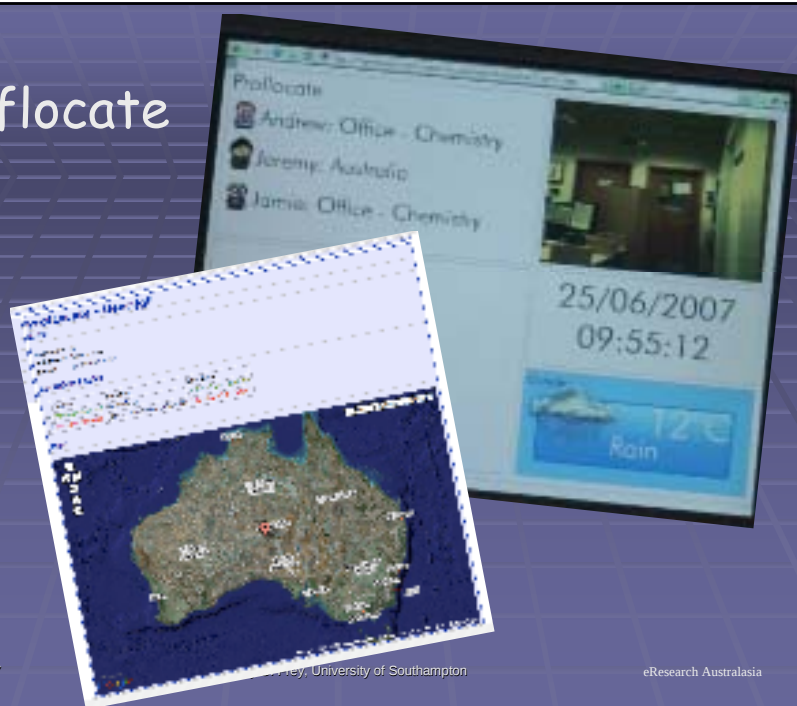
27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia



Proflocate



27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia



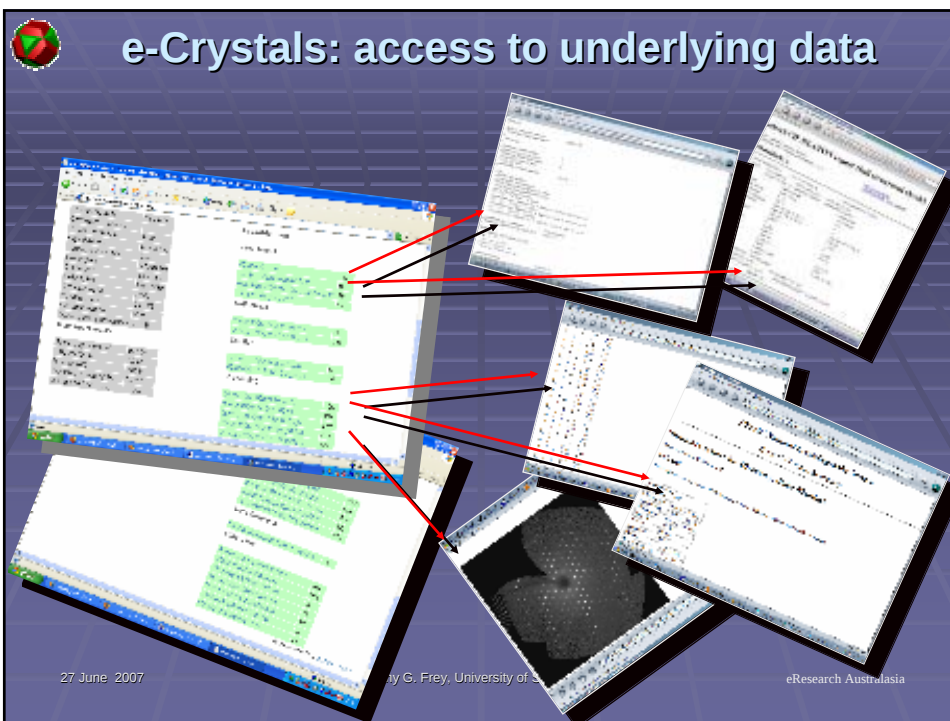
Validation

- Increasing the value of data
- How to bring all the necessary information together to enable appropriate validation
- Increasingly difficult & expensive to achieve
- Need provenance and context
- Essential step otherwise just a collection of items

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia



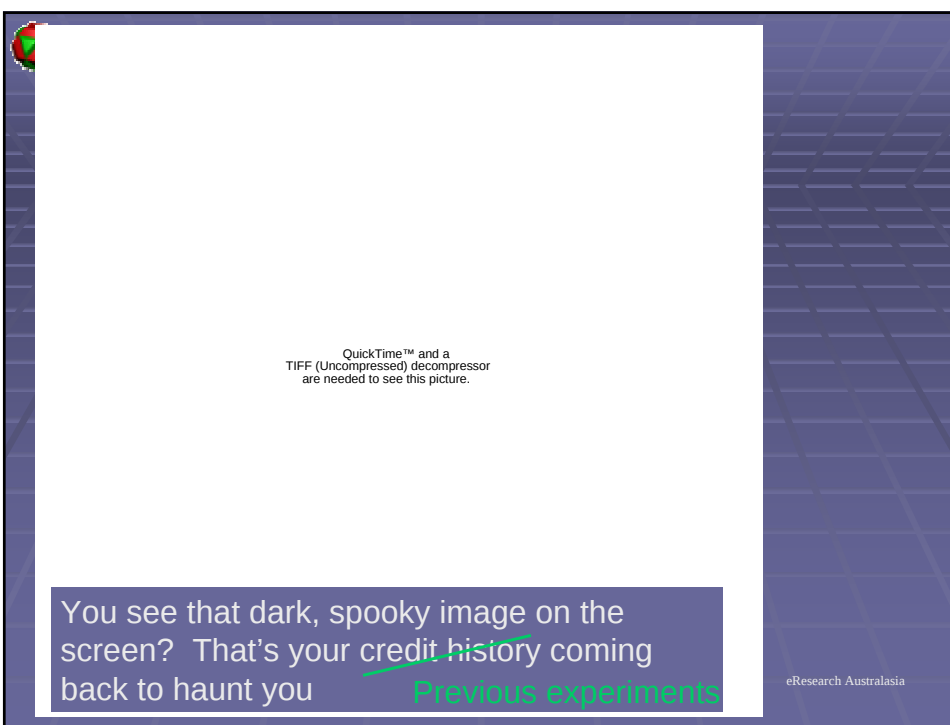
e-Crystals: access to underlying data

The image shows a collage of several computer windows. One window displays a list of data with green highlights. Another window shows a dark, abstract image with white speckles. Red arrows point from the data list window to the other windows, indicating a flow of information or data access.

27 June 2007

by G. Frey, University of

eResearch Australasia

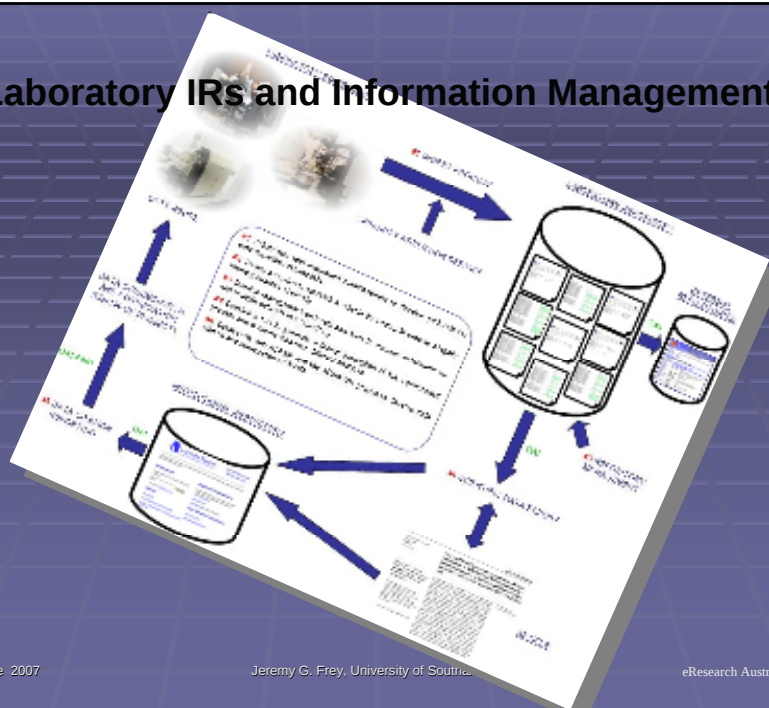


QuickTime™ and a TIFF (Uncompressed) decompressor are needed to see this picture.

You see that dark, spooky image on the screen? That's your credit history coming back to haunt you

[Previous experiments](#)

eResearch Australasia



Jeremy G. Frey, University of Southern



The eCrystals Public Data Archive

Crystal Structure Report Archive

6,7,8,10,12,13,15,16-Octahydro-benzo-1,4,7,10,13-pentaoxacyclopentadecin

Available Files

12,13,15,16-Octahydro-benzo-1,4,7,10,13-yelopentadecin

Crystal Structure Report Archive

12,13,15,16-Octahydro-benzo-1,4,7,10,13-yelopentadecin

Available Files

File

Date

Time

Author

Title

Abstract

Keywords

References

Citation

Download

View

27 June 2007

Southampton

eResearch Australasia

Southampton



Links

- The problem of providing and maintaining the links
 - Provenance & Navigation
 - Update & Notification
 - Update, re-calculate (choice, possibilities)
 - must work over web
 - Link management
 - Operations on the links - search
- Live Copy

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia



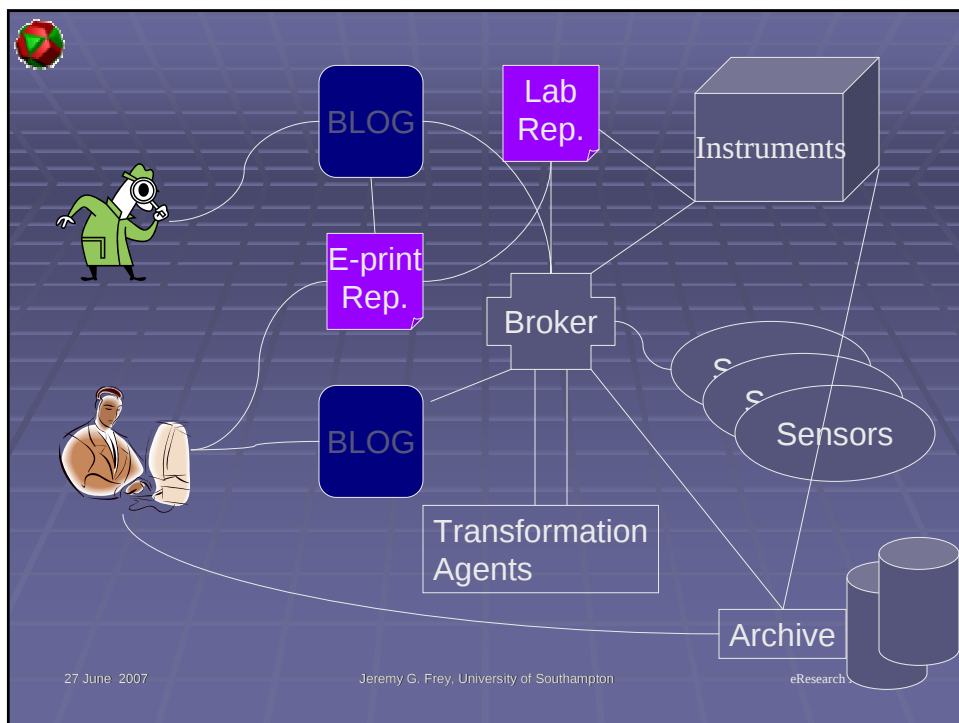
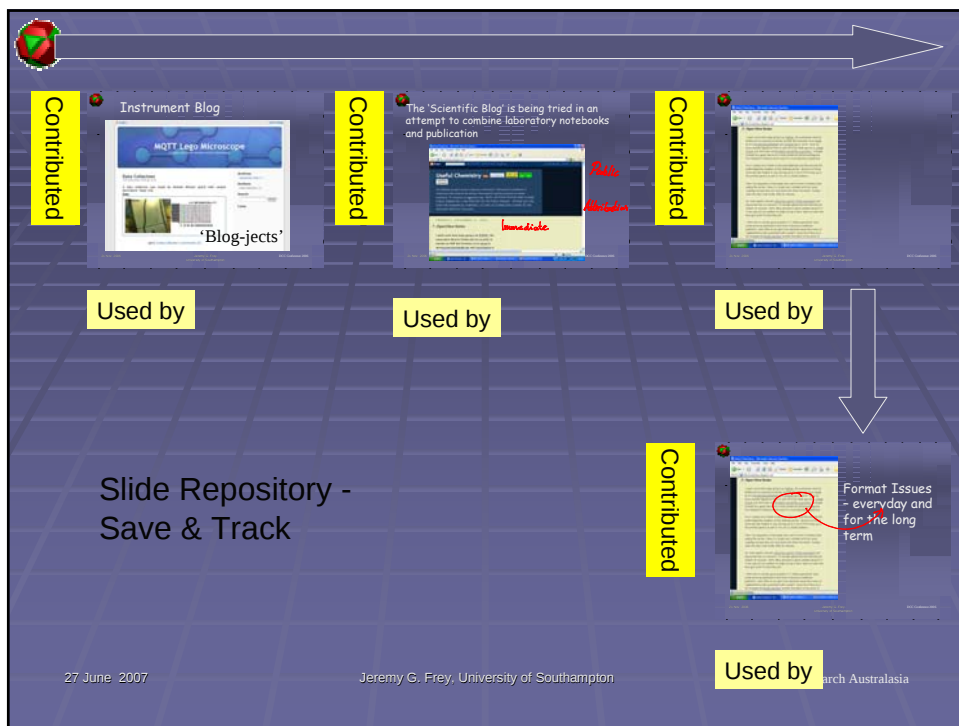
Need to know who is using your data

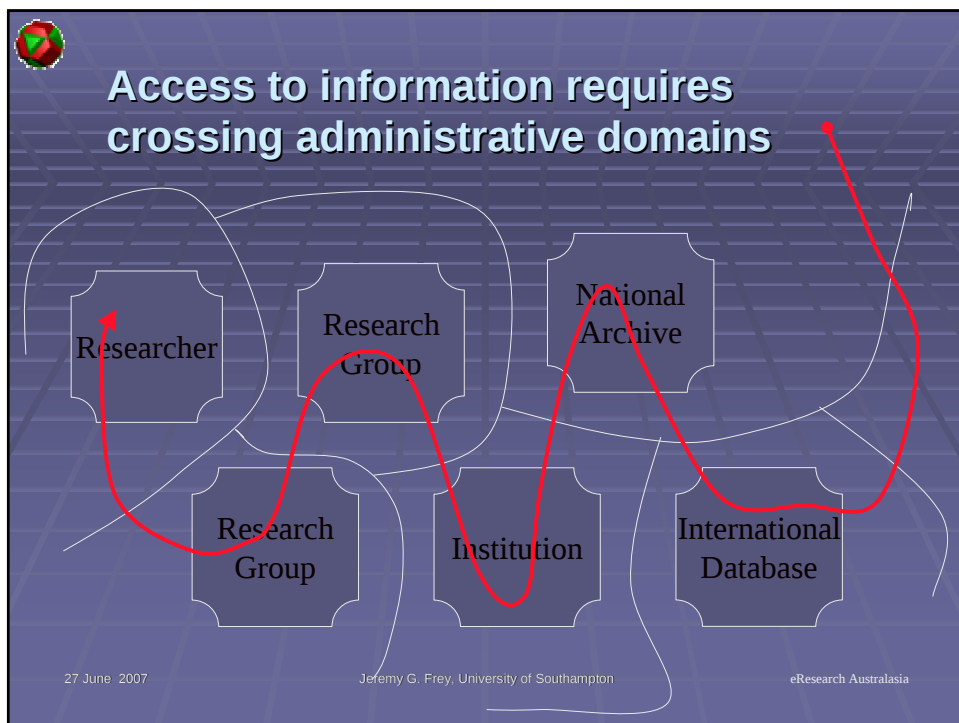
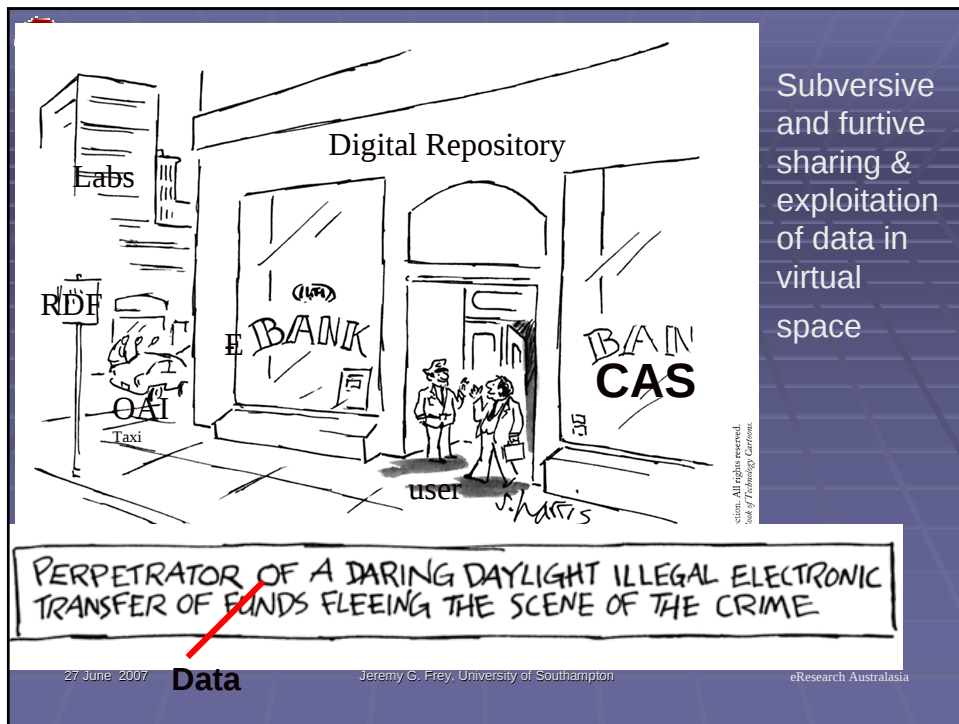
- Citations of all types of research output
- Track PowerPoint Slides
 - Special type of CC License that Contains a 'call back'
 - Report use of ppt slide
- Repository of ppt slides
 - Send back updated slides
 - Track evolution of slides

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia

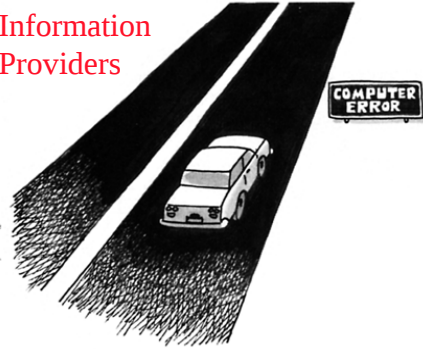




These are the same people – if we can ‘talk’ to ourselves efficiently over time then that is a good start to be able to ‘talk’ to others

Information
Providers

© 1999 New York University. All rights reserved.
From The New Yorker, June 17, 1999



Information
Consumers

ZIEGLER

27 June 2007

Jeremy G. Frey
University of Southampton

eResearch Australasia



QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

Surfing the
data?

Following
endless
trails?

Forgotten
why you
started?

*Are we thinking here, or is this just so much
pointing and clicking*

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia



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*

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia



People & Funding

- Southampton
ECS, MATHS & CHEMISTRY,
IT-INNOVATION
 - BRISTOL
 - UKOLN
 - CCLRC
 - INDIANA
 - SYDNEY
 - MANCHESTER
 - EPRSC e-Science &
Chemistry
Programmes
 - JISC, DTI
- **See web site for full details and links**
- **www.combechem.org**

27 June 2007

Jeremy G. Frey, University of Southampton

eResearch Australasia

