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.

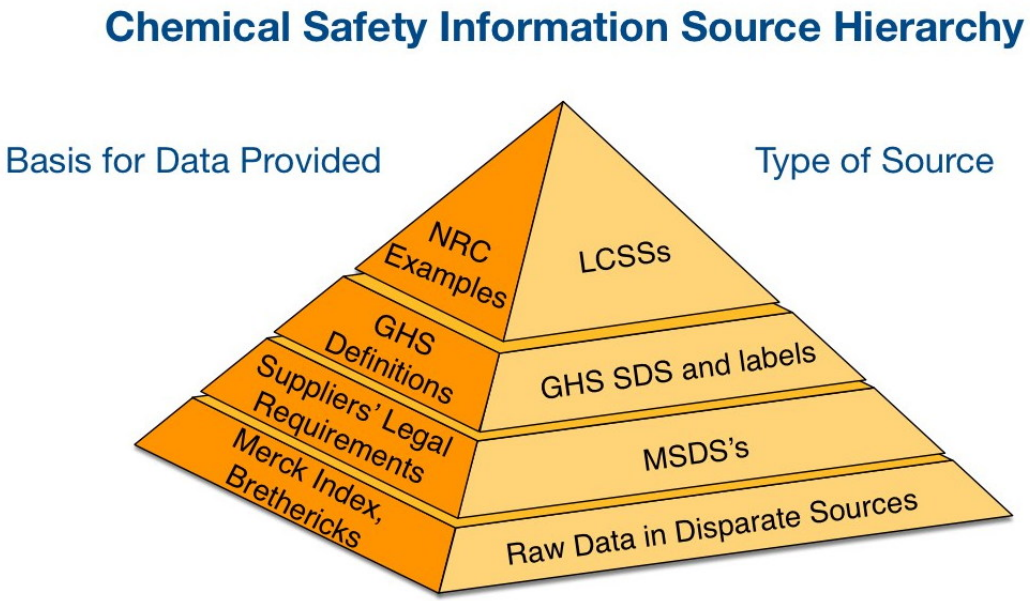
The Structure of Chemical Safety Information

1990's

2000's

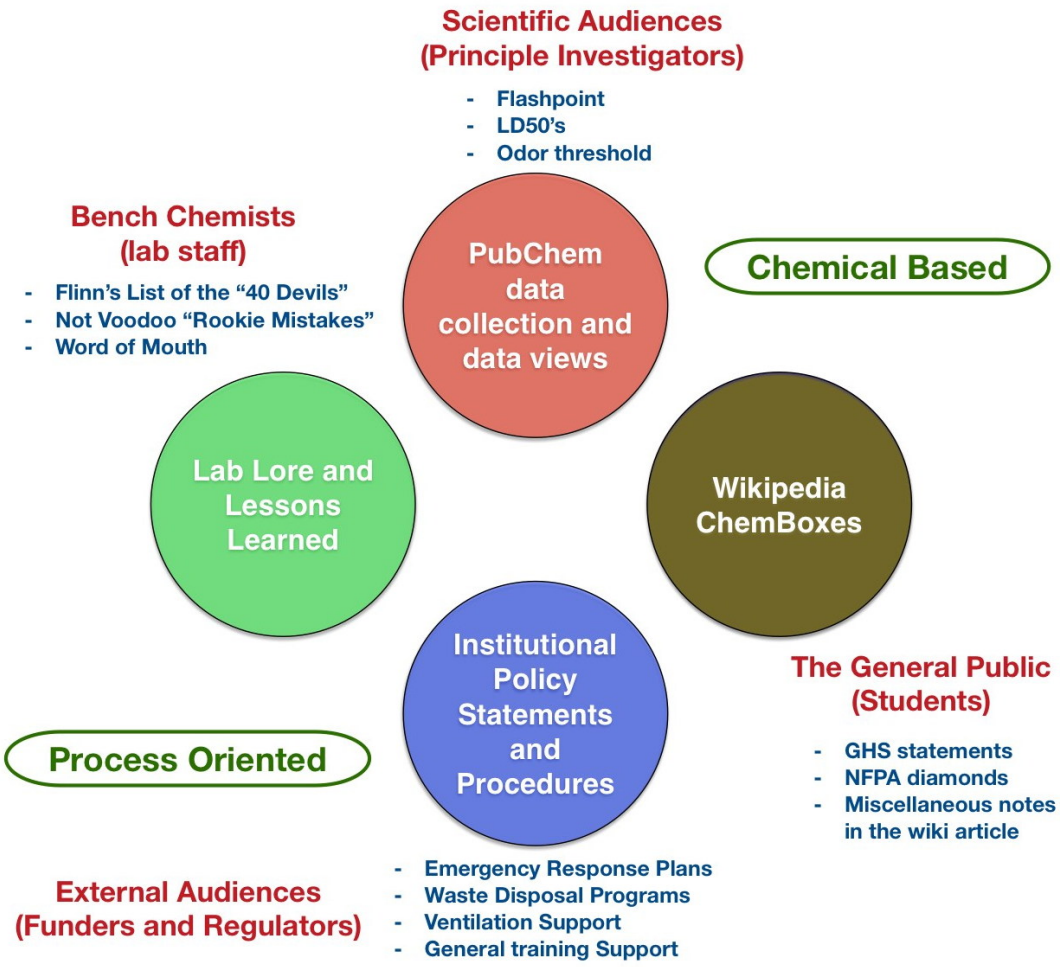
1980's

1970's



Increasingly well-defined; also increasing voluminous

Less Structured Chemical Safety Information

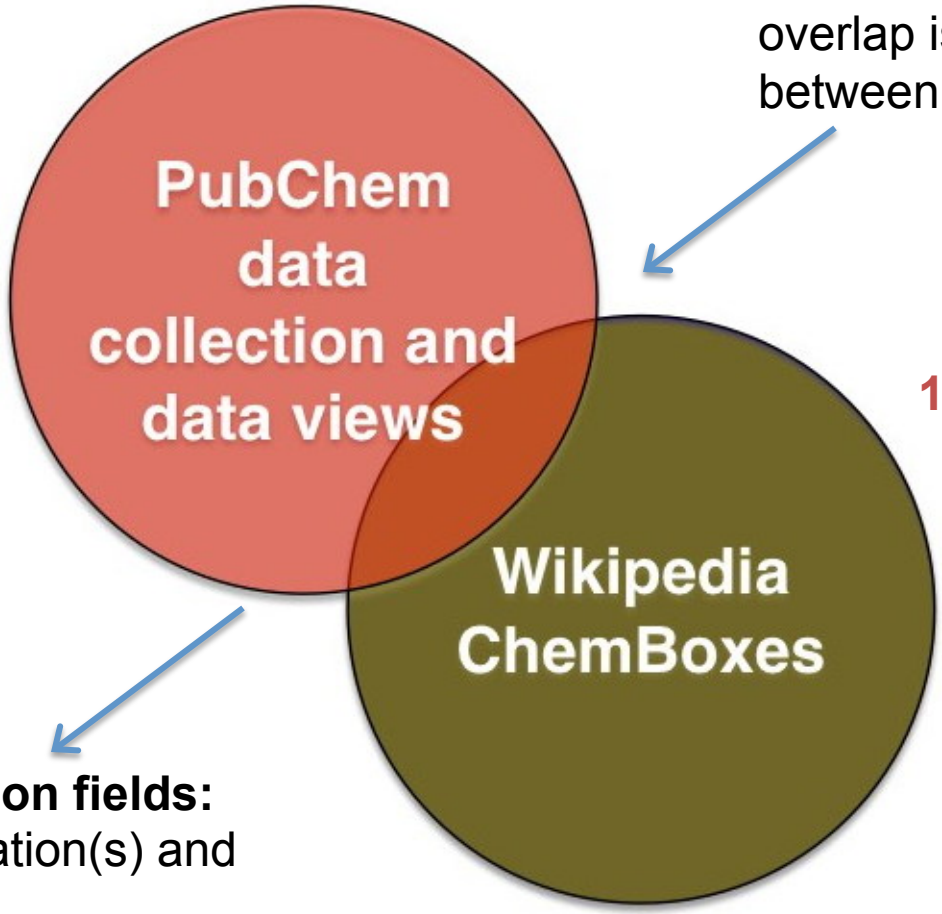


Looking for Structure in the Electronic Data

- 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?



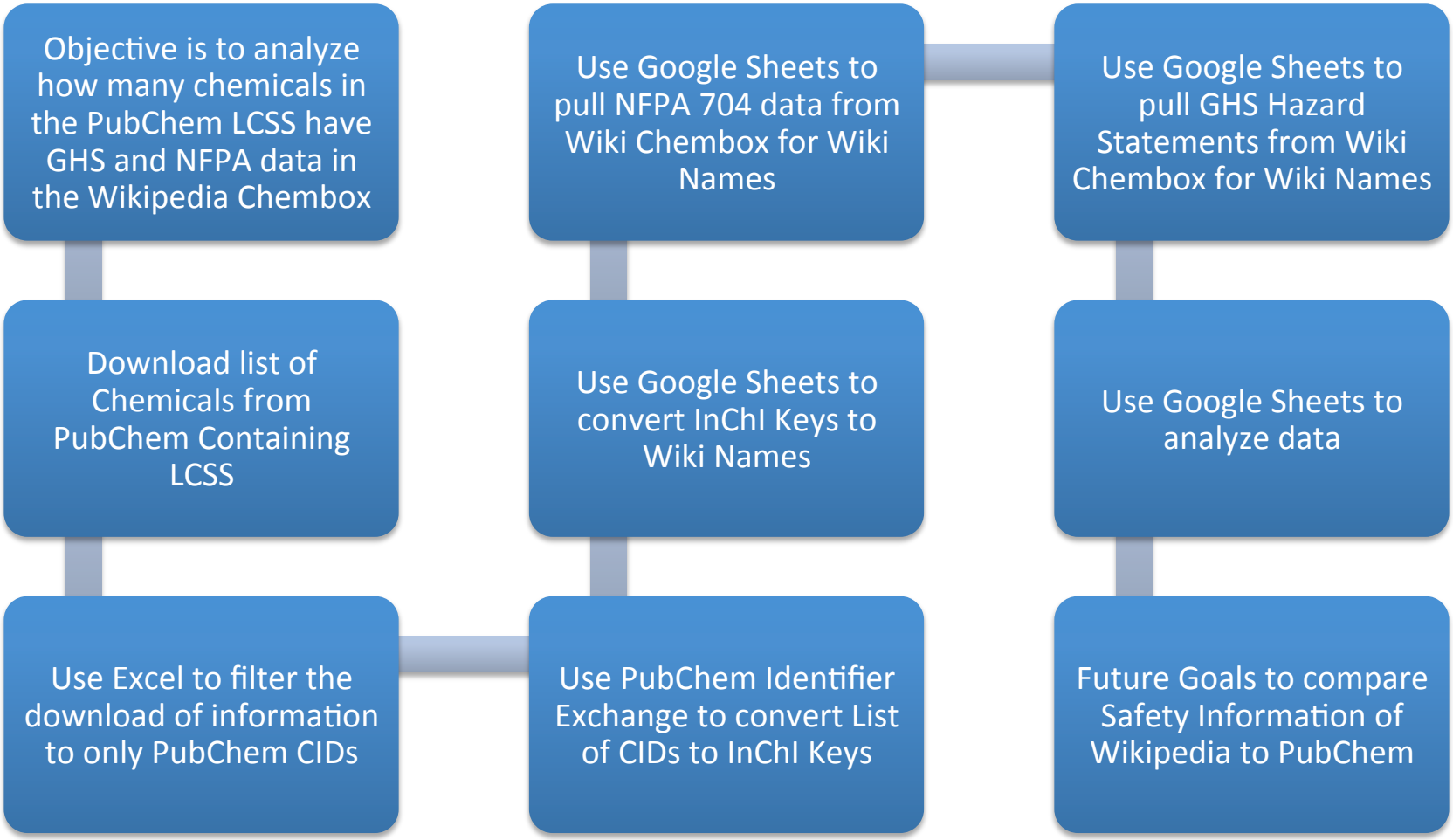
10,000 Chemboxes

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

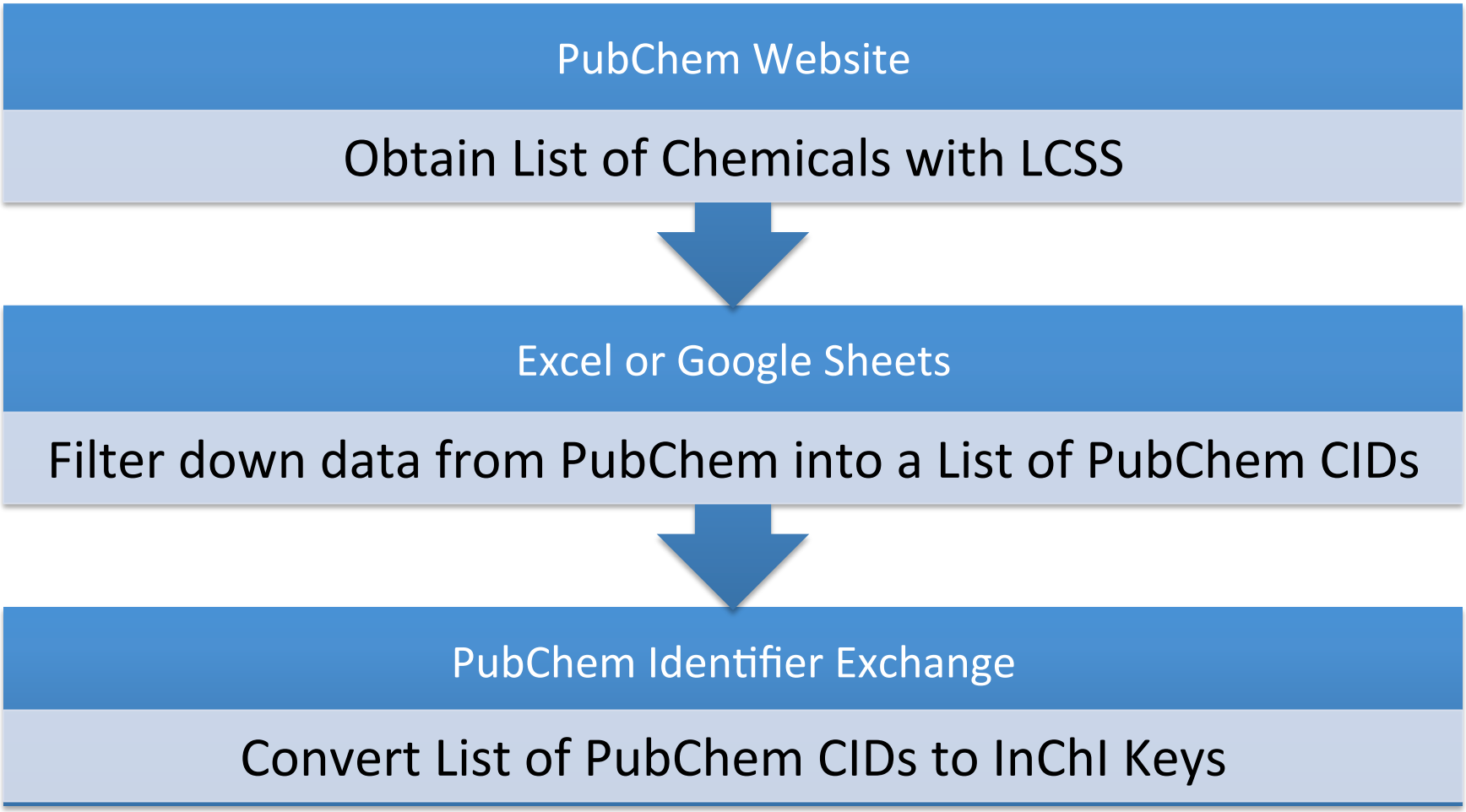
Key safety information fields:

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

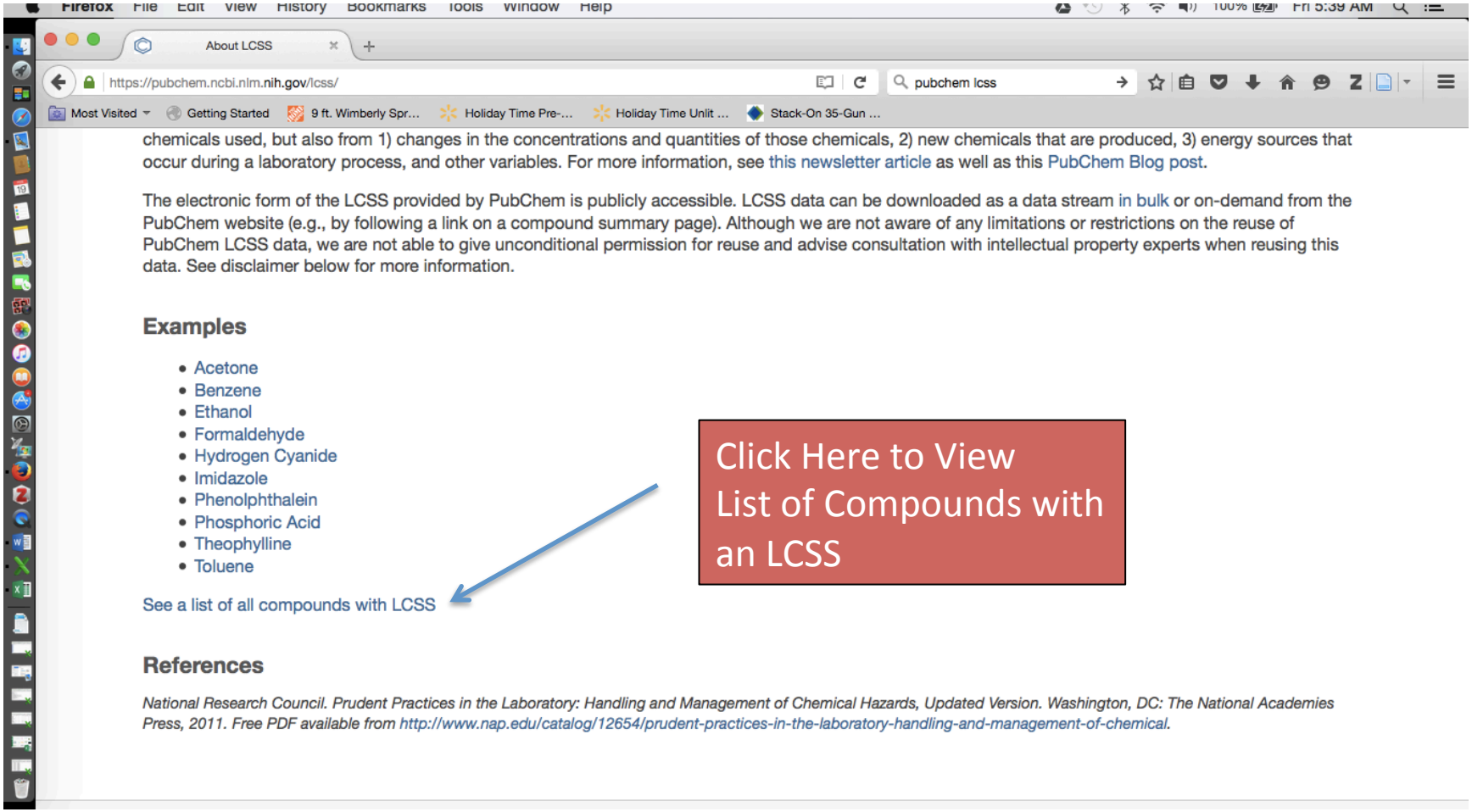
Project Overview



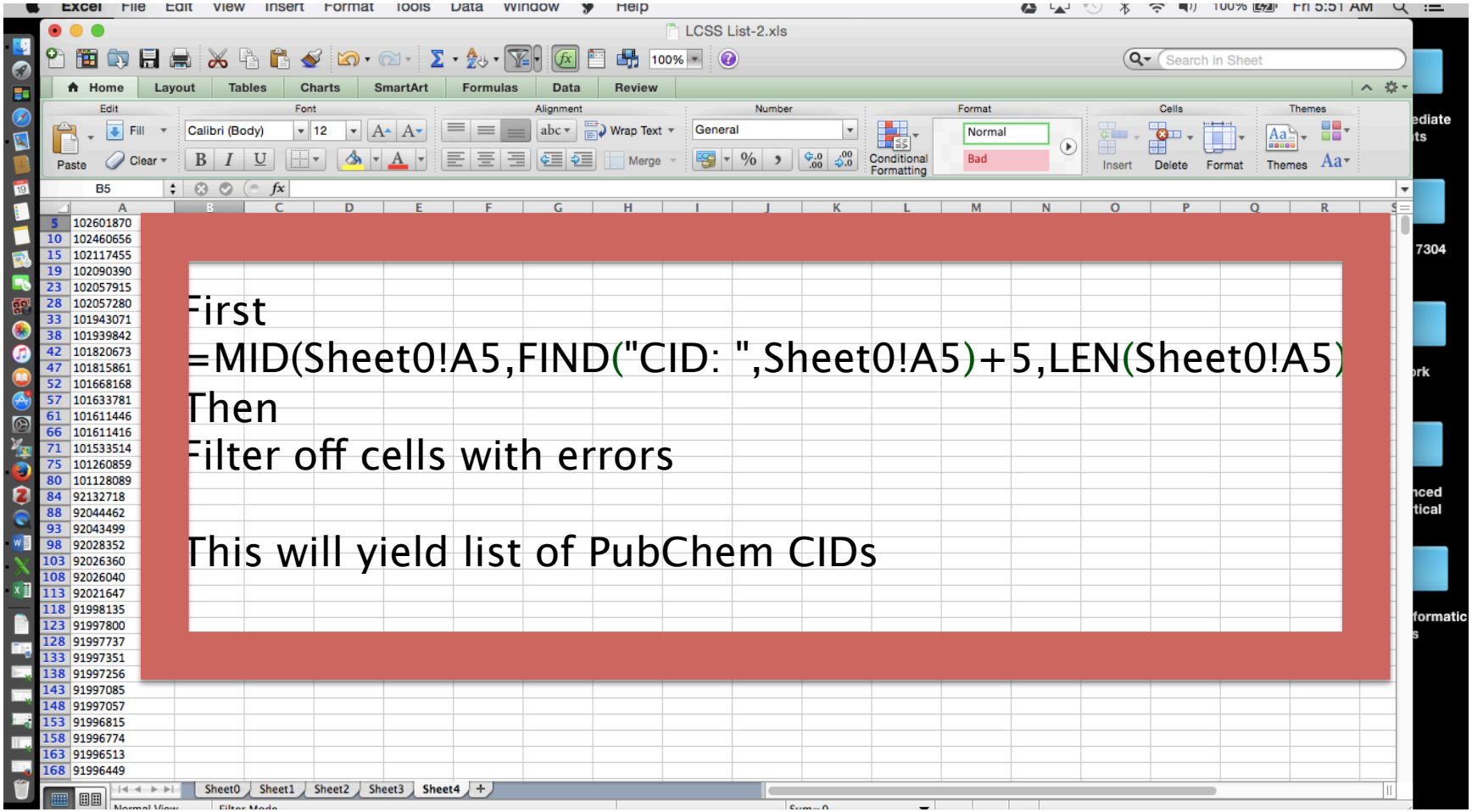
LCSS Data to InChI Key



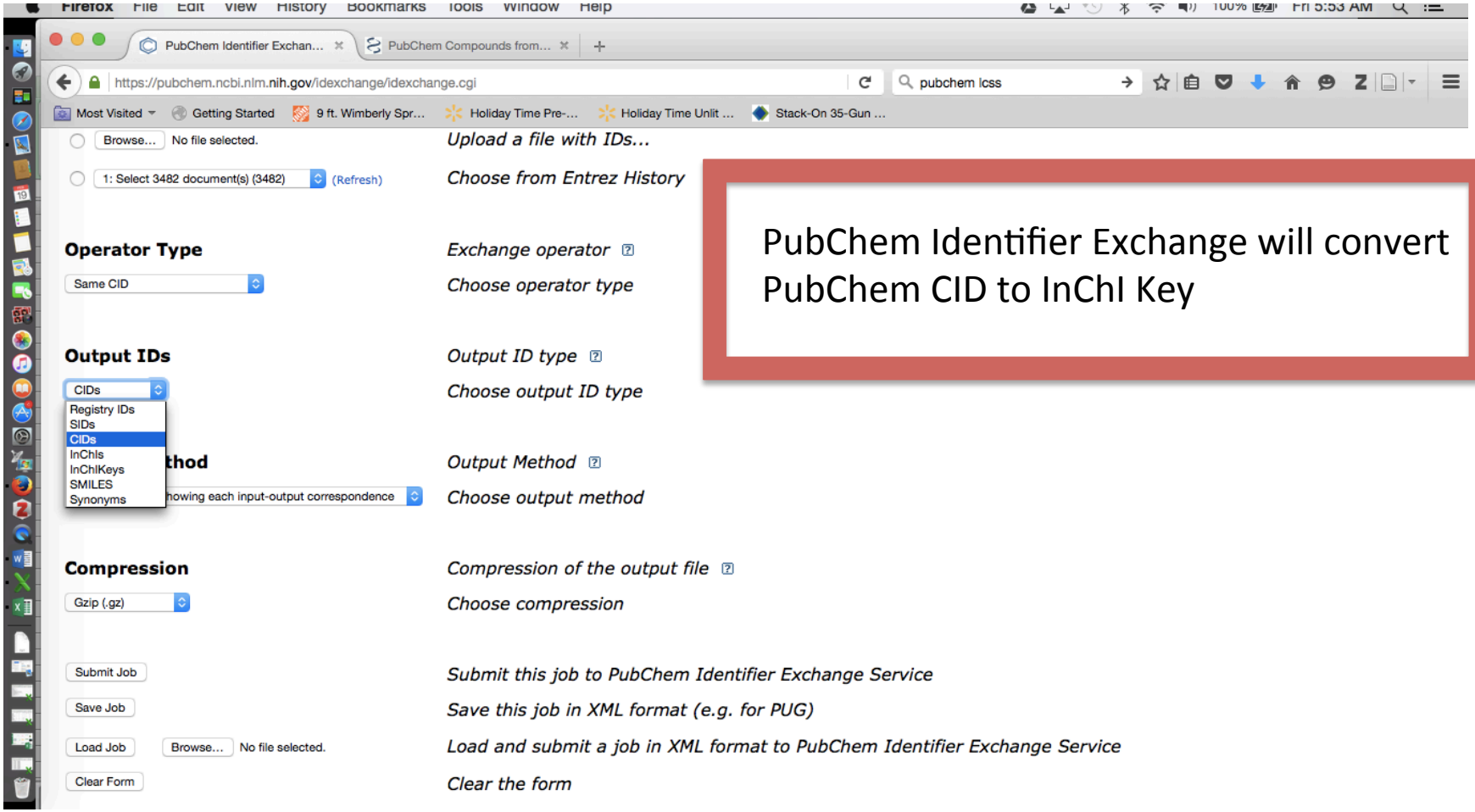
Obtaining LCSS List of Chemicals



Filter PubChem Information into CIDs

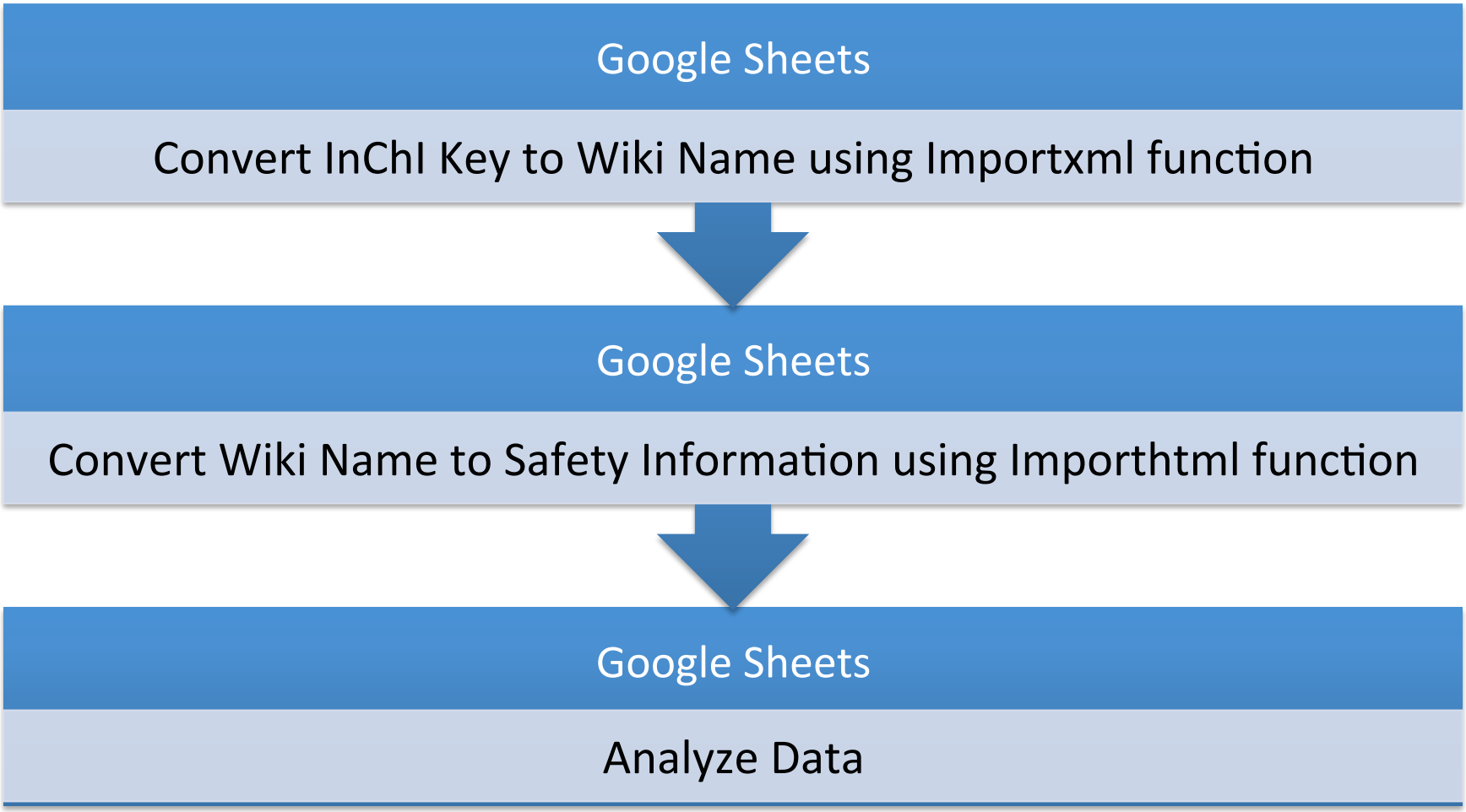


Converting CID to InChI Key

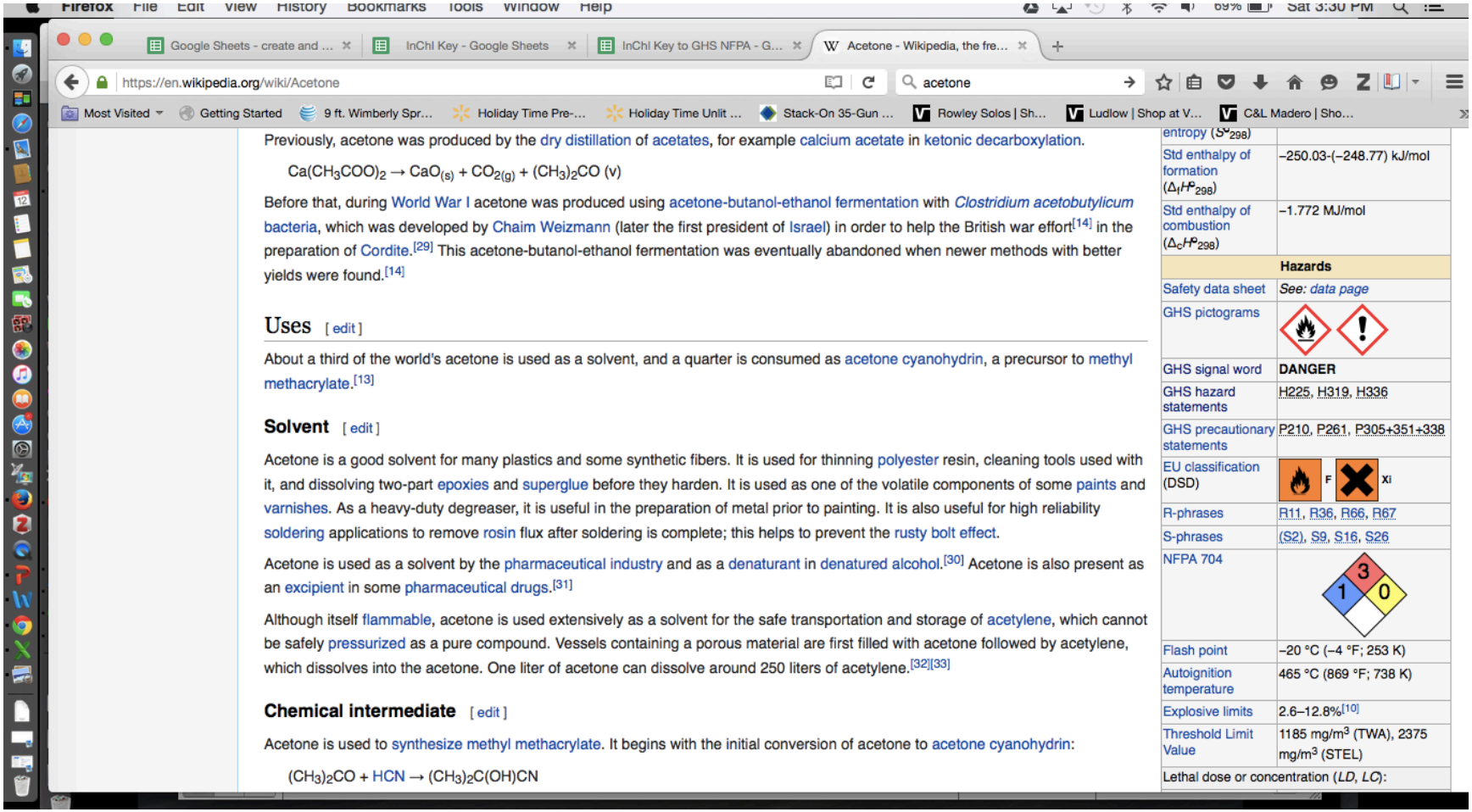


PubChem Identifier Exchange will convert PubChem CID to InChI Key

InChI Key to Safety Information



Wikipedia Chembox



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

$$Ca(CH_3COO)_2 \rightarrow CaO_{(s)} + CO_{2(g)} + (CH_3)_2CO\ (v)$$

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]

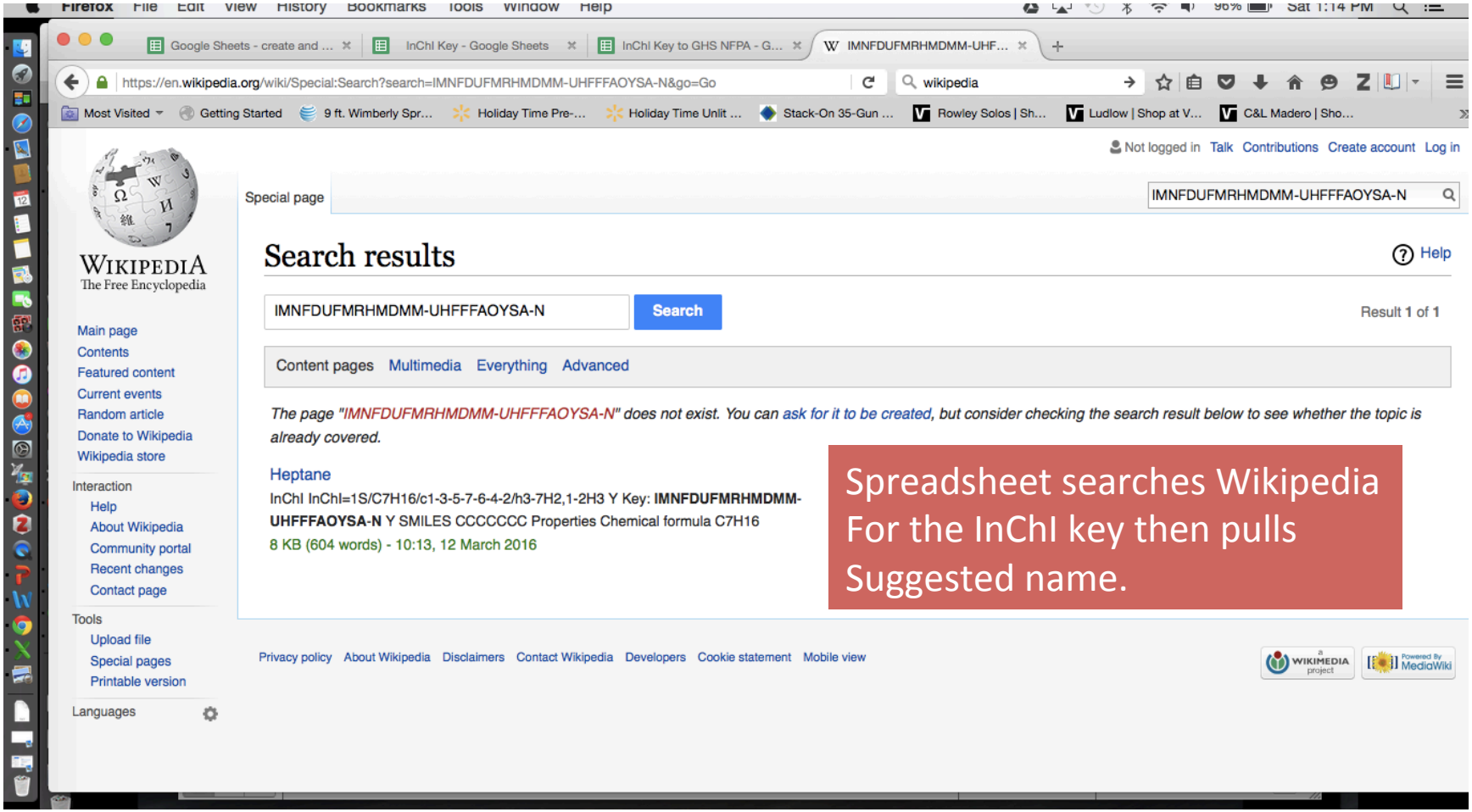
Chemical intermediate [edit]

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

$$(CH_3)_2CO + HCN \rightarrow (CH_3)_2C(OH)CN$$

entropy (S°_{298})	
Std enthalpy of formation ($\Delta_f H^\circ_{298}$)	−250.03-(−248.77) kJ/mol
Std enthalpy of combustion ($\Delta_c H^\circ_{298}$)	−1.772 MJ/mol
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)	F Xi
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/m³ (TWA), 2375 mg/m³ (STEL)
Lethal dose or concentration (*LD*, *LC*):	

InChI Key to Wiki Name



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

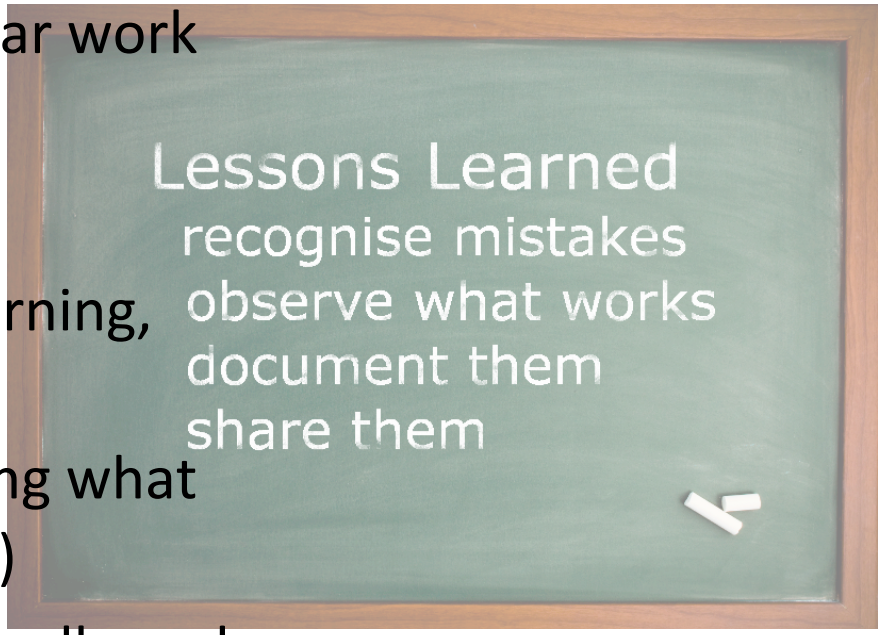
	Not in Wikipedia	In Wikipedia	GHS Hazard Statement	NFPA 704	Total
n	2441	1038	157	431	3486
%	70.02%	29.78%	4.50%	12.36%	

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 provence 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)	F Xi
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