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

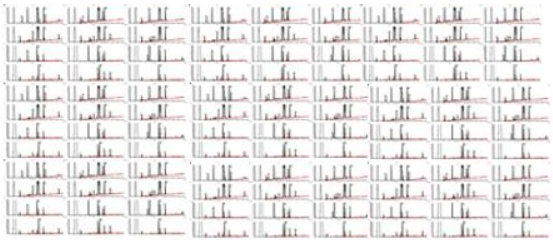
Oliver He, Wes Schafer

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

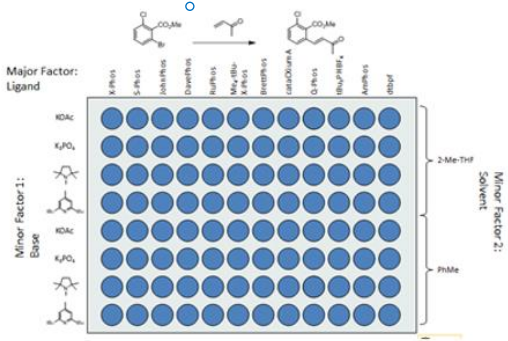
The Importance of Context

e.g., mining analytical parameters across experiments



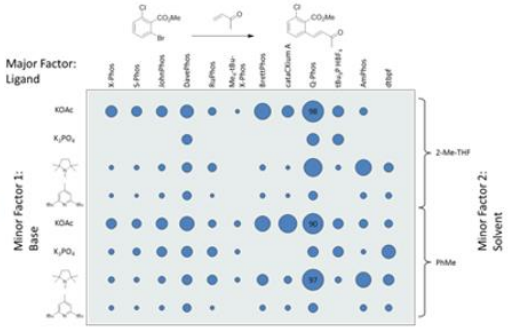
Analytical Data

e.g., mining synthesis parameters across experiments



HTE Synthesis Parameters

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



Complete Data

- 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.

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*

System 1	
Lot	
001J013	
001J014	

System 2	
Batch	
001J013	
001J014	

System 3	
Lot ID	
001J013	
001J014	

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'))

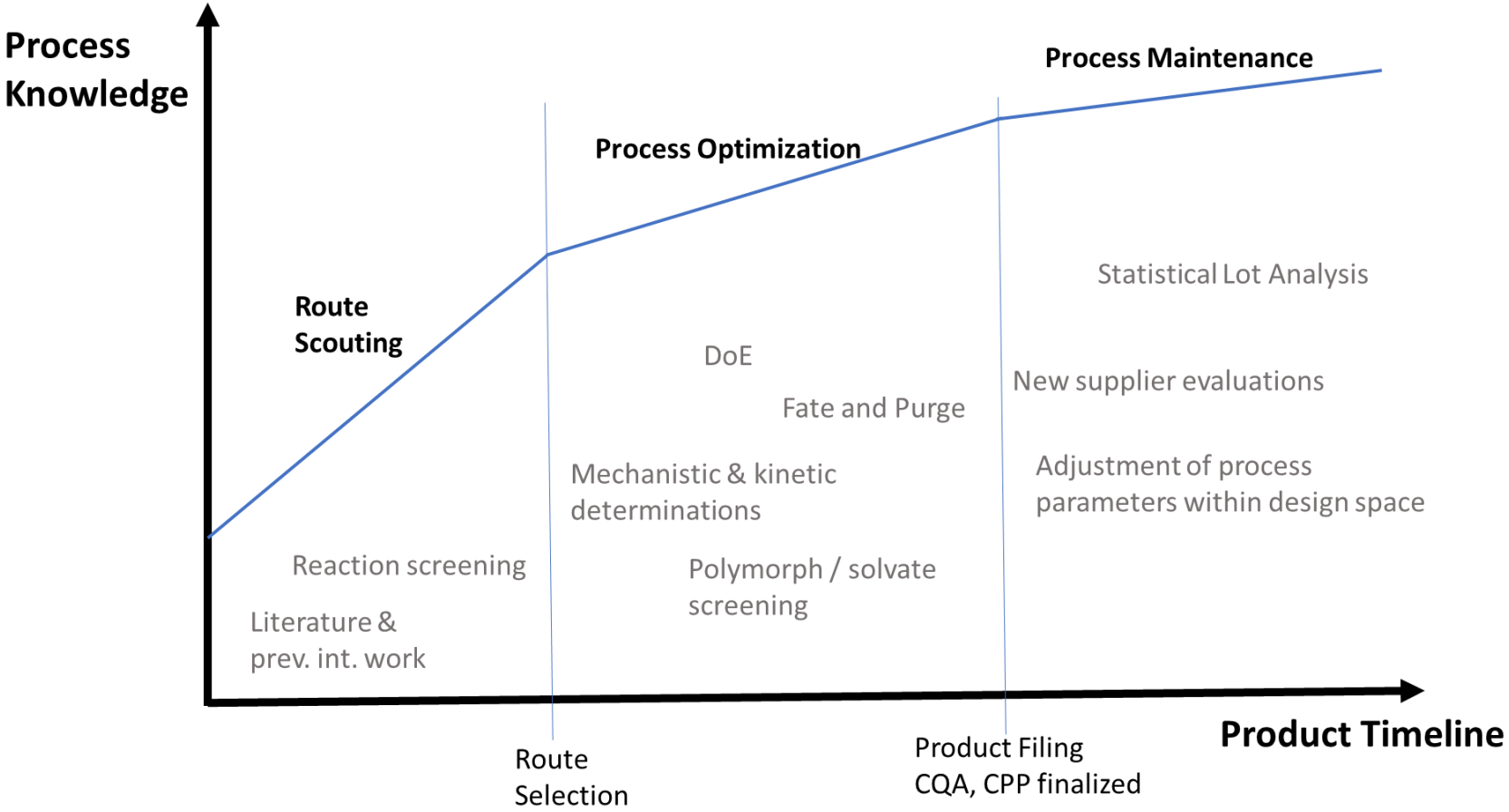
Public reference:
http://purl.allotrope.org/ontologies/result#AFR_0001120

System 1	
batch identifier	
001J013	
001J014	

System 2	
batch identifier	
001J013	
001J014	

System 3	
batch identifier	
001J013	
001J014	

Process Chemistry Lifecycle



Survey of BFO (OBO) Ontologies for Process Chemistry Terms

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

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.
  - **ISA-88**: a standard that addresses batch process control (https://en.wikipedia.org/wiki/ISA-88)
- Several ontology publications in process chemistry
  - E.g., overview of chemical ontologies: https://arxiv.org/abs/2002.03842
  - 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/).
  - 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”

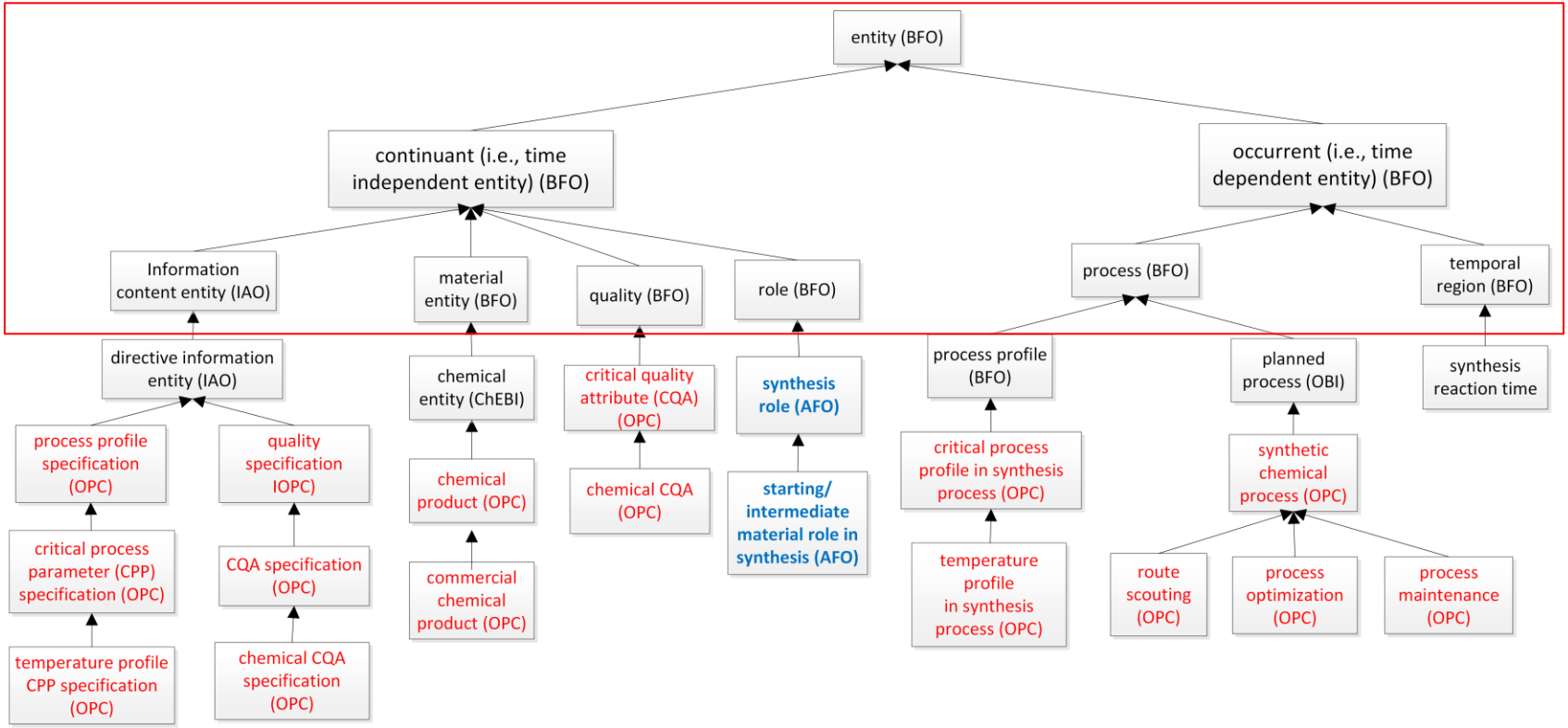
- A new ontology, OPC with the focus on the domain of process chemistry.
- OPC covers the entire timeline from **route scouting** to **reaction optimization** to **process maintenance**.
- Example components of the proposed OPC:
  - Reaction kinetics and mechanism (incl stoichiometry)
  - Polymorphism, 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).

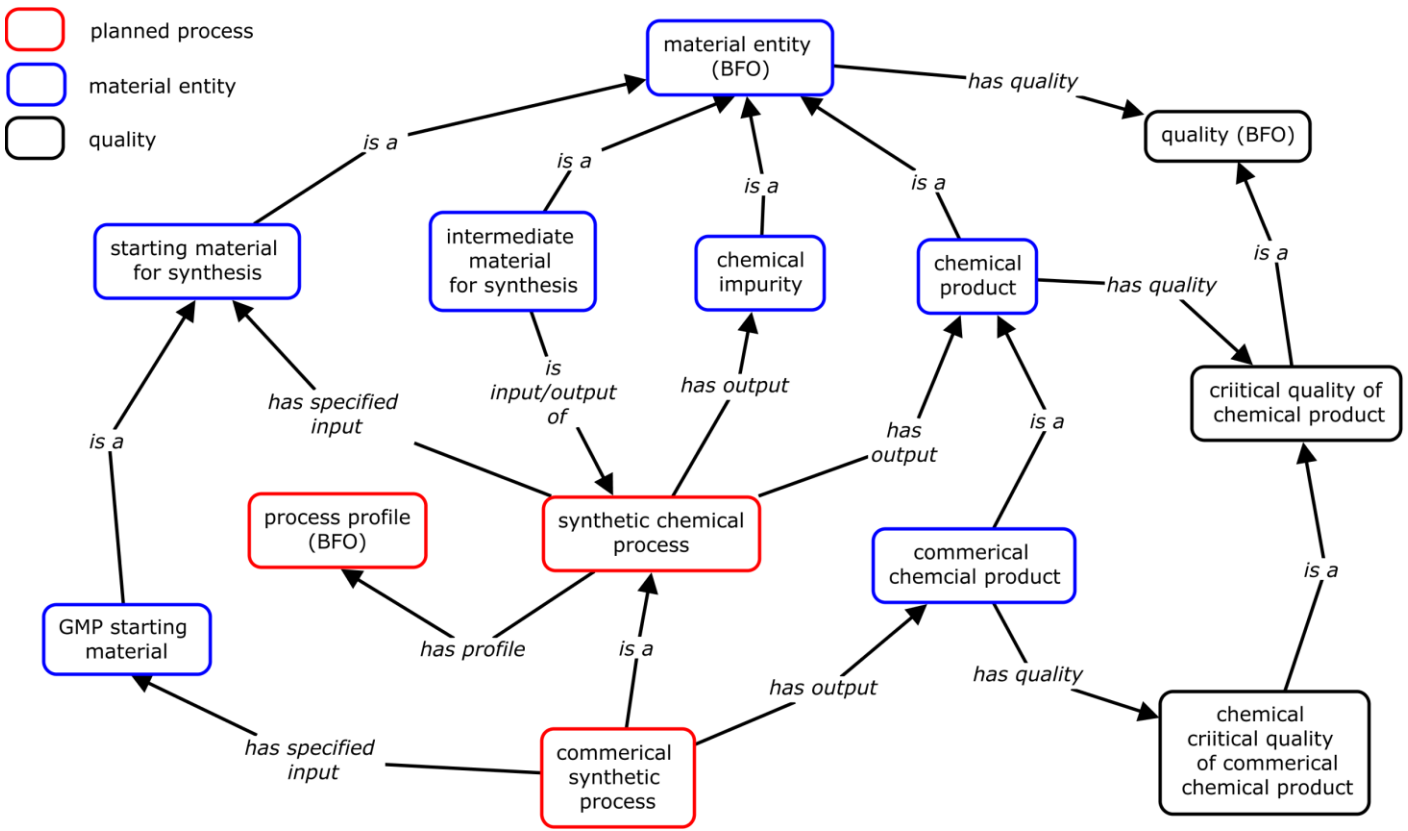
Top-Down: Define Upper level of OPC

- Align with BFO



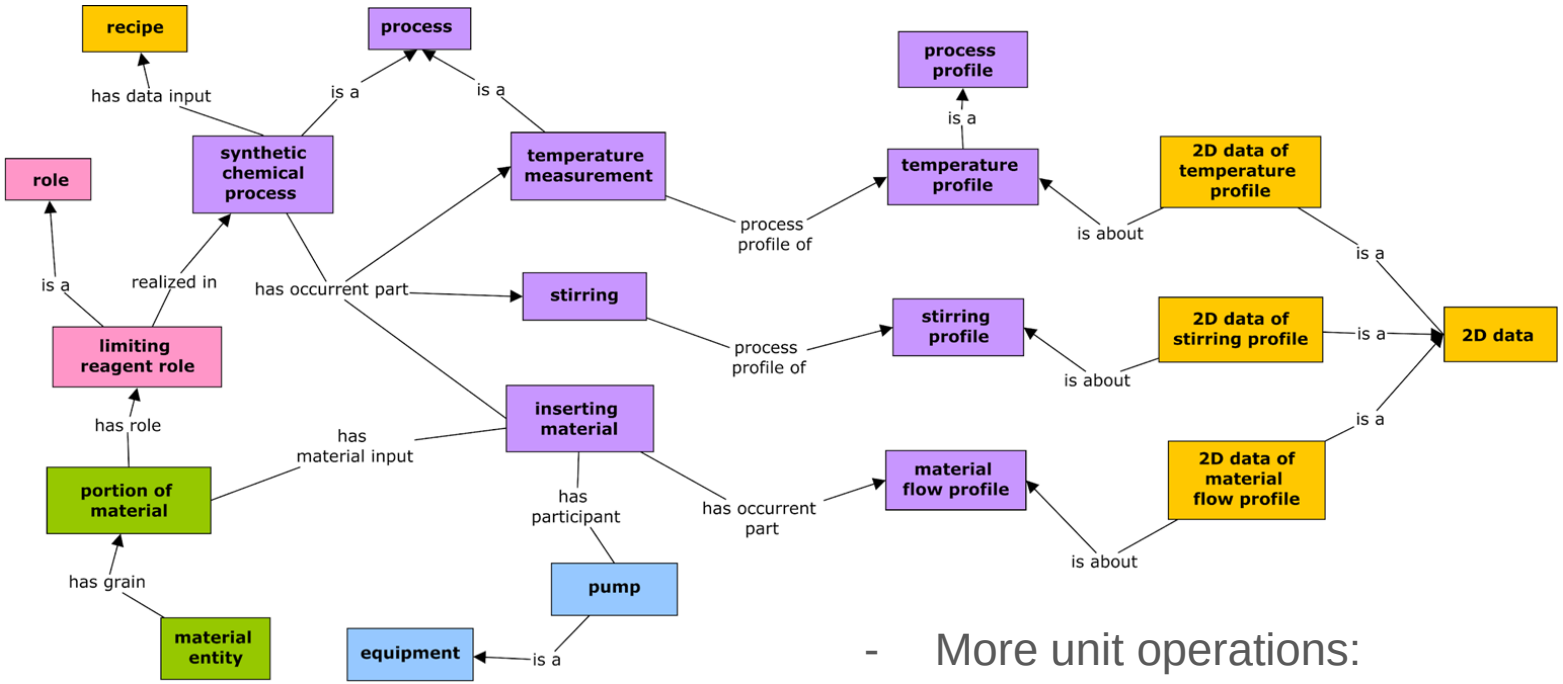
OPC design pattern for a chemical process

BFO compliant, consistent with AFO



OPC unit operation processes detail syntheses

- Leveraging existing vocabularies (ontologies when available)

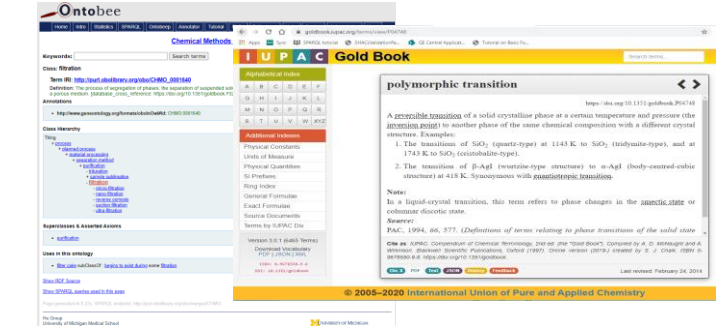


- More unit operations: crystallization, salt formation, pH adjustment, etc.

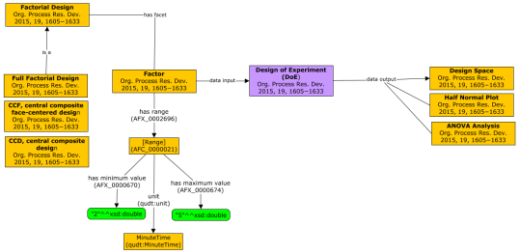
Vetting Workflows - Bottom-Up General Process

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

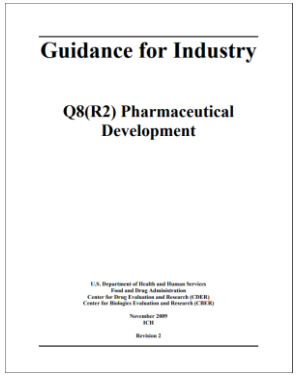
- Survey existing terms (OBO Foundry, AFO, IUPAC) / identify gaps.



- Build "scaffold" CMAP to facilitate ontology development (OWL definitions)



Use Case 1: DoE Studies


Guidance for Industry

Q8(R2) Pharmaceutical Development

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

November 2009
ICH

Revision 2

"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."

e.g. Allotrope PAT and chromatography data models

Factorial Design

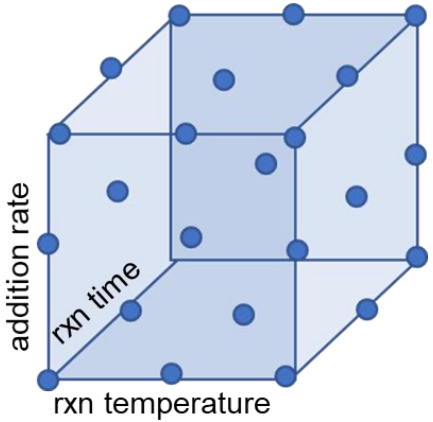
Multi-dimensional approach to understanding key factors / variables and their interdependence

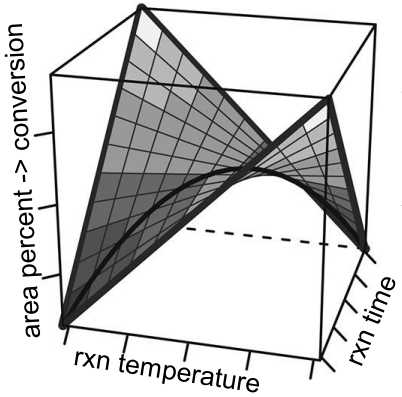
Response Curves

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

DoE Study Implementation

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

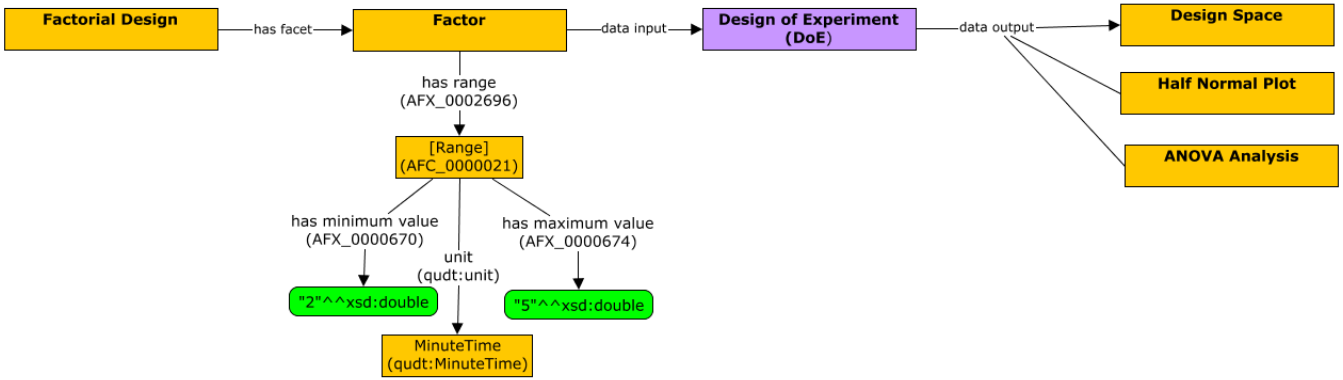


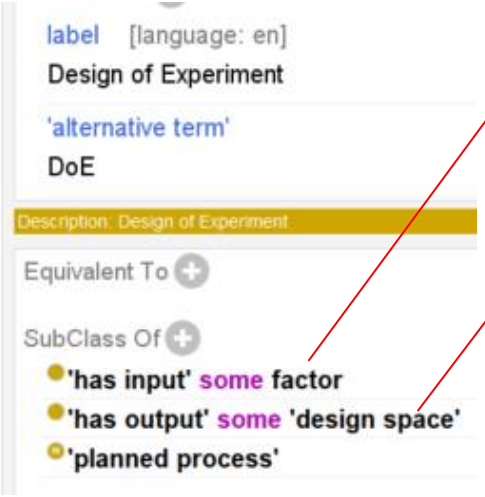


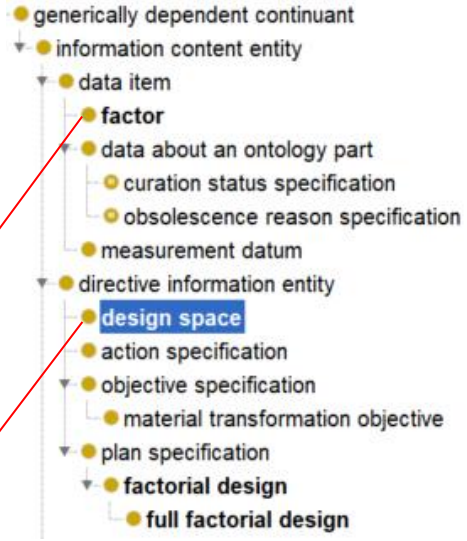
Variable interactions can be modeled by graphs

$A_i i + A_{j1j2} j + A_{f(k1,k2)} k$

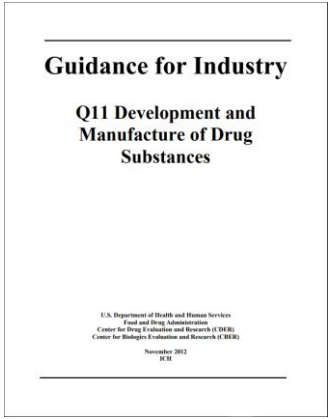
Example terms from Use Case 1: DoE Studies







Use Case 2: Fate and Purge Studies


Guidance for Industry

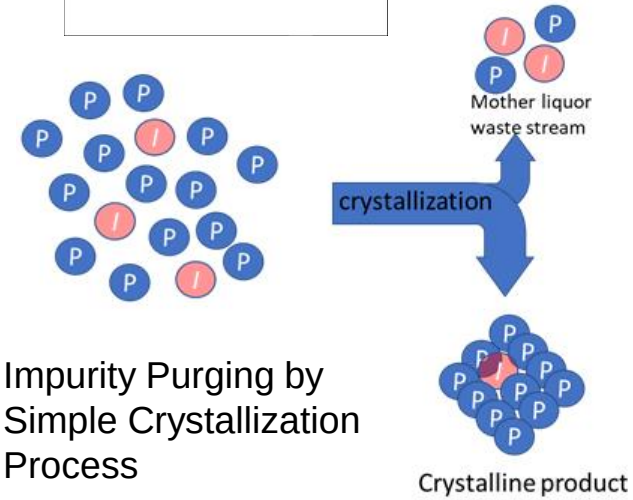
Q11 Development and Manufacture of Drug Substances

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

November 2012
ICH

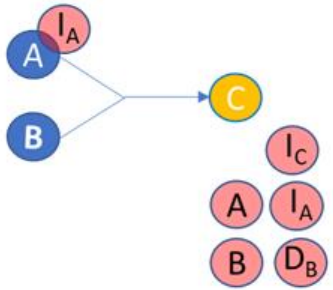
"It is important to understand the formation, fate (whether the impurity reacts and changes its chemical structure), and purge (whether the impurity is removed by, for example, crystallization, extraction), as well as their relationship to the resulting impurities that end up in the drug substance as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

e.g. Allotrope chromatography and SQD models



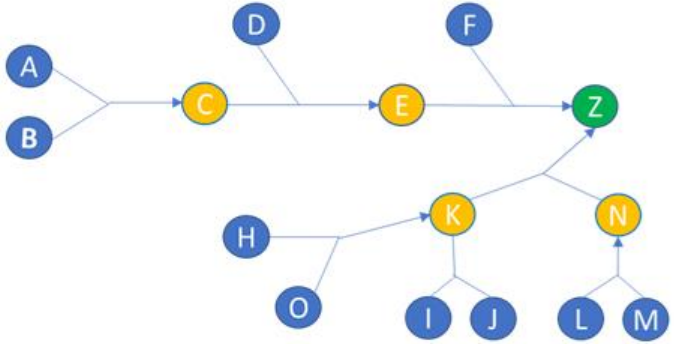
Impurity Purging by Simple Crystallization Process

Analytical analyses of process streams determines purge factors of individual steps



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

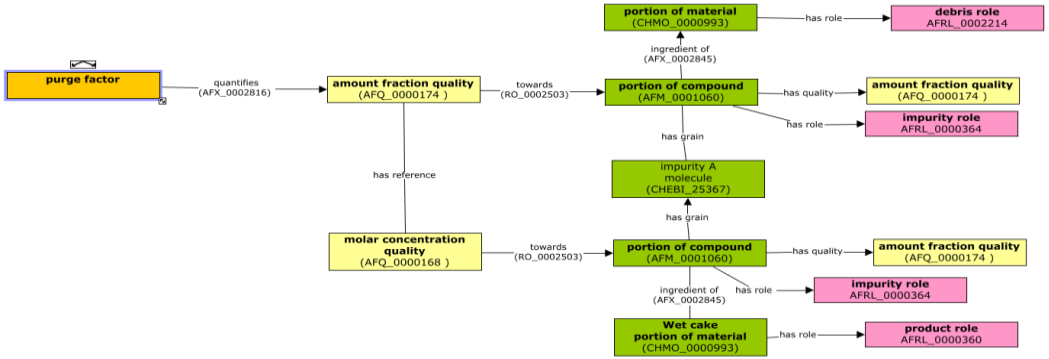
Aggregation of experimental data tells the whole story of the ultimate fate of impurities and overall purge factors

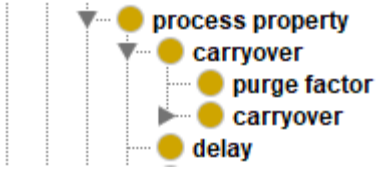


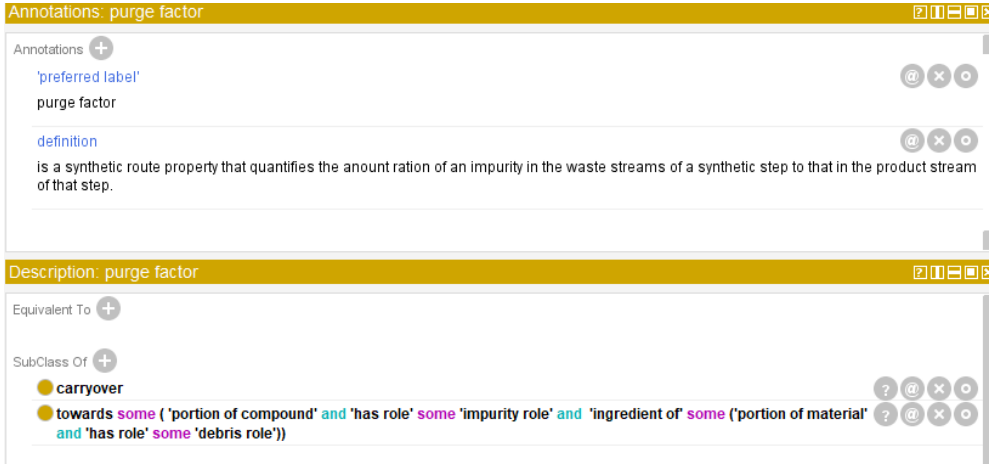
Complicated Synthetic Processes push the capabilities of relational models (Graphs are Scalable)

Example terms from Use Case 2: Fate and Purge Studies

Purge Factor
Overall Purge Factor
Carryover
Cumulative Carryover





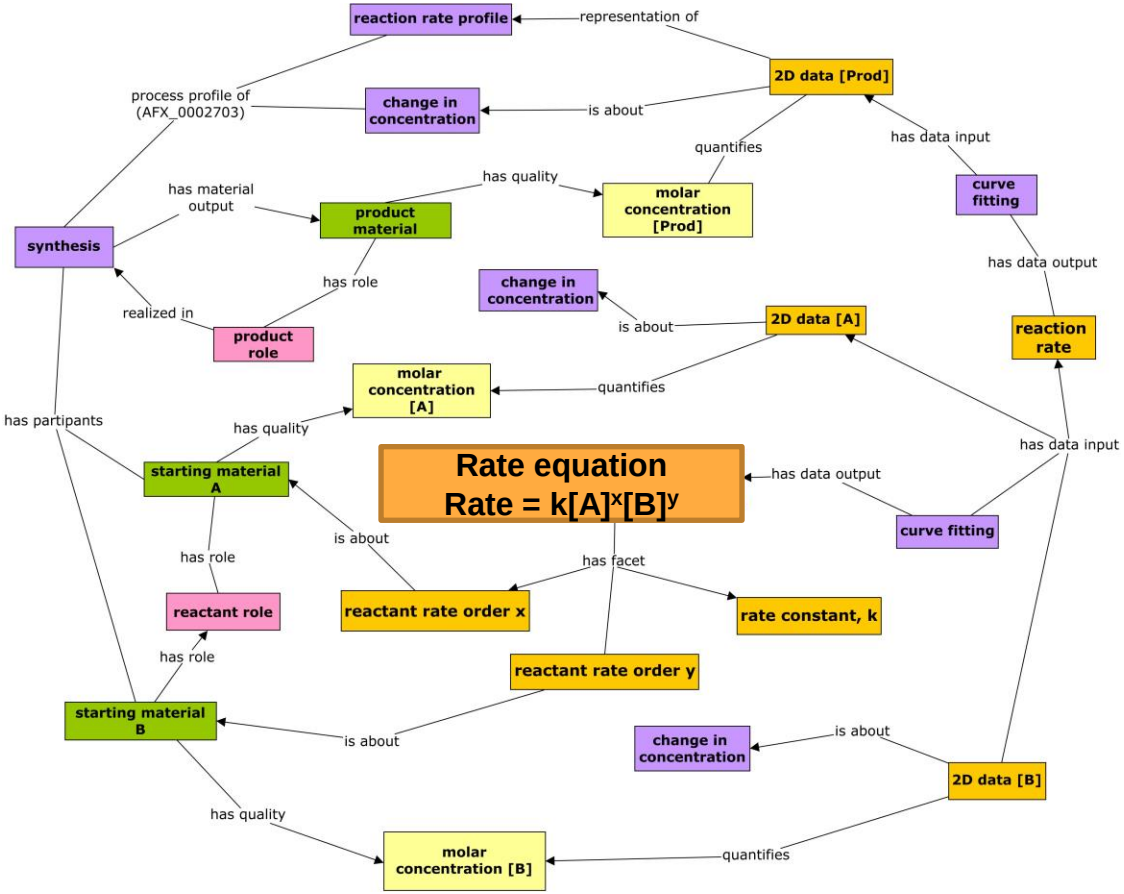


“Pressure Test”: Does Ontology Support Reaction Kinetics?

- Modeling of a reaction that has two input materials A and B. The reaction rate kinetics formula is then:

  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.