



# “The CombeChem Experience”

KM Workshop  
AHM 2005, Nottingham  
Jeremy Frey [j.g.frey@soton.ac.uk](mailto:j.g.frey@soton.ac.uk)

# Chemical Data

- Chemical data, information & knowledge
- Exponential growth in generation of data
- Need for automatic capture of meta data



# Chemical Grids

- Start in the laboratory – pervasive physical grid
- Computational chemistry very significant source – *in silico* grid
- Used by chemists so must be simple to support & maintain - autonomic

# Chemical Semantic Grid

- RDF
- Large scale triple stores (so far up to 50 Million triples of molecular structures and properties)

# ChemLab

The Chemistry 3/5 & 6  
Laboratories

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

DARTMOUTH COLLEGE

| 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*

### Safety

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

### Resources

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

[ChemLab Home](#)

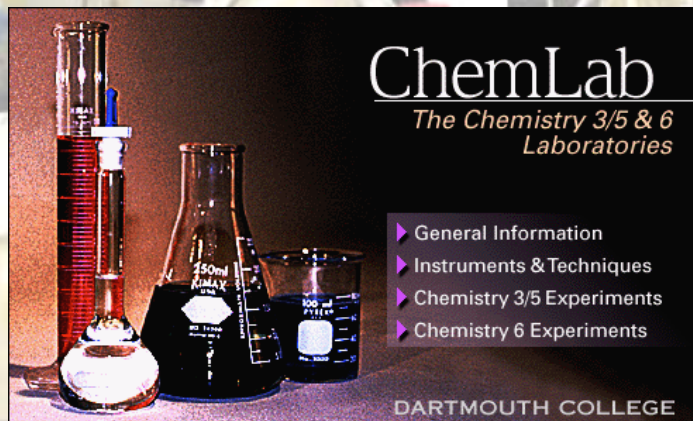
necessary if a calculation or discussion is changed; the section to be deleted is simply removed by drawing a neat "x" through it.

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



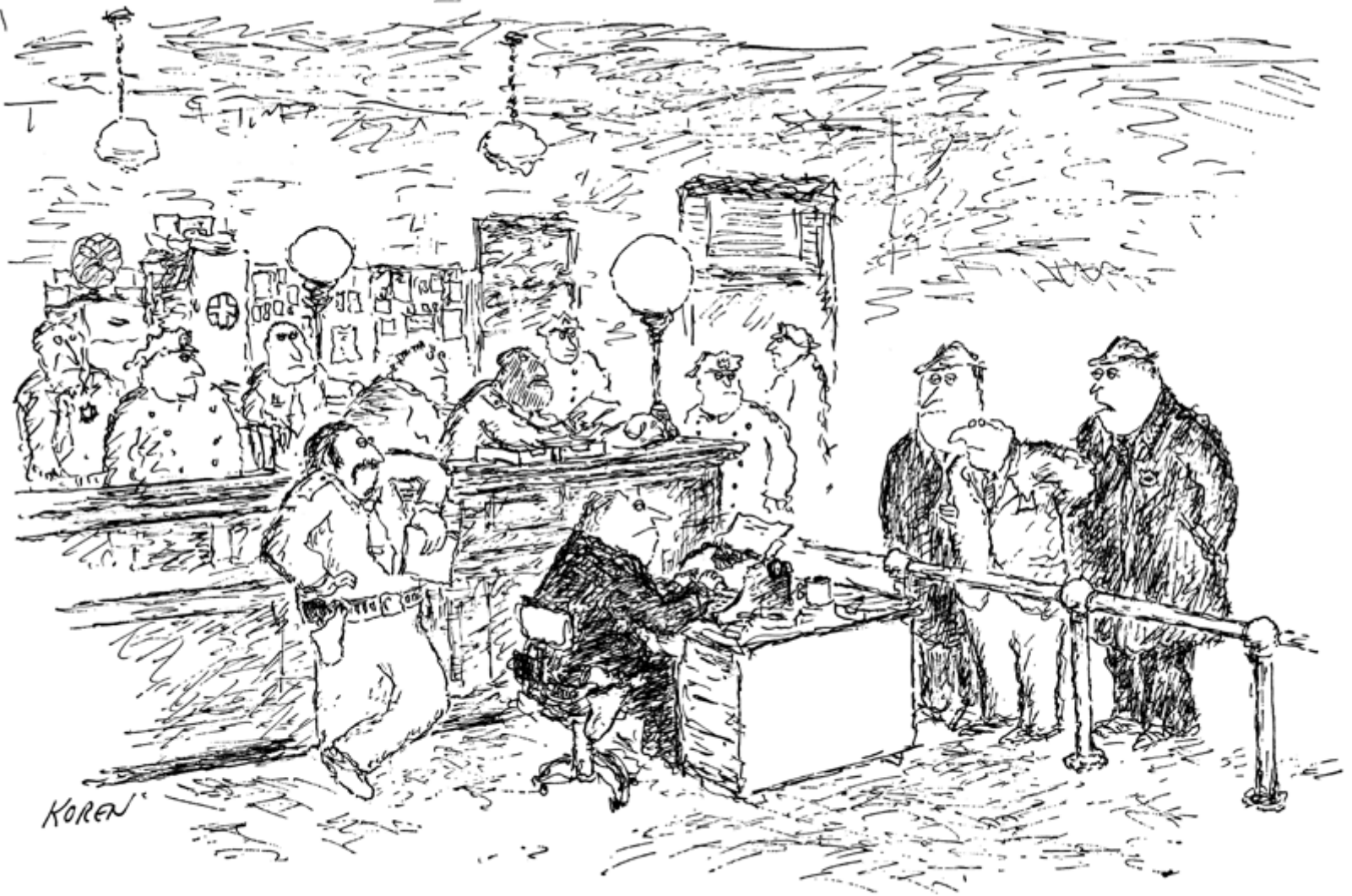


New Yorker collection. All rights reserved.  
from the New Yorker Book of Technology Cartoons.

Digital record at source don't try to add metadata after the fact

*“We can pause, Stu—we can even try fast-forwarding—but we can never rewind.”*

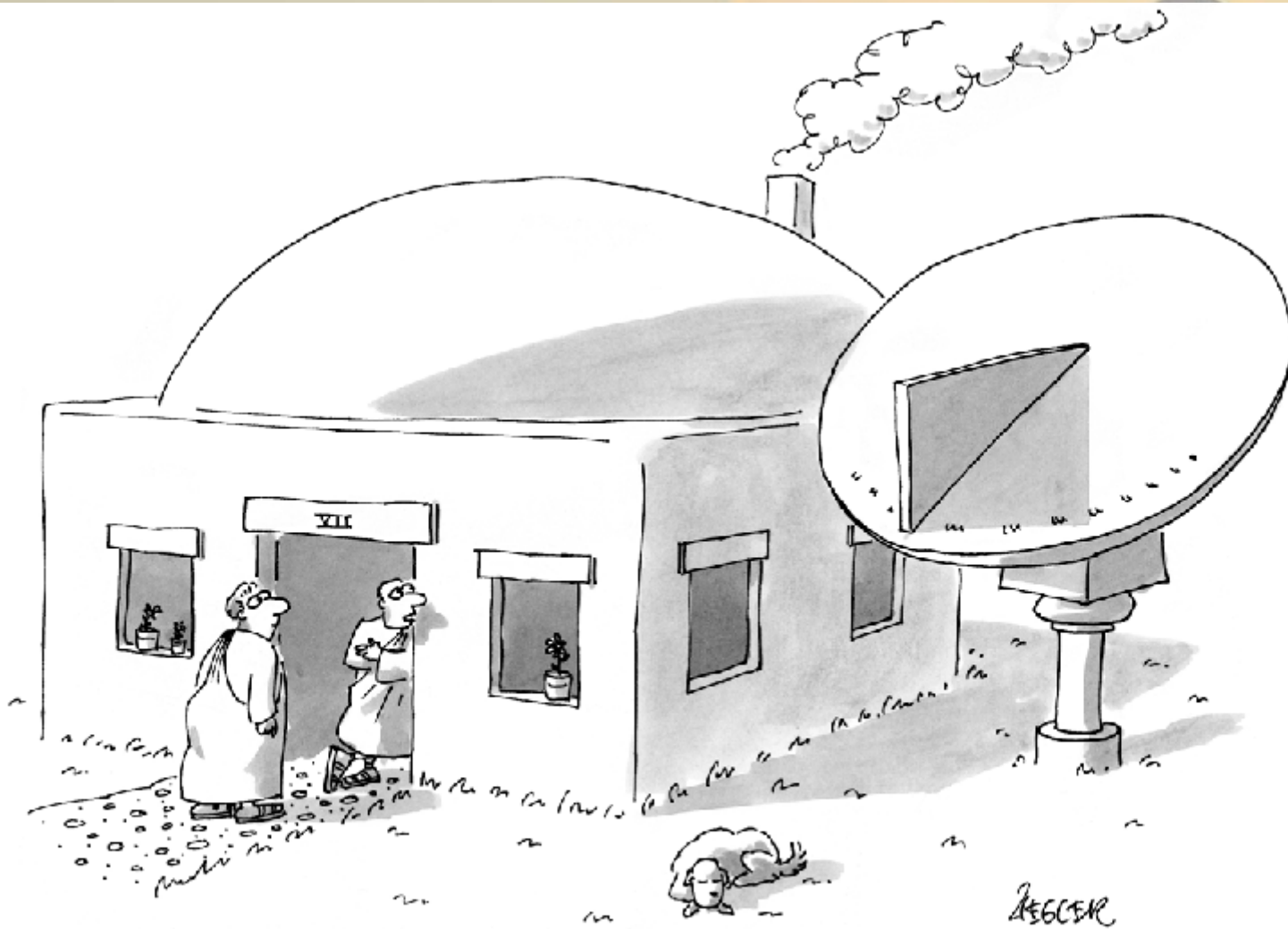




*"He's charged with expressing contempt for data-processing."*

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

He is charged with expressing contempt for meta-data



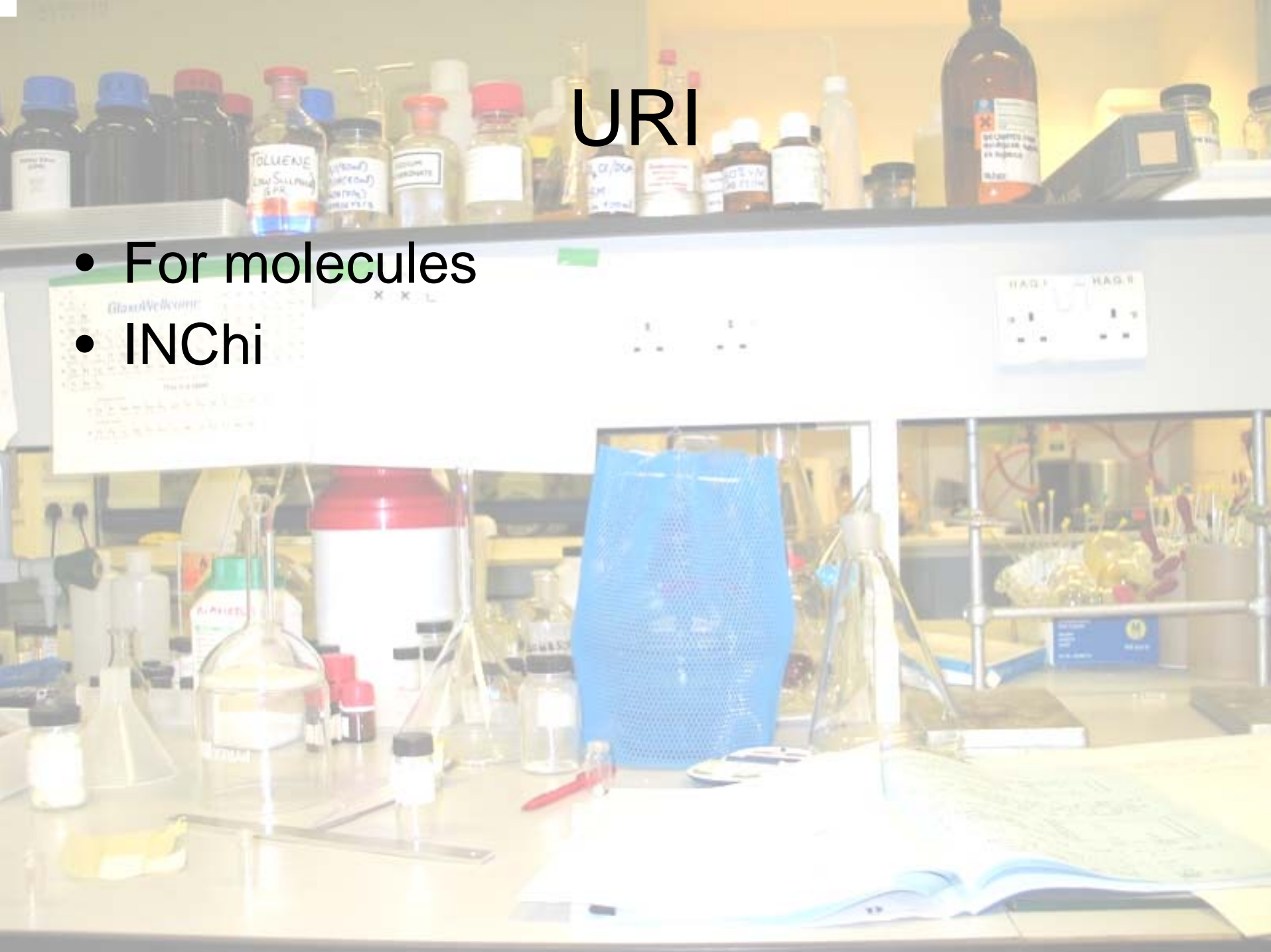
© The New Yorker collection. All rights reserved.  
From *The New Yorker Book of Technology Cartoons*.

*"Sure, it's an eyesore, but we get better time than anyone else in the neighborhood."*

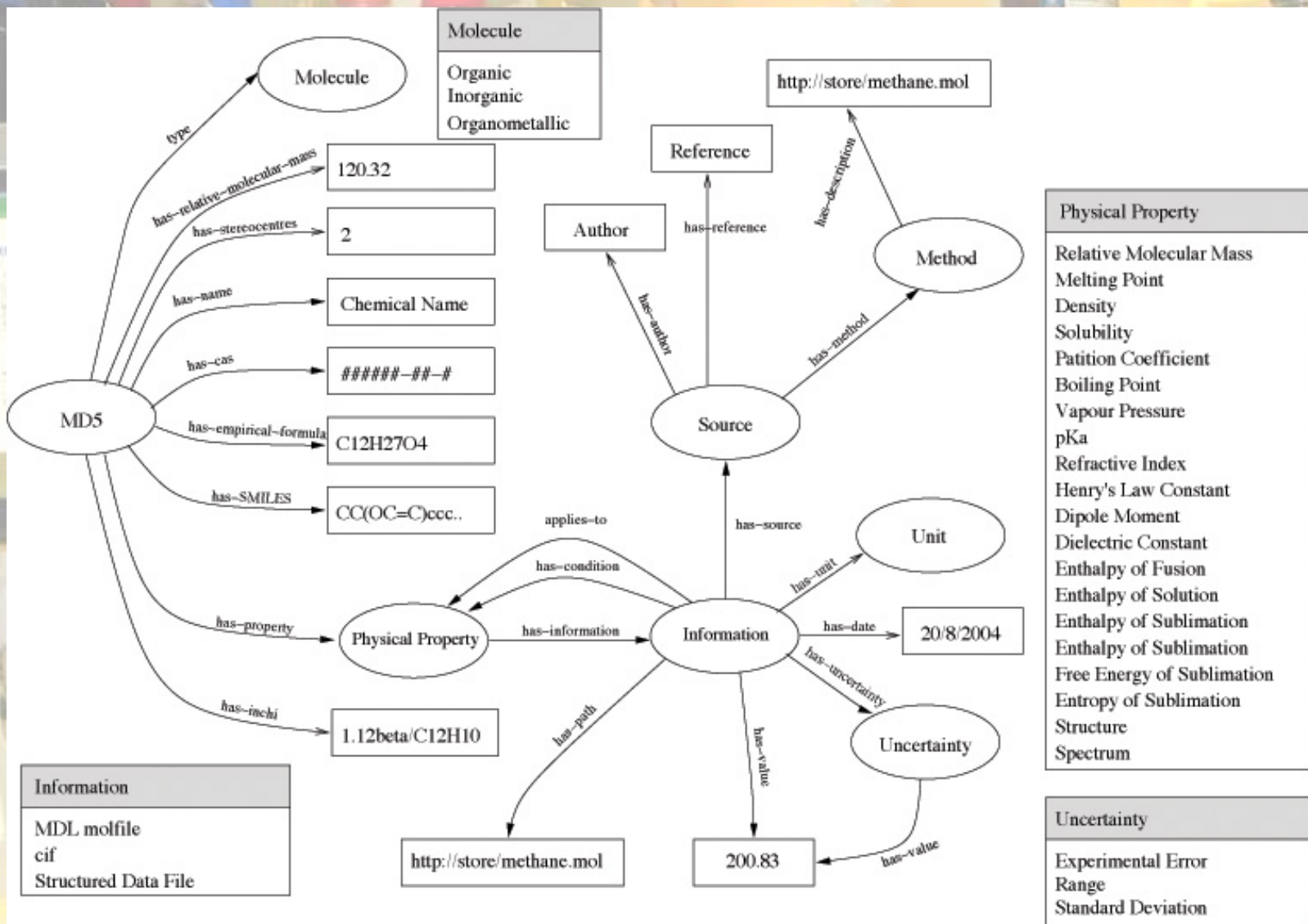
**Old technology does not scale**

# URI

- For molecules
- INChi



# Create large semantically rich database of structures and properties



- <c:OrganicMolecule  
rdf:about="file:///storage/ba8efc2ce0edada69d63b02d1b8630c6.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>

## Property in RDF

*Currently testing on  
200,000 compounds but  
about to go up by order of  
magnitude*

*3Store is a scaleable  
solution*



Provenance  
Record experiments  
Make data available  
(e-crystals, e-Bank)

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

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

You see that dark spooky image on the screen?

That's your credit history coming back to haunt you?

Security  
and trust  
for  
experiments  
and data



*"On the Internet, nobody knows you're a dog."*

# Chemistry Data

private or public,

open or controlled access



Subversive  
and furtive  
exploitation  
of data



tion. All rights reserved.  
ok of Technology Cartoons.

PERPETRATOR OF A DARING DAYLIGHT ILLEGAL ELECTRONIC  
TRANSFER OF FUNDS FLEEING THE SCENE OF THE CRIME

**Data**

# Importance of using & maintaining distributed data



Yorker collection. All rights reserved.  
re: Yorker Book of Technology Cartoons.



*"Someday, son, all this will belong to Bill Gates."*

# Chemistry and Life-Sciences

A photograph of a laboratory bench. In the foreground, there is a large Erlenmeyer flask containing a yellow liquid, a smaller flask with a white substance, and a red pipette. A blue mesh safety shield is positioned in the center. In the background, a shelf holds numerous bottles of chemicals, including one labeled 'TOLUENE'. An electrical outlet panel is visible on the wall behind the bench. An open book or document lies on the bench in the lower right.

- Need to link up
- small and large molecule chemistry
- Bio-Informatics
- Environmental Informatics



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

Making  
sure  
Chemistry  
will not  
suffer  
from a  
data  
crunch

*"All I'm saying is now is the time to develop the technology to deflect an asteroid."*

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