

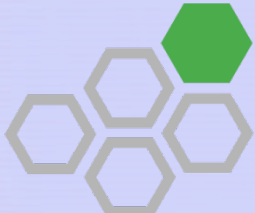
# Experience with data handling in large chemical databases

**NEAL LANGERMAN**

Advanced Chemical Safety  
San Diego, California, USA



# Large chemical information databases

**Hazmat**  *Navigator*

- Provide integrated chemical health and safety information with a strong emphasis on reactive chemistry hazards.



# Large chemical information databases



## Bretherick's Handbook of Reactive Chemical Hazards

Content: The resource for reactivity

# of records: 5764



## Sittig's Handbook of Toxic & Hazardous Chemicals and Carcinogens

Content: 2,100 most heavily used, transported & regulated chemicals

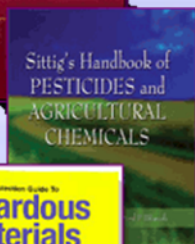
# of records: 1377



## Encyclopedia of Toxicology

Content: Toxicology

# of records: 551



## Sittig's Handbook of Pesticides & Agricultural Chemicals

Content: Chemicals used in food, cosmetics, and personal care products

# of records: 674



## NFPA's Guide to Hazardous Materials

Content: Hazardous, flammable chemicals

# of records: 320



# Large chemical information databases



Knovel. Know More. Search Less.

Substitute Chemical Database

- Help Chemical Process Engineers, Safety Officers and R&D Chemists identify and evaluate chemicals and their properties for performance, environmental, health and safety issues.





# Large chemical information databases



- Knovel Solvents - A Properties Database
  - 1800 records
  - 145 properties
- Existing manually constructed DB
- *Bretherick's Handbook of Reactive Chemical Hazards*
  - 344 records overlap
- *Sittig's Handbook of Toxic and Hazardous Chemicals and Carcinogens*
  - 279 records overlap



# Large chemical information databases

- Haz-Mat Navigator
  - Initial Merge
    - Use CAS # as key
    - Map fields from each source into DB
    - Add discrepancies rather than resolve
    - Report units rather than standardize
    - Text fields (Brethrick narratives) OK



# Large chemical information databases

- Haz-Mat Navigator
  - New or revised records
    - Forced to use data form from contractor

## **Record title**

### **Record Title**

FILL IN CHEMICAL NAME – IUPAC IF POSSIBLE OR COMMON NAME IF BEST RECOGNIZED

URL <http://ilab.acdlabs.com>

## **Properties**

### **Reaxys Registry Number**

H/N WILL FILL IN



# Large chemical information databases

- Final data entry form ran 75 pages

TLV-TWA (ACGIH): xx.xx mg/m<sup>3</sup> / ppm/NE

TLV-STEL (ACGIH): xx.xx mg/m<sup>3</sup> / ppm/NE

PEL- TWA (OSHA): xx.xx mg/m<sup>3</sup> / ppm/NE

PEL-STEL (OSHA): xx.xx mg/m<sup>3</sup> / ppm/NE

REL-TWA (NIOSH): xx.xx mg/m<sup>3</sup> / ppm/NE

REL-STEL (NIOSH): xx.xx mg/m<sup>3</sup> / ppm/NE

IDLH (NIOSH): xx.xx mg/m<sup>3</sup> / ppm/NE

MAK (DFG): xx.xx mg/m<sup>3</sup> / ppm/NE

Manufacturer's OEL

No data are available.

BIOLOGICAL EXPOSURES INDICES (BEIs; ACGIH) FOR THIS SECTION REFER TO (MOST CURRENT DATE) ACGIH "TLVs AND BEIs": There are no BEI's currently established for this chemical. The following BEI's are established for this product:





# Large chemical information databases

- References associated with major TABS

International Agency for Research on Cancer, <http://monographs.iarc.fr/> (accessed FILL IN DATE)

EPA Human Health <http://www.epa.gov/ebtpages/humanhealth.html>

NIOSH International Chemical Safety Cards, <http://www.cdc.gov/niosh/ipcs/icstart.html> (accessed FILL IN DATE)

U.S. National Library of Medicine *ToxNet* <http://toxnet.nlm.nih.gov/> (accessed FILL IN DATE)

U.S. Food and Drug Administration *Drug Info Rounds* <http://www.fda.gov/Drugs/default.htm> (accessed FILL IN DATE)

Chemical Book (online), [http://www.chemicalbook.com/ProductIndex\\_EN.aspx](http://www.chemicalbook.com/ProductIndex_EN.aspx) (accessed FILL IN DATE)

*Encyclopedia of Toxicology*, Elsevier

*Sittig's Handbook of Toxic and Hazardous Chemicals and Carcinogens*, Elsevier



# Large chemical information databases

- Haz-Mat Navigator
  - Initial Goal 50 new records per year
  - Less than 20 were actually added over 3 year period
- Retained Contractor 1 to actually generate the new records
  - Highly skilled chemist; MS level; experience in authoring MSDSs
- Retained Contractor 2 to review work for technical quality
- Cost about \$1000 USD per record



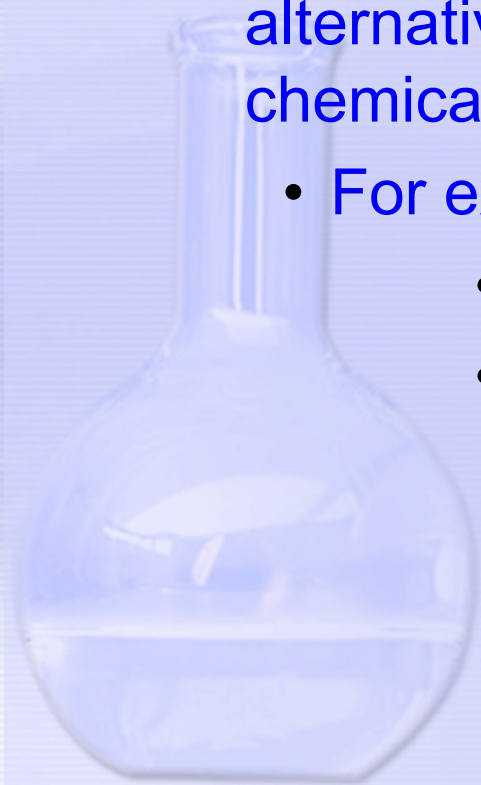
# Large chemical information databases

- IT Contractor (whose form was used) had significant difficulty converting WORD XML chemical information into format that would appear correctly on web
- ELS had difficulty with DB management and business model
- Project ultimately shut-down with the entire Navigator series



# Large chemical information databases

- SubChem
  - A work in progress
  - Purpose is to answer the question: “What are alternative solvents, based on specific physical and chemical properties?”
    - For example:
      - Flash point  $> 150^{\circ}\text{C}$
      - Viscosity 100 – 300 cp





# Large chemical information databases

- SubChem
  - Adding records
    - Identify “solvents” from source files
      - Manual task
    - Map fields in source file to 89 fields in existing file
      - Manual task
    - Develop rules to guide machine parsing of records
    - For example:

A chemical following the word “See” is listed. (the listing stopped at the “:”)

See Potassium dioxide: 2-Aminophenol. Tetrahydrofuran  
See Borane - tetrahydrofuran



# Large chemical information databases

- SubChem
  - Work with IT contractor to address questions
    - Unexpected formats
    - Unit inconsistencies



# Large chemical information databases

- Lessons Learned
  - Chemists must be part of the lead team
  - Information specialists who have some knowledge of chemistry are essential
  - Selected resources **MUST** be CAS Number based
  - If reactivity is to be included, narrative must be allowed
  - Unit consistency must be managed



# Large chemical information databases

- Lessons Learned – Business
  - Project will be much more difficult than anticipated
    - Adequate time and financial resources must be allocated
  - While subscription based service provides a predictable cash flow, a per search option must be available
  - Tying the DB into a widely used resource will help with success
    - For example –
      - Reaxys or ChemDraw

