MIXTURES INCHI OVERVIEW

Alex M. Clark
Since Last Meeting

- Paper published in *Journal of Cheminformatics*
  - 11, article 33 (May 2019)
- iCorps program completed
  - customer research
- Phase 2 grant (NIH/SBIR) reviewed...
  - ... fingers crossed
Previous Work Summary

- Mixtures didn't have any effective standard data format, and urgently needed one
- Defined a simple *Mixfile* format, analogous to *Molfile* (which it encapsulates)
- Open source tool for editing/manipulating, and text-extracted data
- Designed to *feed* the Mixtures InChI (*MlnChI*)
- Proof of concept MInChI assembly algorithm
The MinChl

Mixtures InChI
3 colleagues at Collaborative Drug Discovery signed on to do 100 interviews with potential customers

Learned a lot. Some highlights…

Bliss: most small drug discovery groups don’t think it’s an informatics problem, unless it’s pointed out

- generally enthusiastic conditional on tooling
- 1% refused to concede pen & paper
iCorps, ctd

- **Vendors**: best way is to capture mixtures on the way in, purchase → ELN; for ad hoc notebooks, MInChI as a footnote?

- **Value**: besides features like searching & inventory, machine readable mixtures will be noticed during acquisition or filing (e.g. FDA)

- **Properties**: not everyone cares what’s in a mixture, but they often care about it’s properties, like...
  - pH, density, boiling, melting, vapour pressure, solubility, stability, handling/safety, …
Labelling: lack of standard format often means that samples are labelled badly or not at all; everyone has a story...

Tracking: in all industries, often nobody really knows what a substance actually is - used as an excuse for lax quality control

Biomedical: formulations, screening samples, natural products (real world samples & synthetic mixing) all tend to be underdefined
iCorps, ctd

- **Software**: informaticians (like us) want a common interchange format

- **Identifier**: obtaining a CASRN for every mixture permutation is inconvenient; could use MInChI instead of a numeric database sequence

- **Markup**: suggestions that someone should grab content from important vendors and mark it all up, without asking permission
Next Phase for CDD

Assuming that omens check out...

1. Extend Mixfile to handle more molecule types (inorganics, ceramics, polymers, variations...)
   - in sync with downstream MInChI

2. Major enhancements for text-to-mixture: both public and private collections... much content

3. Integration into ELN software
Acknowledgments

- Leah McEwen & Gerd Blanke
  - and IUPAC/InChI Trust
- Hande Küçük McGinty
  - and Collaborative Drug Discovery