Event: WorldFAIR Chemistry Webinar: “What is a Chemical? User Perspectives on Digital -Machine Readable Depictions”. December 08, 2022. 11-12:30 pm EST.

- ***This space is to share notes/comments and discuss the topic of the event above. This document is available to be downloaded and shared.***

- **Agenda:**

Flash Talks (Moderated by: Ian Bruno)

- ***InChI: Greg Landrum***
  - Greg asks ChatGPT: what is InChI? What is chemical representation? What is a chemical identifier? Is InChI a representation or an identifier?
  - Key feature of InChI: comparing chemical compounds to determine if they are the same
  - Is InChI a representation or an identifier? ChatGPT confused and confusion in the community also…
  - InChI primarily designed as identifier, and based on the way it built, also encodes representation (with some caveats on how it functions as a standard)

- ***SMILES: Vincent Scalfani***
  - Vin uses SMILES for searching in chemistry databases (SciFinder, Pubchem), SMARTS for substructure and pattern searching, programmatic searching in web services
  - Many extensions to SMILES (dative bonds, polymers, machine learning)
  - Interesting history for managing exchange with different implementations
  - Removing aromaticity helps reduce ambiguities…
  - May be interesting to consider how well notations based on computing paradigms of past are suited for ML

- ***HELM: Dana Vanderwall***
  - Wiki.openhelm.org
  - HELM community built and supported, originally via Pistoia Alliance

**WorldFAIR “Global cooperation on FAIR data policy and practice” is funded by the EC HORIZON-WIDERA-2021-ERA-01-41 Coordination and Support Action under Grant Agreement No. 101058393.*

Contact us : FAIRChemistry@iupac.org

@FAIRChemistry *FAIRChemistry*

- Hierarchical approach enables description from high level macro to specific atom level
- Ambiguity prevalent in biological scenarios; linkages are key to describe, but how notate if variable or uncharacterized?
- HELM notation suggests the nature of what is connected at what ratios, without specifying where

- ***Graphical Representation: Jonathan Goodman***
  - IUPAC graphical representation recommendations
    - doi: 10.1351/pac200678101897
    - doi: 10.1351/pac200880020277
  - Many ways to represent molecular entities graphically, but may need to sort through full papers to interpret
  - Line drawings streamlined and have been defined in IUPAC recommendations above
  - Not always to draw consistently, can surface issues ambiguity for machines (e.g., unspecified stereo) and how different systems have been set up to address

- ***Systematic Representation: Michelle Rogers***
  - Nomenclature for capturing unique chemical identity, different approaches via structure, but also source-based nomenclature (e.g., natural products) and process-based nomenclature (e.g., UVCBs)
  - Capturing every unique isomer may not address what is put on the market in industry context
  - Source-based nomenclature may not be compliant for reporting, does not address potential genetic modifications of biological sources
  - Process-based nomenclature not very systematic, and nuances may be relevant in industry contexts but not always considered from chemical perspective

**WorldFAIR “Global cooperation on FAIR data policy and practice” is funded by the EC HORIZON-WIDERA-2021-ERA-01-41 Coordination and Support Action under Grant Agreement No. 101058393.*

Contact us : FAIRChemistry@iupac.org

@FAIRChemistry *FAIRChemistry*

Event: WorldFAIR Chemistry Webinar: “What is a Chemical? User Perspectives on Digital -Machine Readable Depictions”. December 08, 2022. 11-12:30 pm EST.

Panel Discussion

- What are current key gaps for current chemical representation
    - Biomolecules, can of worms, need something to help define/determine uniqueness, capturing process (impacts ratios), other areas of biomolecules (glycans, lipids)
    - Small, need adequate organometallics (graphical rep not standardized), isomers
    - Collecting critical information to make determinations on consistent representation
    - Users should not have to worry about toolkit level inconsistencies, in meantime, please cite software for downstream users
    - Pursuit of excellence is moving target
    - Standardization of description for non-standard structures
    - Variability in interpretation of structures in different jurisdictions
- Michelle:
    - Do you know of any indexing/systems that attempt to link the three representations that you spoke of? So that if you search for one, you'll get pointed to the other two?
        - Depends on what can be called the same and searching across, no good cross-linking option
        - How to cross link across different possibilities in different spaces
    - How do you represent a created mixture of unique chemicals and differentiate this to a "natural product", for example cocoa oil?
        - Hard to enumerate, infinite combinations often;
        - Modeling depends on structure chosen to put in
        - Formulation very local system dependent
- What do users need to understand about chemical depiction in a machine-readable environment? Bridging humans aspiring to visual and machines interpreting precise notation
    - General awareness of common depictions in users community to appreciate how to use these

**WorldFAIR “Global cooperation on FAIR data policy and practice” is funded by the EC HORIZON-WIDERA-2021-ERA-01-41 Coordination and Support Action under Grant Agreement No. 101058393.*

Contact us : FAIRChemistry@iupac.org

Event: WorldFAIR Chemistry Webinar: “What is a Chemical? User Perspectives on Digital -Machine Readable Depictions”. December 08, 2022. 11-12:30 pm EST.

    - Clarification of best way to use depictions in different applications
    - Understand what can and cannot pass from sketching tools into standard formats (and thus lost to further machine interpretation)
    - Tools are critical to bridge visual and digital representations, what can be
- Where should we be focusing on for improved representations?
    - As losslessly as possible, to support interoperability of FAIR, similar to approach with SMILES+
    - InChI good for Findability, especially when it expands to other areas (e.g., organometallics)
    - Education - tables of what can be done with different representations, document common use cases
    - Determining what is critical about different classes of entities (small well characterized, what do we need to know about proteins, other macromolecules, nanomaterials)
    - What are the key parameters to capture in representations? May need different approaches for different applications
    - Classification is another area that intersects with describing chemical moieties
- Interplay between the “standards” organizations, the “customers” (big pharma for example) and the vendors/skechers (ChemDraw, ChemAxon, BioviaDraw etc) - who should have the final say, or should we just have multiple options to keep everyone happy?
    - Example: cases with mol whereby different applications interpret the same mol files as different molecules (This relates to where the mol file has ambiguity as different areas of the file contradict each other).
    - Test suites for interoperability
    - Common, agreed upon baseline with authoritative definition (e.g., IUPAC document), well supported, with options for extensions to share at community level - need this for mol formats

**WorldFAIR “Global cooperation on FAIR data policy and practice” is funded by the EC HORIZON-WIDERA-2021-ERA-01-41 Coordination and Support Action under Grant Agreement No. 101058393.*

Contact us : FAIRChemistry@iupac.org

    - Always a bit of community development around how we share information, standardizing will lag this; standards often building on grass-roots efforts
- Education, training
    - Cookbooks for users to use in courses
    - May depend on role on how representation will be used in work
    - Basic guides that describe what is or not included in different representations
    - Drawing resources are generally pretty good, how to introduce the challenges with the nuances?
        - Precise structures?
        - Different interpretations translating to machine-readable depictions?
        - Validation suites that demonstrate potential ambiguity back to users
    - General competencies
    - Gamifying courses to identify “correct structure”

**WorldFAIR “Global cooperation on FAIR data policy and practice” is funded by the EC HORIZON-WIDERA-2021-ERA-01-41 Coordination and Support Action under Grant Agreement No. 101058393.*

Contact us : FAIRChemistry@iupac.org

@FAIRChemistry *FAIRChemistry*