Extending the Allotrope Foundation Ontology: An Ontological Representation and Analysis of Process Chemistry

Oliver He, Wes Schafer
Outline

- Acknowledgements
- The Importance of Context
- The Scope of Process Chemistry
- Survey of Existing Ontologies and Terminologies
- Proposal for Process Chemistry Ontology
  - OPC development strategy and methodology
  - Upper level hierarchy and design patterns
- 2 Use Cases
  - Fate and purge
  - DoE: Design of Experiment
- Example OWL Definitions
- Summary & Discussion
The Importance of Context

- Experimental and workflow details need to be linked to the analytical data in order for it to have context and to enable data mining.
- We need the language to do this.

E.g., mining analytical parameters as a function of synthesis parameters across experiments.

E.g., mining synthesis parameters across experiments.

E.g., mining analytical parameters across experiments.
**Ontology Terms Provide Entry-level Standardization**

- Implementable for systems not semantically enabled.
- Provides standardized data labels
- Still provides link to enduring definitions and machine-readable semantics

<table>
<thead>
<tr>
<th>System 1</th>
<th>System 2</th>
<th>System 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lot</strong></td>
<td><strong>Batch</strong></td>
<td><strong>Lot ID</strong></td>
</tr>
<tr>
<td>001J013</td>
<td>001J013</td>
<td>001J013</td>
</tr>
<tr>
<td>001J014</td>
<td>001J014</td>
<td>001J014</td>
</tr>
</tbody>
</table>

**Preferred label** "batch identifier"

**Definition** "The batch identifier is measurement metadata that identifies the batch where a sample is taken from for being measured. [Allotrope]"

**Semantic Definition (machine readable)**

Subclass of identifier and 'measurement metadata' and (identifies some entity ('has role' some 'batch role'))

**Public reference:**

[http://purl.allotrope.org/ontologies/result#AFR_0001120](http://purl.allotrope.org/ontologies/result#AFR_0001120)
Process Chemistry Lifecycle

- **Route Scouting**
  - Literature & prev. int. work
  - Reaction screening

- **Route Selection**
  - Polymorph / solvate screening
  - Mechanistic & kinetic determinations
  - Fate and Purge
  - DoE

- **Process Optimization**
  - Statistical Lot Analysis
  - New supplier evaluations
  - Adjustment of process parameters within design space

- **Product Filing**
  - CQA, CPP finalized

- **Process Maintenance**
<table>
<thead>
<tr>
<th>Domain</th>
<th>Ontology</th>
<th>Application to process chemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top level ontology</td>
<td>BFO (Basic Formal Ontology)</td>
<td>Aligns other ontologies together</td>
</tr>
<tr>
<td>Relations</td>
<td>Relation Ontology (RO)</td>
<td>Provides standard commonly used relations</td>
</tr>
<tr>
<td>Information</td>
<td>Information Artifact Ontology (IAO)</td>
<td>Provides general information-related terms</td>
</tr>
<tr>
<td>Lab measurements</td>
<td>OBI (Ontology for Biomedical Investigations)</td>
<td>Describe laboratory values related to patient diagnostics</td>
</tr>
<tr>
<td>Chemical compounds</td>
<td>ChEBI (Chemical Entities of Biological Interest)</td>
<td>Describe metabolites and other chemical entities</td>
</tr>
<tr>
<td>Chemical methods</td>
<td>CHMO</td>
<td><a href="https://www.ebi.ac.uk/ols/ontologies/chmo">https://www.ebi.ac.uk/ols/ontologies/chmo</a></td>
</tr>
<tr>
<td>Organic reactions</td>
<td>RXNO</td>
<td>Name reactions in chemical processes</td>
</tr>
<tr>
<td>Units</td>
<td>UO</td>
<td>Units used in chemical processes</td>
</tr>
<tr>
<td>Drugs</td>
<td>DrON (Drug Ontology)</td>
<td>Describe patient medications</td>
</tr>
<tr>
<td>Proteins (e.g., enzymes)</td>
<td>PRO (Protein Ontology)</td>
<td>Describe protein-related entities and the relations between these entities</td>
</tr>
</tbody>
</table>

No full BFO ontology for the domain of process chemistry exists.
Survey of Vocabularies in Process Chemistry

- Two Process Chemistry Vocabularies
  - **Pistoia** (formally Elsevier) Unified Data Model (**UDM**) - Storage and exchange of experimental information about compound synthesis and testing.

- Several ontology publications in process chemistry
  - Overall, good introduction and suggestions, but no solid development.

- Allotrope Foundation Ontologies (**AFO**):
  - AFO covers some results pertinent to process Chemistry ([http://docs.allotrope.org/](http://docs.allotrope.org/)).
  - AFO/AFM covers derivatization which parallels organic synthesis.

*Our thorough survey found no complete ontology(ies) that focus on the domain of process chemistry, a major branch of pharmaceutical chemistry.*
Proposal of OPC: “Ontology of Process Chemistry”

- A new ontology, OPC with the focus on the domain of process chemistry.
- OPC covers the entire timeline from route scouting to reaction optimization to process maintenance.
- Example components of the proposed OPC:
  - Reaction kinetics and mechanism (incl stoichiometry)
  - Polymorphism, solvates
  - Additional material roles (surfactants, flocculants, etc.)
  - Key reaction types: additions, eliminations etc. (coverage by REAXNO)
  - Unit operations such as filtering, refluxing etc.
  - Filings
OPC development strategy and methodology

- eXtensible ontology development (XOD):
  - Term reuse, alignment, design pattern, and community extensibility
  - OBO Foundry ontology development principles
  - Allotrope Foundation Ontology Style Guide
- Up-down design:
  - OPC aligns with the Basic Formal Ontology (BFO).
- Bottom-up design:
  - Common operations, basic reaction pattern (implied by ADM)
  - Use cases based on workflows (often required by regulatory agencies).
Top-Down: Define Upper level of OPC

- Align with BFO
OPC design pattern for a chemical process

BFO compliant, consistent with AFO

- planned process
- material entity
- quality

starting material for synthesis

intermediate material for synthesis

chemical impurity

chemical product

material entity (BFO)

quality (BFO)

critical quality of chemical product

process profile (BFO)

synthetic chemical process

commercial chemical product

GMP starting material

commercial synthetic process

has specified input

is input/output of

has output

has profile

is a

has quality

has output
OPC unit operation processes detail syntheses

- Leveraging existing vocabularies (ontologies when available)

- More unit operations: crystallization, salt formation, pH adjustment, etc.
Vetting Workflows - Bottom-Up General Process

- Engage SME’s: identify key published material to avoid proprietary issues.
- Parse terms from articles (2-4).

- Survey existing terms (OBO Foundry, AFO, IUPAC) / identify gaps.

- Build “scaffold” CMAP to facilitate ontology development (OWL definitions)
Use Case 1: DoE Studies

“Quality by Design (QbD): A systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management.”

DoE Study Implementation

- Objective Definition
- Factor and Range Definition
- Response Definition
- Experimental Data Collection
- Data Analysis
- Confirmatory Reactions

Factorial Design

Multi-dimensional approach to understanding key factors / variables and their interdependence

Response Curves

aggregate experimental meta data with analytical (instrument) results to enable decision making (design space).

Variable interactions can be modeled by graphs

\[ A_i^1 + A_i^{12}j + A_{i(k,l,k2)}^1 \]
Example terms from Use Case 1: DoE Studies
Use Case 2: Fate and Purge Studies

“It is important to understand the formation, fate (whether the impurity reacts and changes its chemical structure), and purge (whether the impurity is removed by, for example, crystallization, extraction), as well as their relationship to the resulting impurities that end up in the drug substance as CQAs.”

Impurity Purging by Simple Crystallization Process

Analytical analyses of process streams determines purge factors of individual steps

Aggregation of experimental data tells the whole story of the ultimate fate of impurities and overall purge factors

e.g. Allotrope chromatography and SQD models

Unreacted reagents, reaction by-products and degradation of impurities complicate fate analyses

Complicated Synthetic Processes push the capabilities of relational models (Graphs are Scalable)
Example terms from Use Case 2: Fate and Purge Studies

Purge Factor
Overall Purge Factor
Carryover
Cumulative Carryover
Modeling of a reaction that has two input materials A and B. The reaction rate kinetics formula is then:

\[ \text{Reaction Rate} = k [A]^x[B]^y \]

The values including A molar concentration [A], B molar concentration [B], and molar concentration of the output product will be measured.
Summary & Discussion

- Summary:
  - A thorough survey on process chemistry ontology
  - A new OPC proposed
    - OPC development strategy, method, top-level design, and design patterns
    - OPC maintains its interoperability with Allotrope Foundation ontologies
  - OPC development just began

- Discussion:
  - OPC could be developed as a standalone ontology and/or become part of Allotrope Foundation Ontologies.
    - Incorporation into AFO subject to governance and acquiring permission sources used in OPC ontology
  - Are others interested in participating?

The University of Michigan acknowledges financial support from Merck for this collaboration.