Towards a Pattern-Based Ontology for Chemical Laboratory Procedures

Cogan Shimizu¹, Leah McEwen², and Quinn Hirt¹

Data Semantics Laboratory, Wright State University, Dayton, OH, USA
Cornell University, Ithaca, NY, USA

Abstract. There is an increasing expectation in the academic sector for chemistry researchers to conduct risk assessment during experimental planning. However, information concerning laboratory scale chemical reactivity hazards can be difficult to parse despite ongoing efforts to compile from reported incidents. Laboratory procedures do not always directly flag possible incompatibilities among constituents or other process factors. In this paper, we present a pattern-based ontology for capturing multiple factors involved in laboratory procedures, including chemical properties, states, conditions, actions, and associated hazard classifications.

1 Motivation

Developing chemical safety risk assessment tools useful for the academic sector will necessitate tapping into digitally curated data in ways that are relevant to the decision-making processes of research chemists, safety professionals, institutional administration, and other stakeholders. For example, a researcher might be looking at two known chemicals in a proposed reaction scheme and want to know of any conditions that might trigger an adverse outcome, if there are any known procedures for minimizing the likelihood of these conditions, and how to mitigate potential harm if something untoward did occur. The relevant data and information may come from a diverse set of sources covering physical properties,³ synthesis protocols,⁴ and previously reviewed incidents,⁵ among other information.

Some of the most relevant information for analyzing risk appears in reports of incidents where safe control was exceeded, and the influence of reactivity and process factors can be considered in retrospect. However, such reports are not the focus of normal research practice and tend to be exceedingly brief mentions found sporadically in letters to editors of journals, or as news items, or occasionally rephrased as caution statements in vetted procedures. Some of these

³ https://pubchem.ncbi.nlm.nih.gov

⁴ https://www.orgsyn.org

⁵ https://www.csb.gov/investigations

⁶ http://pubs.acs.org/cen/safety

⁷ https://dchas.org/the-dchas-1-list

⁸ http://cenblog.org/the-safety-zone/2016/02/oprds-safety-notables-fromthe-literature

reports have been collected into reference sources such as Bretherick's Handbook of Reactive Chemical Hazards, and the Pistoia Chemical Safety Library. Much of this content has been further compiled into an API-processable data stream within the PubChem database, dynamically presented in the Laboratory Chemical Safety Summaries format (LCSS)¹⁰ described by the US National Research Council (NRC) [3]. However, the meaning remains "locked" in unstructured text and not easily parsed for incorporation into digital information workflows.

The ability to make this information discoverable at the time of need will depend in part on more systematic description of these hazard scenarios. There are many factors at play in conducting a laboratory procedure that may contribute to the potential risk of a given situation. There is a body of research dedicated to analyzing the operations and conditions of large scale chemical processes in industrial settings, where these processes are well-defined and carefully specified as part of the planning process [11]. 11 However, such analyses are rarely conducted for chemical procedures developed iteratively at the laboratory level as defined by OSHA regulations in the United States. Analyzing procedures and coupling these with incident data can potentially bring to light incompatible combinations and problematic operations, as well as aid in planning for adjustments to experimental parameters. Domain terminology that describes key factors can enable the systematic analysis of relationships, such as combinations of chemicals, or substances under different conditions. Such approaches have been used for single analysis of M/SDS documents, 12 and chemical procedures. 13 Developing ontology patterns for chemical processes can more systematically represent potential intersections with hazardous situations [10].

Chemical information is predominantly organized by chemical entity, which is a limited perspective for discerning relationships among multiple process factors. The safety literature is no exception, focusing on hazard-related properties of individual chemicals or substances without reference to specific experimental context or to the surrounding laboratory conditions. Scale, concentration, temperature, pressure, flow rate, and many other chemical, process, operator, and environmental factors have the potential to trigger "runaway" hazardous situations. A more complete risk assessment process, as described by the RAMP model, involves a holistic, laboratory level approach to managing risks beyond hazard identification [13]. Complementing the "object-based" index of specific chemical entities with "process-based" modeling could help surface information and data buried in the published literature on how these chemicals are being used under various conditions and combinations, and the potential for subsequent unintentional interactions to arise [9].

⁹ http://www.pistoiaalliance.org/projects/chemical-safety-library

¹⁰ https://pubchem.ncbi.nlm.nih.gov/lcss

www.acs.org/hazardassessment

¹² www.ilpi.com/msds/ref/demystify

¹³ http://chemicaltagger.ch.cam.ac.uk

https://dchas.org/2017/04/05/information-flow-in-environmental-healthsafety

As such, we have begun the construction of a pattern ecosystem for capturing these chemical interactions and laboratory procedures. The foundational pattern is a chemical process pattern, which has been adapted from the State Transition pattern, which, in turn, is a generalization of the Semantic Trajectory pattern [7]. With the pattern, we hope to answer the following competency questions.

- 1. What substances appear in a particular action, together?
- 2. What substances are ever in the same container?
- 3. What temperatures or pressures are associated with these substances (conditions and/or changes)?
- 4. What apparatus or equipment is involved and associated with which substances (eg. glassware, stir-bars, glove-box)
- 5. What substances are co-located after some particular action?

2 Chemical Process Pattern

In this section, we detail the Chemical Process Pattern. A graphical overview of the pattern can be seen in Figure 1.

2.1 State Transition Pattern

The State Transition Pattern is a novel adaptation or *modularization* [5] of the Semantic Trajectory Pattern [7]. We provide a graphical representation of the pattern in Figure 1a.

The State Transition Pattern is a generalization of the Semantic Trajectory Pattern. The Semantic Trajectory deals with some Thing that moves through time and space which are captured as Fixes. In the State Transition Pattern, we have abstracted time and location to be Conditions of some State.

However, for our use case, we must further modularize the State Transition Pattern. At this time, the alignment is a set of subclass relations between the patterns, as follows.

ChemicalSystem $\sqsubseteq \top$ ChemicalActivity \sqsubseteq StateTransition ChemicalProcess \sqsubseteq Process

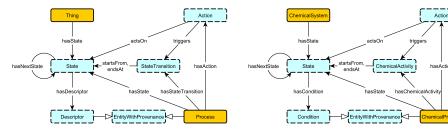
Graphically, we see the results of these equivalences in Figure 1b.

2.2 Patterns Overview

Scoped Domain and Range. One of the primary goals of modelling with ontology design pattern is to lower the number of required ontological commitments required of an ontology engineer adopting the ontology. As such, we *scope* or *guard* many of the range and domain restrictions [6].

$$A \sqsubseteq \forall R.B$$
 (1)

$$\exists R.B \sqsubseteq A$$
 (2)



- (a) A graphical representation of the State Transition Ontology Design Pattern.
- (b) We modularize the State Transition ODP to construct the Chemical Process pattern.

Fig. 1: These two figures illustrate the modularization of the State Transition Pattern to Chemical Process Pattern.

Axiom (1) is a scoped range restriction. This allows us to say "when we relate A to something via R, that something must be a B." Axiom (2) follows the same for scoped domain restriction.

Structural Tautologies. These axioms are intended for human consumption; they do not add anything to the ontology. Essentially, these axioms, taking the below form, simply inform the reader of the intended use of a property [6].

$$A \sqsubseteq \ge 0R.B$$

OPLa Annotations. The provided OWL file is annotated with the appropriate OPLa annoations [5]. We note, in particular, the classes marked as opla:ExternalClass: Action, Condition, and State. ChemicalActivity and EntityWithProvenance are defined later in the paper. The annotations were generated with the OPLa plugin for Protégé [12].

Standard Disjointness. In the following sections, all classes which are not in direct or inferred subclass relationship are declared to be mutually disjoint.

2.3 Action

Additionally, we provide graphical representations of the Stir Action and Heat Action subpatterns, as well as an expanded view of the Action Pattern in Figure 2. In the diagram, we use MethodTypes.txt and Apparatus.txt to denote that these values are *individuals* from a controlled vocabulary. An individual appearing the controlled vocabulary is an individual of type MethodType or Apparatus, for

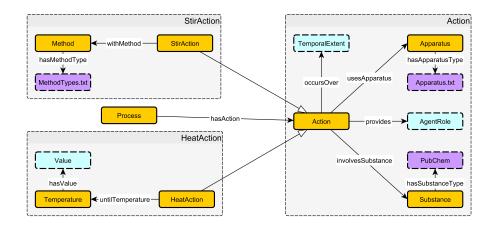


Fig. 2: Graphical overviews of the Action sub-patterns.



Fig. 3: Graphical overview of the Simultaneous Action Pattern.

example. The Simultaneous Action is shown in Figure 3.

$Action \sqsubseteq = 1 triggers. ChemicalActivity$	(1)
$Action \sqsubseteq = 1 actsOn.State$	(2)
$\top \sqsubseteq \forall occursOver.TemporalExtent$	(3)
$Action \sqsubseteq = 1 occursOver. TemporalExtent$	(4)
$ op \sqsubseteq orall$ uses $Apparatus.Apparatus$	(5)
Action $\sqsubseteq \ge 1$ usesApparatus.Apparatus	(6)
$\top \sqsubseteq \forall hasApparatusType. ApparatusType$	(7)
$\forall hasApparatusType. \top \sqsubseteq ApparatusType$	(8)
$Action \sqsubseteq = 1 provides. AgentRole$	(9)
$\top \sqsubseteq \forall involvesSubstance.Substance$	(10)
$Action \sqsubseteq \ge 1 involvesSubstance. Substance$	(11)
$\top \sqsubseteq \forall hasSubstanceType.PubChem$	(12)
$\forall hasSubstanceType. \top \sqsubseteq Substance$	(13)

- 1. An Action triggers exactly one ChemicalActivity. However, we currently leave it to the ontology engineer to specify the exact complexity of a ChemicalActivity.
- 2. An Action acts on exactly one state.
- 3. The range of occursOver is strictly limited to TemporalExtent.
- 4. An Action occurs over exactly one TemporalExtent.
- 5. The range of usesApparatus is strictly limited to Apparatus.
- 6. An Action uses at least one Apparatus.
- 7. The range of hasApparatusType is strictly limited to ApparatusType.
- 8. The domain of hasApparatusType is strictly limited to Apparatus.
- 9. An Action provides exactly one AgentRole.
- 10. The range of involvesSubstance is strictly limited to Substance.
- 11. An Action always involves at least one Substance.
- 12. The range of hasSubstanceType is strictly limited to SubstanceType.
- 13. The domain of hasSubstanceType is strictly limited to Substance.

StirAction

$StirAction \sqsubseteq Action$	(14)
$\top \sqsubseteq \forall withMethod.Method$	(15)
${\sf StirAction} \sqsubseteq = 1 {\sf withMethod}. {\sf Method}$	(16)
$\top \sqsubseteq \forall hasMethodType. MethodType$	(17)
\forall has M ethod T ype. $\top \sqsubseteq M$ ethod	(18)

- 14. All StirActions are Actions.
- 15. The range of with Method is strictly limited to Method.
- 16. A StirAction is completed with exactly one Method.
- 17. The range of hasMethodType is strictly limited to MethodType.
- 18. The domain of hasMethodType is strictly limited to Method.

HeatAction

$HeatAction \sqsubseteq Action$	(19)
$HeatAction \sqsubseteq =1 untilTemperature.Temperature$	(20)
$ op \sqsubseteq orall hasValue.Value$	(21)
Temperature $\sqsubseteq =1$ has $Value$.	(22)

- 19. All HeatActions are Actions.
- 20. A HeatAction has exactly one limiting Temperature.
- 21. The range of has Value is strictly limited to Value.
- 22. A Temperature has exactly one Value.

SimultaneousAction	
$SimultaneousAction \sqsubseteq Action$	(23)
$ op \sqsubseteq orall hasSimultaneousAction.$	Action
_	(24)
$ op \sqsubseteq orall hasSimultaneousAction.$	$\neg SimultaneousAction\ (25)$
\forall hasSimultaneousAction. $\top \sqsubseteq$ SimultaneousAction	(26)
$has Simultaneous Action \circ occurs Over \sqsubseteq occurs Over$	(27)
$has Simultaneous Action \circ involves Substance \sqsubseteq involves Substance$	(28)
 All SimultaneousActions are Actions The range of hasSimultaneousAction is strictly limited to Action. A SimultaneousAction may not have another SimultaneousAction taneous action. The domain of hasSimultaneousAction is strictly limited to Simultion. The Actions that co-occur must, in fact, occur simultaneously. Any Substance that is involved in a "subaction" is involved in t neousAction. 	ltaneousAc-
2.4 Chemical Activity	
${\sf ChemicalActivity} \sqsubseteq = 1 {\sf startsFrom.State}$	(1)
$ChemicalActivity \sqsubseteq = 1 endsAt.State$	(2)
$ op \sqsubseteq \forall startsFrom.State$	(3)
$ op \sqsubseteq orall endsAt.State$	(4)
	(5)
 A ChemicalActivity always begins in some State and results in so supra. The range of startsFrom is strictly limited to States. The range of endsAt is strictly limited to States. 2.5 ChemicalProcess	me State.
$ op \sqsubseteq orall hasAction.Action$	(1)
$\top \sqsubseteq \forall hasChemicalActivity.ChemicalActivity$	(2)

(3)

(4)

(5)

 ${\sf ChemicalProcess} \sqsubseteq {\geq} 1 {\sf hasAction.Action}$

 $\mathsf{ChemicalProcess} \sqsubseteq \geq 1 \mathsf{hasState}. \mathsf{State}$

 ${\sf ChemicalProcess} \sqsubseteq {\geq} 1 {\sf hasChemicalActivity}. {\sf ChemicalActivity}$

- 1. The range of hasAction is strictly limited to Activity.
- 2. The range of hasChemicalActivity is strictly limited to ChemicalActivity.
- 3. A ChemicalProcess must have at least one Action.
- 4. A ChemicalProcess must have at least one ChemialActivity.
- 5. A ChemicalProcess must have at least one State.

2.6 ChemicalSystem

ChemicalSystem
$$\sqsubseteq \ge 1$$
hasState.State (1)

$$\top \sqsubseteq \forall \mathsf{hasState}.\mathsf{State}$$
 (2)

$$\mathsf{State} \sqsubseteq \le 1 \mathsf{hasState}^-. \top \tag{3}$$

- 1. A ChemicalSystem always has at least one State.
- 2. The range of hasState is strictly limited to State.
- 3. Any State is associated with exactly one Thing.

2.7 Condition

Condition
$$\sqsubseteq$$
 EntitywithProvenance (1)

$$\top \sqsubseteq \forall \mathsf{hasCondition}.\mathsf{Condition} \tag{2}$$

(3)

- 1. All Conditions must have provenance. In this use-case this is reasonable as every condition is measured by someone or some device.
- 2. The range of hasCondition is strictly limited to Conditions.

2.8 EntityWithProvenance

The EntityWithProvenance Pattern is extracted from the PROV-O ontology. At the pattern level, we do not want to make the ontological committment to a full-blown ontology. It suffices to align a sub-pattern to the core of PROV-O. Further discussion on the EntityWithProvenance pattern, as well as its specification (as below) in an OWL file may be found on the online portal. ¹⁵

 $^{^{15}\ \}rm https://ontologydesignpatterns.org/wiki/Submissions:$ EntityWithProvenance

$EntityWithProvenance \sqsubseteq \forall wasDerivedFrom.EntityWithProvenance$	(1)
$\forall attributedTo.Agent \sqsubseteq EntityWithProvenance$	(2)
${\sf EntityWithProvenance} \sqsubseteq \forall {\sf attributedTo}. {\sf Agent}$	(3)
$\forall generatedBy. Provenance Activity \sqsubseteq EntityWithProvenance$	(4)
${\sf EntityWithProvenance} \sqsubseteq \forall {\sf generatedBy.ProvenanceActivity}$	(5)
$\forall used. EntityWithProvenance \sqsubseteq ProvenanceActivity$	(6)
${\sf ProvenanceActivity} \sqsubseteq \forall {\sf used.EntityWithProvenance}$	(7)
$\forall performedBy.Agent \sqsubseteq ProvenanceActivity$	(8)
${\sf ProvenanceActivity} \sqsubseteq \forall {\sf performedBy.Agent}$	(9)

- 1. The scoped range of wasDerivedFrom, scoped by EntityWithProvenance, is EntityWithProvenance.
- The scoped domain of attributedTo, scoped by Agent, is EntityWithProvenance.
- 3. The scoped range of attributedTo, scoped by EntityWithProvenance, is Agent.
- 4. The scoped domain of generated By, scoped by Provenance Activity, is Entity With Provenance.
- The scoped range of generatedBy, scoped by EntityWithProvenance, is ProvenanceActivity.
- 6. The scoped domain of used, scoped by EntityWithProvenance, is ProvenanceActivity
- 7. The scoped range of used, scoped by ProvenananceActivity, is EntityWith-Provenance.
- 8. The scoped domain of performed By, scoped by Agent, is Provenance Activity.
- 9. The scoped range of performedBy, scoped by ProvenanceActivity, is Agent.

2.9 State

$$\top \sqsubseteq \forall \mathsf{hasNextState}.\mathsf{State}$$
 (1)

$$\mathsf{State} \sqsubseteq \leq 1 \mathsf{hasNextState}.\mathsf{State} \tag{2}$$

- 1. The range of hasNextState is strictly limited to State.
- 2. A State will always follow at most one State.

3 Worked Example

The following incident report is extracted from [4,1]. Formatting and language have been modified in order to make it clear exactly how the information was obtained. In the interest of brevity, we have used a simple incident report. However, even such a simple application of the pattern requires a high level of detail from the report. Thus, in our worked example, we aim to provide an illustration of the foundational concepts of our ontological ecosystem and note

certain aspects will be addressed in future work. In the following, we use the cpp: namespace as an abbreviation for "Chemical Process Pattern" in the URI https://daselab.org/chemicalprocesspattern/.

```
The Incident Report.
```

5-ethyl-2-methyl-pyridine and 70% nitric acid were placed in a small auto-clave.

They were heated and stirred for 40 minutes.

The emergency vent was opened due to a sudden pressure rise. A violent explosion occurred 90 seconds later.

From the first statement, we extract the following triples regarding the substances and apparatus. The placement of the chemicals will also constitute an Action subclass, as it is developed.

From the next sentence we extract the StirAction and HeatAction. In order to capture their simultaneity, we use the SimultaneousAction.

```
cpp:te1 rdf:type cpp:TemporalExtent .
cpp:sa1 rdf:type cpp:StirAction .
cpp:ha1 rdf:type cpp:HeatAction .
cpp:sim1 rdf:type cpp:SimultaneousAction cpp:hasSimultaneousAction cpp:sa1 cpp:hasSimultaneousAction cpp:hasCimultaneousAction cpp:hasCimulta
```

From the next sentence, we extract the apparatus and resulting state of the action. The Condition is provided an asText property for illustrative purposes.

```
cpp:ap2
          rdf:type
                                 cpp:Apparatus
          cpp:hasApparatusType
                                 "fume hood"
          rdf:type
                                 cpp:Condition
cpp:c1
          ewp:isAttributedTo
                                 cpp:ap2
          cpp:asText
                                 "high pressure"
cpp:s2
          rdf:type
                                 cpp:State
cpp:s1
          rdf:type
                                 cpp:State
          cpp:hasNextState
                                 cpp:s2
```

In the last step, we note that a hazardous state has been entered. However, the development of this part of the ontological ecosystem is still planned in future work. We note possible integration the Modified Hazardous Material Pattern [2] to help model this aspect. Finally, we may wrap it all together into the Chemical Process.

4 Conclusions

In this paper, we have described a foundational pattern to building a ontology design pattern ecosystem for modelling chemical processes. The core pattern is based on the State Transition Pattern, which in turn, is adapted from the Semantic Trajectory Pattern. The intent of this pattern and the surrounding ecosystem is to provide chemists—and their students— with a resource for analyzing experiments and potentially finding unforeseen interactions that can result in hazardous states, events, or situations.

A sufficiently populated ontology of chemical processes can also be used as background knowledge for training a more sophisticated learning model or could be used to explain the decisions *made* by such a system (deep learning models and explainable AI, respectively).

In the future, we expect to integrate more closely with the large chemistry based datasets, such as PubChem and M/SDS. In addition, there are existing patterns that may be integrated to enhance the functionality of the core pattern and complete other pieces, such as QUDT¹⁶ for measurements and units, the ModifiedHazardous Material Pattern [2] for modelling hazardous states, and the Material Transformation [8] for extending ChemicalActivity.

Acknowledgement. Cogan Shimizu acknowledges support by the Dayton Area Graduate Studies Institute (DAGSI).

¹⁶ https://qudt.org/

References

- Nitric acid. National Center for Biotechnology Information. PubChem Compound Database; CID=944, datasheet=lcss. Accessed May 30th, 2018.
- 2. M. Cheatham, H. Ferguson, C. Vardeman, and C. Shimizu. A modification to the hazardous situation ODP to support risk assessment and mitigation. In K. Hammar et al., editors, *Advances in Ontology Design and Patterns*, volume 32 of *Studies on the Semantic Web*, pages 97–104. IOS Press, 2017.
- N. R. Council. Prudent Practices in the Laboratory: Handling and Management of Chemical Hazards, Updated Version. The National Academies Press, Washington, DC, 2011.
- 4. R. L. Frank. volume 30:, pages 33-48. 1952.
- P. Hitzler, A. Gangemi, K. Janowicz, A. A. Krisnadhi, and V. Presutti. Towards a simple but useful ontology design pattern representation language. In E. Blomqvist et al., editors, Proceedings of the 8th Workshop on Ontology Design and Patterns (WOP 2017) Vienna, Austria, October 21, 2017, volume 2043 of CEUR Workshop Proceedings. CEUR-WS.org, 2017.
- 6. P. Hitzler and A. Krisnadhi. On the roles of logical axiomatizations for ontologies. In P. Hitzler, A. Gangemi, K. Janowicz, A. Krisnadhi, and V. Presutti, editors, Ontology Engineering with Ontology Design Patterns - Foundations and Applications, volume 25 of Studies on the Semantic Web, pages 73–80. IOS Press, 2016.
- Y. Hu, K. Janowicz, D. Carral, S. Scheider, W. Kuhn, G. Berg-Cross, P. Hitzler, M. Dean, and D. Kolas. A geo-ontology design pattern for semantic trajectories. In T. Tenbrink, J. G. Stell, A. Galton, and Z. Wood, editors, Spatial Information Theory - 11th International Conference, COSIT 2013, Scarborough, UK, September 2-6, 2013. Proceedings, volume 8116 of Lecture Notes in Computer Science, pages 438-456. Springer, 2013.
- 8. C. F. V. II, A. A. Krisnadhi, M. Cheatham, K. Janowicz, H. Ferguson, P. Hitzler, and A. P. C. Buccellato. An ontology design pattern and its use case for modeling material transformation. *Semantic Web*, 8(5):719–731, 2017.
- M. Leah. ci, volume 39, chapter Chemical Health and Safety Data Management, page 31. 2018 2017.
- 10. L. McEwen and R. Stuart. Meeting the google expectation for chemical safety information. *Chemistry International*, 37(5-6):12–16, 2015.
- M. B. Mulcahy, C. Boylan, S. Sigmann, and R. Stuart. Using bowtie methodology to support laboratory hazard identification, risk management, and incident analysis. *Journal of Chemical Health and Safety*, 24(3):14 20, 2017.
- 12. C. Shimizu, Q. Hirt, and P. Hitzler. A protégé plugin for annotating OWL ontologies with opla. ESWC 2018, June 2018. To Appear.
- 13. R. B. Stuart and L. R. McEwen. The safety "use case": Co-developing chemical information management and laboratory safety skills. *Journal of Chemical Education*, 93(3):516–526, 2016.