Effective chemical information

Jonathan M Goodman
jmg11@cam.ac.uk

Department of Chemistry
Cambridge University
Effective Chemical Information

Can we have information sources:
- which do not need updating
- which can validate themselves
- which use almost no data
- which are rarely accessible
- which are wrong

... and still be useful?
Cambridge Chemistry Webserver

http://www.ch.cam.ac.uk/

Cambridge University
Chemical Laboratory

Search Cambridge Chemistry

Cambridge e-mail addresses

Chemistry library

- Address | Academic Staff | Introduction | Research Areas | Chem@Cam | Chem@300
- Colloquia: Inorganic | Organic | Physical | Theory | Informatics | Melville | General
- Chemistry Department: WWW servers (Graphical Index) | CPGS | Computer Officers
- Analytical | Atmospheric | CCDC | CCMR | Melville | Physical | Theoretical | Informatics

Chemistry 2000: (People | UK | Europe | USA | World | Societies | Data)
- Chemistry Journals | Periodic Table | Silicon Graphics Laboratory
- Undergraduate (more info) | Teaching Resources | Chemical Calculations
- Postgraduate Applications | Jobs and Studentships in the Department of Chemistry

Information for Cambridge staff and students (restricted access)
- Administration team (forms) | Department: phone book | Stores
- First Aid Information | Technology Transfer Office
- ISI Web of Science | Corporate Associates Scheme

- New Information (23rd July 2002):
  - Chemistry Stores Information
  - Staff Review Forms - on-line (Cambridge Access only)
  - NEWTON - the new online library catalogue
  - PhD: Chemical Biology / Biochemistry
  - PhD: Computational Organic Chemistry

Today: Tuesday 6 August 2002

No events in database

Today in Chemical History: Sir Alexander Fleming

This web server designed and maintained by Dr. I M Goodman
Chemical Informatics Letters

http://www.ch.cam.ac.uk/MMRG/CIL/

• The latest news in chemical informatics, links to relevant web sites, analysis and discussion

• Current Issue:
  – Volume 5, Issue 2; August 2002
Chemistry 2000
http://www.ch.cam.ac.uk/c2k/

Review:
Chemistry 2000

http://www.ch.cam.ac.uk/c2k/

- Is it sustainable?
- Is it reasonable use of information?
- Java/XML programs check and update the list once a month
CAVEAT ARANEA

http://www.ch.cam.ac.uk/c2k/resource.html

- Trust no one
- Who provides the information?
  - Is it too ambitious?
  - Is it within authors’ expertise?
  - Is it within the authors’ resources?
- Is it clear what the site provides?
- What is the author selling?
- Try several resources
Chemistry 2000

http://www.ch.cam.ac.uk/c2k/

• Is it sustainable?
• Is it useful?
• Is it comprehensive?
Chemistry 2000
http://www.ch.cam.ac.uk/c2k/

• Is it sustainable? √ YES!
• Is it useful? √ YES!
• Is it comprehensive? √ YES!
Chemistry 2000

http://www.ch.cam.ac.uk/c2k/

Links for Chemists:
http://www.liv.ac.uk/Chemistry/Links/links.html
8500 links (August 2001); 8075 links (2002)

ChemDex:
http://www.chemdex.org/
5806 links (August 2001); 7192 links (2002); 1 % failure rate

Chemistry 2000:
http://www.ch.cam.ac.uk/c2k/
2783 links (2002); 47 problems (1.7 %)
Chemistry 2000
http://www.ch.cam.ac.uk/c2k/

Extend list of chemistry departments to a list of chemists?
<table>
<thead>
<tr>
<th>Location</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>Aberdeen</td>
</tr>
<tr>
<td>Europe</td>
<td>Bangor</td>
</tr>
<tr>
<td>USA</td>
<td>Bath</td>
</tr>
<tr>
<td>Rest of World</td>
<td>Belfast (Queen's)</td>
</tr>
<tr>
<td></td>
<td>Birmingham</td>
</tr>
<tr>
<td></td>
<td>Bradford - Polymer IRC</td>
</tr>
<tr>
<td></td>
<td>Brighton</td>
</tr>
<tr>
<td></td>
<td>Bristol</td>
</tr>
</tbody>
</table>

Dr J M Goodman: Cambridge (UK)
Joshua L. Goodman: New York - Rochester (USA)
Jonathan Goodman: Tennessee - Vanderbilt University (USA)
MURRAY GOODMAN: California - University of California - San Diego (USA)
Ms. Pat Goodman: Alabama - Samford (USA)

http://www.vanderbilt.edu/AnS/Chemistry/chemmain/Goodman.html
<table>
<thead>
<tr>
<th>Location</th>
<th>Institution Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>Wisconsin - University of Wisconsin - Carroll College</td>
</tr>
<tr>
<td>Europe</td>
<td>Wisconsin - University of Wisconsin - Eau Claire</td>
</tr>
<tr>
<td>USA</td>
<td>Wisconsin - University of Wisconsin - La Crosse</td>
</tr>
<tr>
<td>Rest of World</td>
<td>Wisconsin - University of Wisconsin - Madison</td>
</tr>
<tr>
<td></td>
<td>Wisconsin - University of Wisconsin - Oshkosh</td>
</tr>
<tr>
<td></td>
<td>Wisconsin - University of Wisconsin - Parkside</td>
</tr>
<tr>
<td></td>
<td>Wisconsin - University of Wisconsin - Stevens Point</td>
</tr>
<tr>
<td></td>
<td>Wisconsin - Viterbo College</td>
</tr>
</tbody>
</table>

Barnhard, Ralph J.: Oregon - Oregon
Dennis W. Bamum: Oregon - Portland State
Barofsky, Douglas: Oregon - Oregon State
Susan Barrett: Pennsylvania - Lehigh
Barry, Nancy: Iowa - Luther College
Ludwig Bartels: California - University of California - Riverside
Bartfai, Tamas: California - Scripps Research Institute
Roger Barth: Pennsylvania - West Chester
Professor Paul A. Bartlett: California - University of California - Berkeley
Professor Emeritus Neil Bartlett: California - University of California - Berkeley
Bartlett, Anne: Iowa - Luther College
Mr. Kevin Barton: Nebraska - University of Nebraska - Omaha
Janice Barton: Kansas - Washburn

Jacqueline K. Barton: California - California Institute of Technology
Richard A. Bartsch: Texas - Texas Tech University

http://www.its.caltech.edu/~jkbgp/

**Magnus chemistry**

**c2k: 11094 chemists**

Stephen Allen and

Jonathan M Goodman

Cambridge University

(c) 2001-2002
List of Chemists

http://www.ch.cam.ac.uk/c2k/

• Is it comprehensive? NO (50%?)
• Is it sustainable? √ YES!
• Is it useful? √ YES!
Access to index per day

- Red - new index
- Green - old index
- Black - total
Access to index per day

- Red - new index
- Green - old index
- Black - total
Access to journal index per day

- Red - new index
- Green - old index
- Black - total
Tools for Chemists

http://www.ch.cam.ac.uk/magnus/

J. M. Goodman, P. D. Kirby, and L. O. Haustedt

J. M. Goodman, A.-K. Köhler and S. C. M. Alderton
Tools for Chemists

http://www.ch.cam.ac.uk/magnus/
Tools for Chemists

http://www.ch.cam.ac.uk/magnus/

The variation of boiling point with pressure

Boiling points at 1 atm:
SolventA: Benzene 80°C
SolventB: Acetone 56°C
Solute: 200

Quantities (moles):
SolventA: 10
SolventB: 10
Solute: 2

Temperature: 20°C

Final Evaporation Extent: 100

New Molar Quantities after Evaporation:
SolventA: 0.000
SolventB: 0.000
Solute: 0.000
August 13th

Solomon Zbewski

Born on this day, 1879, Ob. 1951. Nobel Prize in Physiology or Medicine, 1934 for discoveries concerning the replication mechanism and the genetic structure of viruses.

August 14th

Richard R. Ernst

Born on this day, 1933. Nobel Prize in Chemistry, 1991 *for his contributions to the development of the methodology of high resolution nuclear magnetic resonance (NMR) spectroscopy*.

August 15th

Prince Louis-Victor Pierre Raymond de Broglie

Born on this day, 1892, Ob. 1987. Nobel prize in Physics, 1929 *for his discovery of the wave nature of electrons*.

August 15th

Gerty Theresa Cori, née Radnitz

Born on this day, 1896, Ob. 1957. Nobel Prize in Physiology or Medicine, 1947 *for the discovery of the course of the catalytic conversion of glycogen*, a prize shared with her husband Carl Ferdinand Cori.

August 16th

Wendell Meredith Stanley

Born on this day, 1904. Nobel Prize in Chemistry, 1946 *for his preparation of enzymes and virus proteins in a pure form*.

August 19th

Julius Lothar Meyer

Born on this day, 1830, Ob. 1895. Constructed a periodic table of the elements but did not predict the existence of undiscovered elements, unlike his near-contemporary Mendeleev.

August 20th

Hidetosi Shirakawa

Born on this day, 1936. Nobel Prize in Chemistry, 2000 *for the discovery and development of conductive polymers*.

August 23rd

Robert F Curl

Born on this day, 1933. Nobel Prize in Chemistry, 1996 *for their discovery of fullerenes*.

August 23rd

Hamilton O. Smith

Born on this day, 1931. Nobel Prize in Physiology or Medicine, 1978 *for the discovery of restriction enzymes and their application to problems of molecular genetics*.

August 25th

Hans Krebs

Born on this day, 1900, Ob. 1981. Nobel Prize in Physiology or Medicine, 1953 *for his discovery of the citric acid cycle*. After studying medicine and chemistry at Berlin, he worked as a doctor, before leaving Germany in 1933. He moved to the department of biochemistry in Cambridge, then to Sheffield, and finally to Oxford in 1954.

August 26th

Antoine-Laurent Lavoisier

Born on this day, 1743; published his textbook *Elementary Treatise on Chemistry* 1789; guillotined for being a tax collector, May 8th, 1794. Panopticon Lavoisier contains more information about him.
Educational Resources
Synthesis, reactivity and conformational preferences of novel enediynyl peptides: a possible scaffold for β-sheet capping turns

Amit Basak, Kakali Rani Rudra, Subhendu Sekhar Bag and Ajoy Basak


Experimental Analysis

Paste a paragraph of experimental data:

N-tert-Butyloxy carbonyl-L-alaninylaminoundec-5-ene-3,7-diynoyl-L-serine; Yield: 65%; State: viscous oil;

a. R¹ = R² = Me
b. R¹ = R² = CH₂Ph
c. R¹ = Me, R² = CH₂OH

Mass (EI) m/z 615 (M⁺); HRMS calculated for C₃₅H₄₁N₃O₇; 615.2946 found 615.2950.

web-based data-checking
Synthesis, reactivity and conformational preferences of novel enediynyl peptides: a possible scaffold for β-sheet capping turns

Amit Basak, Kakali Rani Rudra, Subhendu Sekhar Bag and Ajoy Basak


Experimental Analysis

Paste a paragraph of experimental data:

(4 H, br s, piperidine CH 2), 1.09 (2 H, br s, piperidine CH 2), 0.95 (4 H, br s, piperidine CH 2), δ (62.90 MHz, CDCl 3) 14 (C), 139.23 (C), 138.00 (C), 133.79 (CH), 131.08 (CH), 130 (CH), 128.19 (CH), 123.18 (C), 123.04 (CH), 54.48 (CH 2), 25.69 (CH 2), 24.00 (CH 2). m/z (APCI) 344.2 (MH, 100%).
data - database - information

World of data

Ordered data

information
data - database - information

World of data

Ordered data

XML

information
How well are we using XML in Chemistry? J. M. Goodman
Effective Chemical Information

Can we have information sources:
- which do not need updating - Journals (CIL)
- which can validate themselves - Chemistry 2000
- which use almost no data - Boiling points, etc
- which are rarely accessible - History of chemistry
- which are wrong - Text books, journals,

... and still be useful?
Acknowledgements

• The Royal Society
• EPSRC
• BBSRC
• EU
• Unilever
Jonathan M Goodman

MA PhD CChem FRSC

• University Lecturer
• Fellow of Clare College
  – Admissions Tutor (to September 2001)
  – Director of Studies in Physical Natural Sciences (from October 2001)

• Other Appointments:
  – Editorial Board of Journal of Molecular Graphics and Modelling
  – ChemWeb Preprint server Scientific Advisory Board
  – PSIgate (http://www.psigate.ac.uk/) Advisory Board
  – IUPAC Division VIII (Chemical Nomenclature and Structure Representation) Advisory Subcommittee
Current Undergraduate Teaching

• Undergraduate Lectures:
  – IB: Shape and Organic Reactivity
  – II: Molecular Modelling and Informatics
  – III: New Synthetic Methods

• Demonstrating:
  – Synthetic Organic Chemistry
  – Molecular Modelling

• Examining:
  – IA examiner

• Supervising:
  – All years of the course in organic chemistry and molecular modelling
Teaching: Unilever Courses

- Molecular Data on the Internet
- Data Analysis
- Java Programming
Conferences 2001

• Second Joint Sheffield Conference on Chemoinformatics.
  – University of Sheffield, April 9th - 11th 2001.

• RSC Annual Conference, 2001
  – Specialist Symposium: How new developments in computational chemistry and molecular modelling can help synthetic and combinatorial chemists
  – Birmingham, August 2001

• 222nd ACS National Meeting
  – Science Portals on the Internet
  – Chicago, USA, August 2001

• The 2001 International Chemical Information Conference
2001 Publications


2002 Publications

