

Reviewing PubChem laboratory chemical safety summaries for different user types

Brian Murphy

University of Arkansas, Little Rock

Ralph Stuart

Chemical Hygiene Officer, Keene State College

R.E. Belford

University of Arkansas, Little Rock

Leah McEwen

Chemistry Librarian, Cornell University

March, 2016

My Experience with On-Line Distance Education

- Developed Lab-XL project with EPA Region 1 and DC from 1996-2000
- Taught HAZWOPER (chemical safety for field students) with an on-line component from 2000 to 2011
- Participated in the 2004 OLCC On Lab Safety
- Used Mediawiki platform for student assignments and reports

The Current Context of Chemical Safety Information

- **Caveat emptor:** Chemistry textbooks and laboratory manuals provide a overview of generic rules, followed by "see the MSDS".
- Wikipedia provides links to random MSDS sources with no evidence of why that source is selected; some sources are kaput, many are dated
- Reports by the **National Research Council, the ACS, NFPA** after specific laboratory safety incidents found this approach to chemical safety education and information inadequate.



Material Safety

the handling of this c
cautions. It is high
safety Datasheet (M
and follow its directio

- Mallinckrodt Bake
- Science Stuff

October 2014
Key Lessons for Preventing Incidents from Flammable Chemicals in Educational Demonstrations
Eliminating Flash Fire Hazards by Substituting or Minimizing the use of Flammable Chemicals and Performing an Effective Hazard Review Will Prevent Injuries

Key Lessons Summaries

- Due to flash fire hazards and the potential for serious injuries, do not use bulk quantities of flammable chemicals in educational demonstrations when small quantities are sufficient.
- Employers should implement strict safety protocols when demonstrations include handling hazardous chemicals — including written procedures, effective training, and the removal of inappropriate personal protective equipment for all participants.
- Conduct a comprehensive hazard review prior to performing any educational demonstration.
- Provide a safety barrier between the demonstration and the audience.

October 2014
Key Lessons for Preventing Incidents from Flammable Chemicals in Educational Demonstrations
Eliminating Flash Fire Hazards by Substituting or Minimizing the use of Flammable Chemicals and Performing an Effective Hazard Review Will Prevent Injuries

le safety
u seek the Material
om a reliable source



The Structure of Chemical Safety Information

Chemical Safety Information Source Hierarchy

1990's

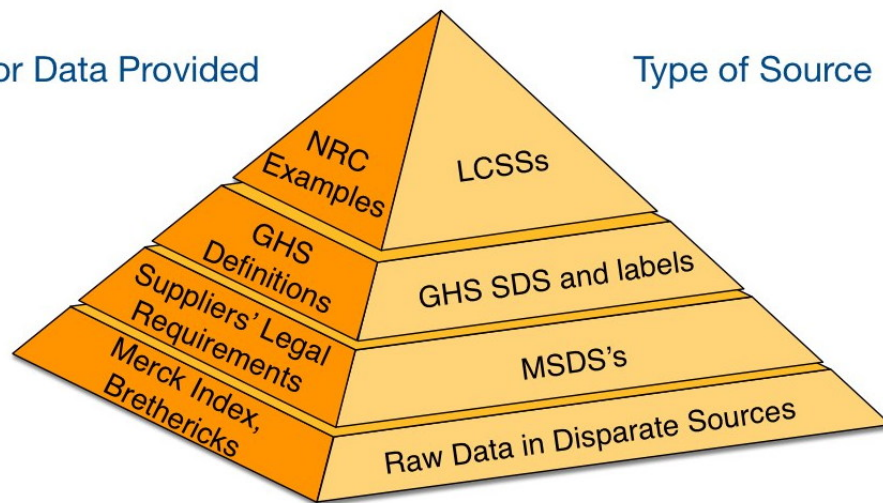
Basis for Data Provided

Type of Source

2000's

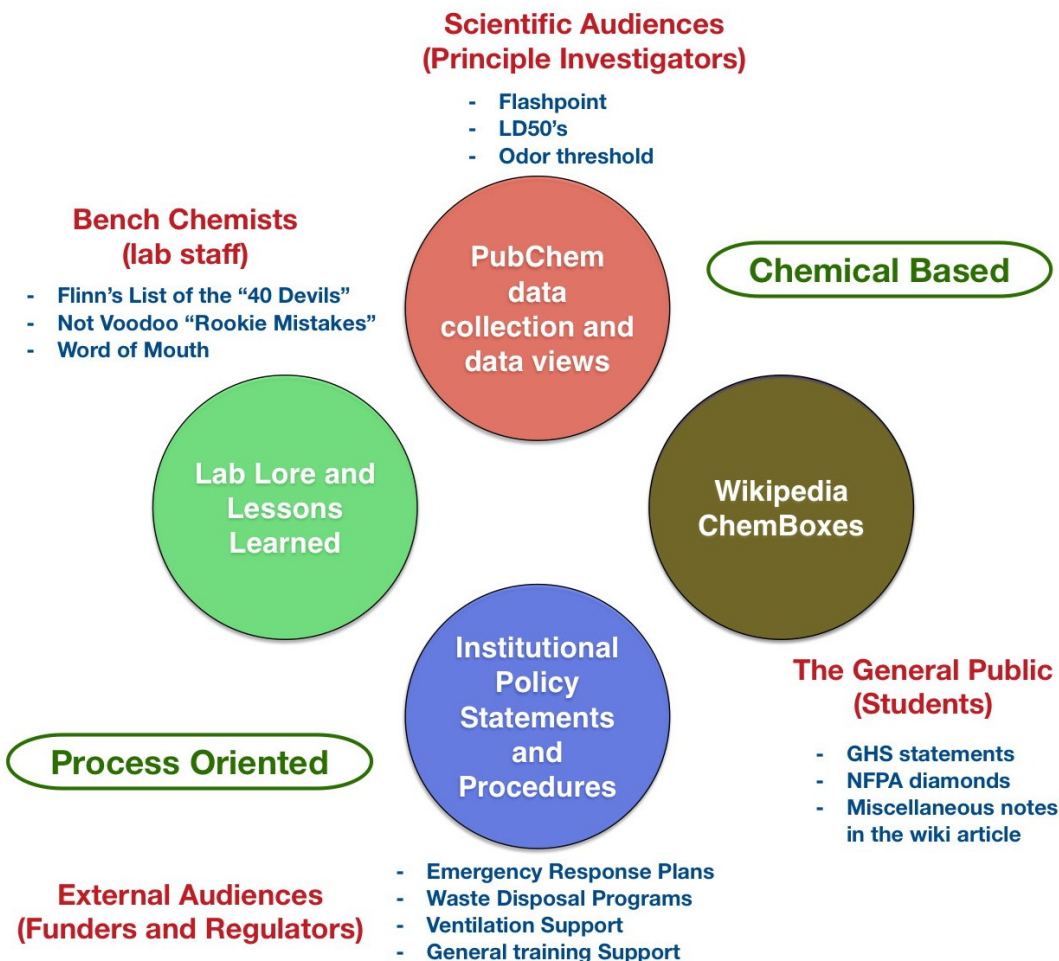
1980's

1970's



Increasingly well-defined; also increasing voluminous

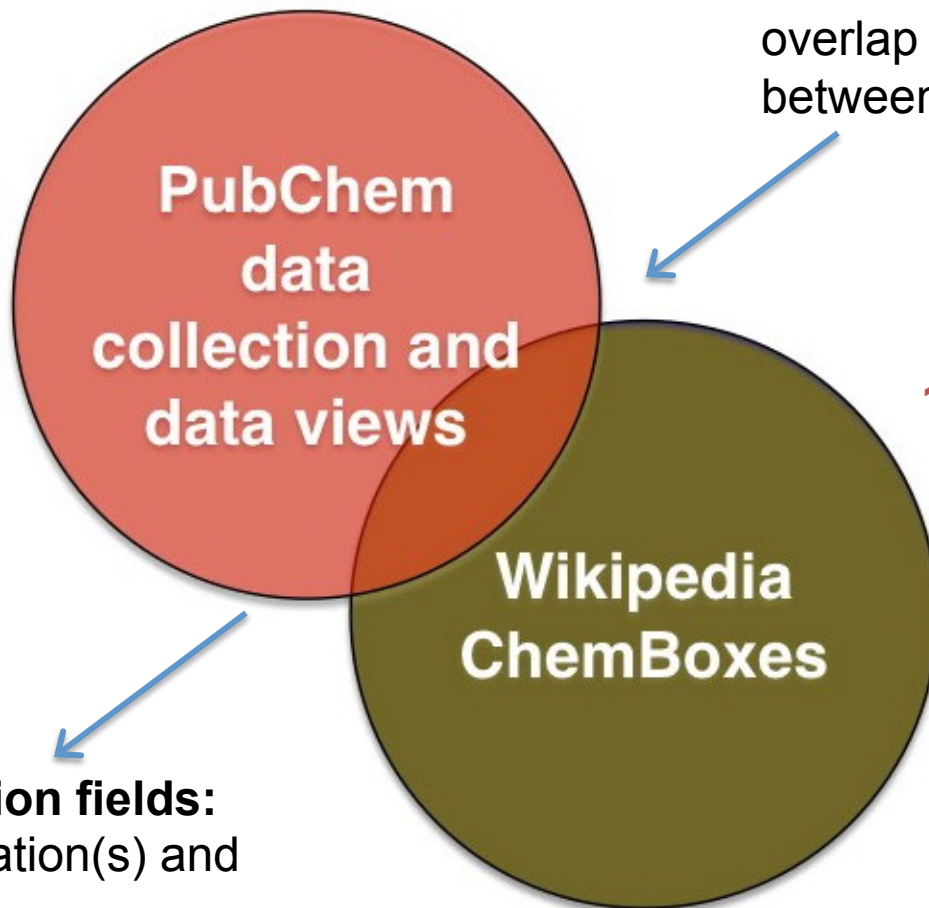
Less Structured Chemical Safety Information



Looking for Structure in the Electronic Data

- How large is the PubChem chem safety information universe?
- How high quality is it (including consistency & provenance information)?

Millions of chemicals;
3500 with GHS info



How much overlap is there between the two?

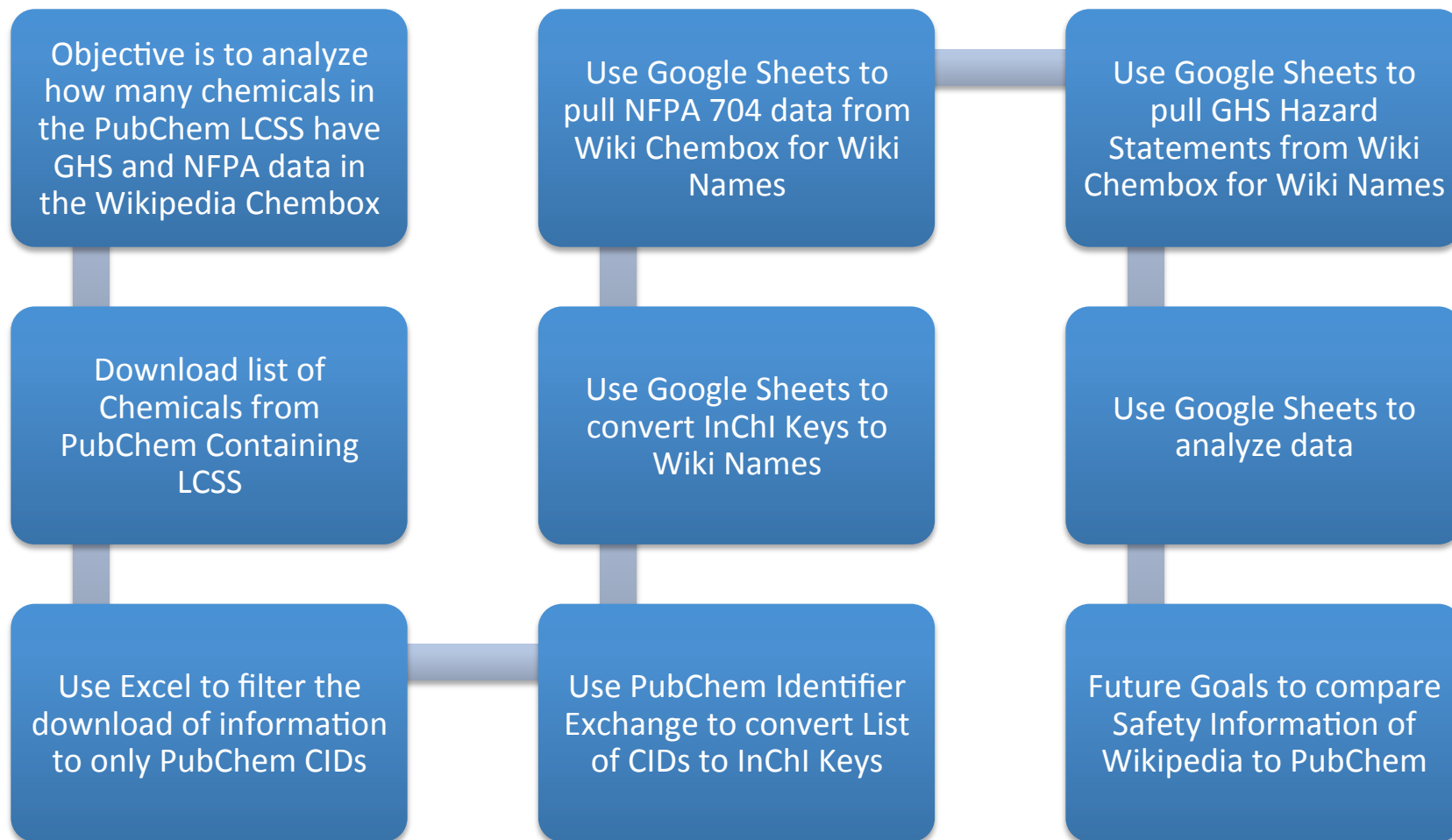
10,000 Chemboxes

- How large is the Wikipedia chem safety information universe?
- How high quality is it?

Key safety information fields:

- GHS class designation(s) and signal word
- NFPA diamond information

Project Overview



LCSS Data to InChI Key

PubChem Website

Obtain List of Chemicals with LCSS



Excel or Google Sheets

Filter down data from PubChem into a List of PubChem CIDs



PubChem Identifier Exchange

Convert List of PubChem CIDs to InChI Keys

Obtaining LCSS List of Chemicals

chemicals used, but also from 1) changes in the concentrations and quantities of those chemicals, 2) new chemicals that are produced, 3) energy sources that occur during a laboratory process, and other variables. For more information, see [this newsletter article](#) as well as this [PubChem Blog post](#).

The electronic form of the LCSS provided by PubChem is publicly accessible. LCSS data can be downloaded as a data stream [in bulk](#) or on-demand from the PubChem website (e.g., by following a link on a compound summary page). Although we are not aware of any limitations or restrictions on the reuse of PubChem LCSS data, we are not able to give unconditional permission for reuse and advise consultation with intellectual property experts when reusing this data. See disclaimer below for more information.

Examples

- [Acetone](#)
- [Benzene](#)
- [Ethanol](#)
- [Formaldehyde](#)
- [Hydrogen Cyanide](#)
- [Imidazole](#)
- [Phenolphthalein](#)
- [Phosphoric Acid](#)
- [Theophylline](#)
- [Toluene](#)

[See a list of all compounds with LCSS](#)

References

National Research Council. Prudent Practices in the Laboratory: Handling and Management of Chemical Hazards, Updated Version. Washington, DC: The National Academies Press, 2011. Free PDF available from <http://www.nap.edu/catalog/12654/prudent-practices-in-the-laboratory-handling-and-management-of-chemical>.

Click Here to View
List of Compounds with
an LCSS

Filter PubChem Information into CIDs

The screenshot shows an Excel spreadsheet with a list of PubChem IDs in column A. A red-bordered text box is overlaid on the spreadsheet, containing the following text:

First
=MID(Sheet0!A5,FIND("CID: ",Sheet0!A5)+5,LEN(Sheet0!A5))
Then
Filter off cells with errors
This will yield list of PubChem CIDs

The spreadsheet shows a list of PubChem IDs in column A, ranging from 102601870 to 91996449. The formula bar shows the formula: =MID(Sheet0!A5,FIND("CID: ",Sheet0!A5)+5,LEN(Sheet0!A5)).

Converting CID to InChI Key

The screenshot shows the PubChem Identifier Exchange web interface. The browser address bar displays `https://pubchem.ncbi.nlm.nih.gov/idxexchange/idxexchange.cgi`. The search bar contains `pubchem lcss`. The interface includes several sections:

- File Upload:** Options for "Browse..." (No file selected), "1: Select 3482 document(s) (3482)", and "Choose from Entrez History".
- Operator Type:** A dropdown menu set to "Same CID".
- Output IDs:** A dropdown menu with "CIDs" selected. Other options include Registry IDs, SIDs, InChIs, InChIKeys, SMILES, and Synonyms.
- Output Method:** A dropdown menu set to "Showing each input-output correspondence".
- Compression:** A dropdown menu set to "Gzip (.gz)".
- Actions:** Buttons for "Submit Job", "Save Job", "Load Job", and "Clear Form".

A red-bordered box highlights the text: "PubChem Identifier Exchange will convert PubChem CID to InChI Key".

InChI Key to Safety Information

Google Sheets

Convert InChI Key to Wiki Name using Importxml function



Google Sheets

Convert Wiki Name to Safety Information using Importhtml function



Google Sheets

Analyze Data

Wikipedia Chembox

Previously, acetone was produced by the **dry distillation** of **acetates**, for example **calcium acetate** in **ketonic decarboxylation**.

$$\text{Ca}(\text{CH}_3\text{COO})_2 \rightarrow \text{CaO}_{(s)} + \text{CO}_{2(g)} + (\text{CH}_3)_2\text{CO}_{(v)}$$

Before that, during **World War I** acetone was produced using **acetone-butanol-ethanol fermentation** with *Clostridium acetobutylicum* bacteria, which was developed by **Chaim Weizmann** (later the first president of **Israel**) in order to help the British war effort^[14] in the preparation of **Cordite**.^[29] This acetone-butanol-ethanol fermentation was eventually abandoned when newer methods with better yields were found.^[14]

Uses [edit]

About a third of the world's acetone is used as a solvent, and a quarter is consumed as **acetone cyanohydrin**, a precursor to **methyl methacrylate**.^[13]

Solvent [edit]

Acetone is a good solvent for many plastics and some synthetic fibers. It is used for thinning **polyester** resin, cleaning tools used with it, and dissolving two-part **epoxies** and **superglue** before they harden. It is used as one of the volatile components of some **paints** and **varnishes**. As a heavy-duty degreaser, it is useful in the preparation of metal prior to painting. It is also useful for high reliability **soldering** applications to remove **rosin** flux after soldering is complete; this helps to prevent the **rusty bolt effect**.




Acetone is used as a solvent by the **pharmaceutical industry** and as a **denaturant** in **denatured alcohol**.^[30] Acetone is also present as an **excipient** in some **pharmaceutical drugs**.^[31]

Although itself **flammable**, acetone is used extensively as a solvent for the safe transportation and storage of **acetylene**, which cannot be safely **pressurized** as a pure compound. Vessels containing a porous material are first filled with acetone followed by acetylene, which dissolves into the acetone. One liter of acetone can dissolve around 250 liters of acetylene.^{[32][33]}

Chemical intermediate [edit]

Acetone is used to **synthesize methyl methacrylate**. It begins with the initial conversion of acetone to **acetone cyanohydrin**:

$$(\text{CH}_3)_2\text{CO} + \text{HCN} \rightarrow (\text{CH}_3)_2\text{C}(\text{OH})\text{CN}$$

entropy (<i>S</i> ₂₉₈)	
Std enthalpy of formation (<i>Δ</i> _f <i>H</i> ₂₉₈)	−250.03-(−248.77) kJ/mol
Std enthalpy of combustion (<i>Δ</i> _c <i>H</i> ₂₉₈)	−1.772 MJ/mol
Hazards	
Safety data sheet	See: <i>data page</i>
GHS pictograms	
GHS signal word	DANGER
GHS hazard statements	H225, H319, H336
GHS precautionary statements	P210, P261, P305+351+338
EU classification (DSD)	
R-phrases	R11, R36, R66, R67
S-phrases	(S2), S9, S16, S26
NFPA 704	
Flash point	−20 °C (−4 °F; 253 K)
Autoignition temperature	465 °C (869 °F; 738 K)
Explosive limits	2.6–12.8% ^[10]
Threshold Limit Value	1185 mg/m ³ (TWA), 2375 mg/m ³ (STEL)
Lethal dose or concentration (LD, LC):	

InChI Key to Wiki Name

Firefox File Edit View History Bookmarks Tools Window Help

Google Sheets - create and ... x InChI Key - Google Sheets x InChI Key to GHS NFPA - G... x W IMNFDUFMRHMDMM-UHF... x

https://en.wikipedia.org/wiki/Special:Search?search=IMNFDUFMRHMDMM-UHFFFAOYSA-N&go=Go | wikipedia

Most Visited Getting Started 9 ft. Wimberly Spr... Holiday Time Pre... Holiday Time Unit ... Stack-On 35-Gun ... Rowley Solos | Sh... Ludlow | Shop at V... C&L Madero | Sho...

Not logged in Talk Contributions Create account Log in

Special page

Search results

Result 1 of 1

[Content pages](#) [Multimedia](#) [Everything](#) [Advanced](#)

*The page "**IMNFDUFMRHMDMM-UHFFFAOYSA-N**" does not exist. You can ask for it to be created, but consider checking the search result below to see whether the topic is already covered.*

Heptane
InChI InChI=1S/C7H16/c1-3-5-7-6-4-2/h3-7H2,1-2H3 Y Key: **IMNFDUFMRHMDMM-UHFFFAOYSA-N** Y SMILES CCCCCC Properties Chemical formula C7H16
8 KB (604 words) - 10:13, 12 March 2016

**Spreadsheet searches Wikipedia
For the InChI key then pulls
Suggested name.**

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
Upload file
Special pages
Printable version

Languages

Privacy policy About Wikipedia Disclaimers Contact Wikipedia Developers Cookie statement Mobile view

WIKIMEDIA project
Powered By MediaWiki

Wiki Name to Safety Information

- Google Sheets
 - Importhtml function
 - Pulls Safety Information using Wiki Name
- =IFError(query(IMPORTHTML(CONCATENATE("https://en.wikipedia.org/wiki/",C2719),"Table",1),"SELECT Col2 where Col1='GHS hazard statements'",0),"No GHS Hazard Statement")

The Early Statistics




- PubChem has an LCSS view for about 3500 (2000 more to come soon) chemicals; Wikipedia has Chemboxes for about 10,000 chemicals
- Of those in the PubChem LCSS collection, about 30% have an entry in Wikipedia
- 4% of the Pubchem collection has GHS information; 12% of the PubChem collection have NFPA diamond information

	Not in Wikipedia	In Wikipedia	GHS Hazard Statement	NFPA 704	Total
n	2441	1038	157	431	3486
%	70.02%	29.78%	4.50%	12.36%	

Future Directions

- Understand the Wikipedia Chembox structure to collect information more efficiently
- Develop a Wikipedia – PubChem data link that can provide chemical safety information with provenance data to the ChemBox
- Consider what chem safety data makes sense to put in the Wikipedia Chembox and what can be linked to there

The hazard portion of the **acetone** chembox

Hazards	
Safety data sheet	See: data page
GHS pictograms	
GHS signal word	DANGER
GHS hazard statements	H225, H319, H336
GHS precautionary statements	P210, P261, P305+351+338
EU classification (DSD)	
R-phrases	R11, R36, R66, R67
S-phrases	(S2), S9, S16, S26
NFPA 704	
Flash point	-20 °C (-4 °F; 253 K)
Autoignition temperature	465 °C (869 °F; 738 K)
Explosive limits	2.6–12.8% ^[10]
Threshold Limit Value	1185 mg/m ³ (TWA), 2375 mg/m ³ (STEL)
Lethal dose or concentration (LD, LC):	
LD ₅₀ (Median dose)	5800 mg/kg (rat, oral) 3000 mg/kg (mouse, oral) 5340 mg/kg (rabbit, oral) ^[11]
LC ₅₀ (Median concentration)	20,702 ppm (rat, 8 hr) ^[11]

Lessons Learned about this OLCC project

- Benefits: Cross campus collaboration on "specialty" topics helps:
 - Connect with people doing similar work
 - Provide access to experts
 - Develop real world products
- Challenges: Distance education (learning, teaching and collaboration)
 - Synchronizing schedules (knowing what has happening on the other end)
 - Assessing understanding, both orally and written
 - Maintaining focus on the overall goal while accepting project drift

