

The PolyLab Ontology - A Semantic Model of Laboratory Processes and Their Plans

Marta Dembska^{1,*†}, Martin Held^{2†} and Sirko Schindler^{1,†}

¹German Aerospace Center (DLR), Institute of Data Science, Mälzerstraße 5, 07745 Jena, Germany

²Helmholtz-Zentrum Hereon, Institute of Membrane Research, Max-Planck-Str. 1, 21502 Geesthacht, Germany

Abstract

Documenting laboratory work is shifting from analogue tools like lab notebooks towards digital means using Electronic Laboratory Notebooks and similar systems. This requires an increasing portfolio of well-defined structures, schemas, vocabularies, and other standards to ensure the interoperability of collected data. Domain ontologies in research areas, where laboratory work is a core component, tend to emphasise the components of the experiments – methods, devices, consumables, and the resulting data – rather than providing detailed descriptions of the experimental procedures themselves. Consequently, the workflow is represented only at a high level, omitting the essential procedural details that are critical for reproducible science.

PolyLab was created to bridge this gap in laboratory-based research documentation. Its deliberately generic design makes it applicable to a broad spectrum of laboratory processes across materials science, chemistry, and related disciplines. Building upon well-established ontologies like PROV-O or BFO, PolyLab provides a vocabulary and set of relations that capture both the execution of a lab workflow and its connection to prior planning. Thereby, it allows to document both prospective and retrospective provenance of laboratory work, opening the door for comparative analyses across experimental workflows. In this role, PolyLab provides essential support for the semantic annotation and documentation of laboratory activities across many domains.

Keywords

Ontology, Laboratory Workflows, Electronic Lab Notebook

1. Introduction

Digitization in laboratory environments is still an ongoing process. With the increasing proliferation of Electronic Laboratory Notebooks (ELNs) and Laboratory Information Management Systems (LIMS), data collection is shifting from paper notebooks and ad-hoc solutions to integrated, “digital-first” solutions. This goes beyond merely describing the results of an experiment and covers the entire necessary workflow. Of particular value for reproducibility is the comparison between how a specific workflow was planned versus how it was actually executed. The as-created wealth of information can be leveraged, e.g., by various AI methods including, but not limited to, Large Language Models (LLMs). However, a major challenge remains in structuring and properly enriching the collected data semantically. Current practice typically dumps workflow provenance descriptions in free text that require human interpretation instead of formalised records. Here, ontologies can provide the necessary framework to provide annotations that contextualise individual entries and connect them to other, related experiments. The PROV Ontology (PROV-O) [1] and its extension, The Ontology for Provenance and Plans (P-Plan) [2], represent essentially the semantic standards in the context of workflow provenance due to their general usability.

In this paper, we introduce PolyLab, an ontology that semantically models laboratory processes and their associated plans across a wide range of experimental domains. Extending earlier work on PolyMat [3], a domain ontology in polymer membrane research, PolyLab expands the representation

SeMatS 2026: The 3rd International Workshop on Semantic Materials Science co-located with the 25th International Semantic Web Conference (ISWC 2026), October 25th, 2026, Bari, Italy

*Corresponding author.

† These authors contributed equally.

✉ marta.dembska@dlr.de (M. Dembska); martin.held@hereon.de (M. Held); sirko.schindler@dlr.de (S. Schindler)

ORCID 0000-0002-8180-1525 (M. Dembska); 0000-0003-1869-463X (M. Held); 0000-0002-0964-4457 (S. Schindler)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

to cover complete laboratory workflows in experimental science. Although its development was motivated by the specific requirements of polymer membrane investigations, PolyLab is deliberately engineered as a domain-agnostic framework capable of describing laboratory processes across a selection of domains, including materials science, chemistry and related disciplines. It connects the planned laboratory processes and their constituent elements with the actual execution of those processes and their corresponding components, thereby integrating specific laboratory activities, the associated work documentation, data management artifacts, performance metrics, and the resulting experimental data. PolyLab is built directly on the provenance model defined by PROV-O, with P-Plan employed as its plan-oriented extension. By adhering to the PROV-O standard and leveraging P-Plan's additional constructs, the model can represent both the prospective provenance of a planned laboratory workflow (the intended process and its components) and the retrospective provenance of its actual execution (the realised activities, data, and outcomes). Leveraging PolyMat's definition of an experiment as a central laboratory activity, PolyLab expands the ontology with additional activity classes (e.g., preparation, measurement, or evaluation) and incorporates further workflow components (such as consumables, devices and their configurations, and other experimental variables), thereby allowing entire laboratory workflows to be assembled from these elements. PolyLab is intentionally designed to enable easy annotation of ELN-recorded experiments, streamlining the FAIR-compliant preparation of experimental workflows. The granularity of the process and plan descriptions is left to the user, allowing each researcher to decide how much detail to formalise and retain. By building on the widely accepted PROV-O provenance model, the ontology standardises provenance records and makes them interoperable with other provenance-compliant datasets. Consequently, PolyLab's generic and adaptable design creates connections between otherwise unrelated research fields, opening the possibility for cross-disciplinary discoveries.

Online resources

Resource type Ontology

Licence CC BY 4.0 International

URL <https://w3id.org/polylab/>

GitLab <https://gitlab.com/dlr-dw/poly-ontologies/polylab-ontology>

DOI 10.5281/zenodo.21709031

2. Related work

Several ontologies¹ address research processes in materials science and adopt a strongly aligned upper-ontology foundation with Basic Formal Ontology (BFO) [4]. NFDIcore Ontology (NFDIco) constitutes a common integration layer and cross-domain interoperability for NFDI resources, representing organizations, projects, persons, services, infrastructures, and metadata [5]. Based on that, MaterialDigital PMD Core Ontology (PMDco) provides a mid-level framework for materials science and engineering, covering process chains, material structures, properties, devices, plans, roles, functions and characterization, following the processing–structure–properties paradigm [6]. The NFDI Matwerk Ontology (MWO) ontology [7] reuses NFDIco's organisational model, PMDco's model of experiments and the model of documentation from the Information Artifact Ontology (IAO) [8, 9], and extends this foundation to

¹Here and in the following examples, we will use label-based identifiers for some ontologies to improve readability. Further, we use the following namespace prefixes: bfo: <<http://purl.obolibrary.org/obo/BFO>>; chebi: <<http://purl.obolibrary.org/obo/CHEBI>>; cob: <<http://purl.obolibrary.org/obo/COB>>; iao: <<http://purl.obolibrary.org/obo/IAO>>; m4i: <<http://w3id.org/nfdi4ing/metadata4ing#>>; mwo: <<http://purl.helmholtz-metadaten.de/mwo#>>; om: <<http://www.ontology-of-units-of-measure.org/resource/om-2/>>; pims-ii: <<http://www.molmod.info/semantics/pims-ii.ttl#>>; plab: <<https://w3id.org/polylab/>>; pmat: <<https://w3id.org/polymat/>>; prov: <<http://www.w3.org/ns/prov#>>; p-plan: <<http://purl.org/net/p-plan#>>; rdfs: <<http://www.w3.org/2000/01/rdf-schema#>>; ro: <<http://purl.obolibrary.org/obo/RO>>

consortium structures, research infrastructures, software, datasets, workflows, educational activities, and community resources.

The PolyLab ontology and the combination of MWO and PMDco reuse BFO and the corresponding Open Biological and Biomedical Ontologies Foundry (OBO) [10] relations such as `bfo:has part`, `ro:has participant`, `ro:has role`, and `ro:has output`, to which MWO adds relations of agents and PolyLab contributes process chain predicates. The BFO alignment also results in matching concepts for `chebi:Chemical entity`, `bfo:Material entity`, `bfo:Role`, documents, process participation, and process outputs. While MWO's subclasses of `bfo:Role` are more general than PolyLab's roles, their classes for documents and software correspond conceptually, though under different parental classes. Devices are modelled complementarily, as MWO focuses on their description while PolyLab highlights their configuration.

However, the focus of several classes differs in scope and granularity. PMDco and MWO aim at broad coverage of materials-science domains, whereas PolyLab prioritises operational laboratory procedures, experiment-specific process chains and provenance tracking in a specific experimental setting. While some concepts from PolyLab have counterparts in MWO (e.g., `plab:LaboratoryProcessPlan` and `cob:Planned process` or `plab:Preparation` and `mwo:Specimen preparation protocol`), other fine-grained classes remain unique (e.g., `plab:LaboratoryActivityRequestSubmission` or `plab:MeasurementStep`). Especially the choice of modelling measurements and physical quantities via IAO in MWO and the Ontology of Units of Measure (OM) [11] in PolyLab leads to an incompatible architecture of subclasses and object properties for these tasks. In summary, PolyLab opens an operational perspective while MWO remains descriptive of the status.

The Metadata4Ing Ontology (M4I) for research data management in engineering disciplines [12, 13] reuses BFO and other OBO-ontologies with their above-mentioned relations, similar to MWO. However, just like PolyLab, M4I also reuses the PROV-O, leading to identical concepts of roles, agents and processes. The disadvantage of designing specific roles, e.g., `m4i:ProjectManager`, as individuals prevents M4I users from assigning actual individuals to their knowledge graph (KG). Moreover, the description of processes remains superficial in M4I compared to PolyLab, as metrics and detailed process steps are missing, in favour of a generic `m4i:ProcessingStep`. The presence of P-Plan classes in PolyLab further intensifies the detail of procedural specification compared to M4I. Besides, the usage of tools and instruments is shifted into as-titled object properties in M4I rather than employing `prov:used` with explicit classes. Nonetheless, M4I fills the gap for methods in BFO by defining a subclass of an entity, just like PolyLab.

PolyLab's choice of modelling physical quantities with OM creates a less fragmented set of triples:

```
:quant1 a                om:Quantity ;
      om:hasValue          :meas1 .
:meas1 a                  om:Measure
      om:hasNumericalValue "123" ;
      om:hasUnit           :unit1 .
```

M4I requires self-defined artefacts for this task (`m4i:hasAssignedValue`, `m4i:hasValue`, `m4i:hasUnit`, `m4i:NumericalAssignment` etc.), as well as artefacts from the Physicalistic Interpretation of Modelling and Simulation Interoperability Infrastructure Ontology (PIMS-II) [14] like `pims-ii:Variable`, `pims-ii:Value`, `pims-ii:QuantityValue`, etc. Moreover, M4I lacks the definition of parameters, covering this with a `pims-ii:Variable` class and specific subclasses that would usually be used in data properties. In brief, both M4I and PolyLab ontologies address scientific work processes, but M4I lacks the finer-grained semantics of plans, orchestration, and provenance that are implemented in PolyLab. M4I focuses more on the resulting data of a scientific investigation.

Many very similar choices to PolyLab in modelling roles, measurements, samples, equipment, calculations, and data interaction are found in the PRovenance Information in MAterials science (PRIMA) ontology [15]. Hence, it harmonizes very much with PolyLab, reuses many identical classes from BFO and OBO Relation Ontology (RO), and even defines many identical classes and object properties that are missing in those reused ontologies. As it focuses more on the data provenance and data lifecycle compared to PolyLab's workflow-centric representation, PRIMA lacks the reuse of P-Plan. In the absence

of `p-plan:Activity` and `p-plan:Plan`, executed processes and plans are coarser-grained, a bit scattered in `cob:completely` executed planned process and `iao:Directive` information entity, and do not reflect each other systematically. Parameters, measures, units, and values are missing altogether. In contrast, PRIMA defines publication attributions, roles of persons, types of data, and modelling in greater detail. Nevertheless, with their great overlap, PolyLab can describe workflow orchestration details in PRIMA’s specifications of the lifecycle of scientific results.

Unlike PolyLab, other above-mentioned ontologies lack the distinction between nominal vs. real measure and planned vs. observed parameter, which is detailed in [Section 4](#). A comparison to non-BFO-aligned ontologies for lab workflows, such as EMMO General Process Ontology (GPO) [16], requires extensive mapping, exceeding the scope of this paper. In summary, PolyLab complements existing ontologies that incorporate lab workflows by combining process execution, experimental documentation, and data lifecycle perspectives in a unified model, and remaining interoperable through the reuse of established ontology standards.

3. Ontology development

The development of PolyLab is motivated by the central role of laboratory workflows in generating and reusing data. Alongside some ELNs and LIMS offering detailed, systematic process recording, PolyLab seeks to formalise laboratory process descriptions and to standardise procedures and templates. By capturing execution records in a structured form, the ontology enables digital documentation of data acquisition and generation, supporting digitalisation in the laboratory environment.

From a knowledge-representation standpoint, the PolyLab ontology defines a data model that formally describes laboratory processes, their constituent components, and their formal provenance. By extending W3C’s PROV-O with laboratory specific concepts, it supplies a machine-interpretable vocabulary that is applicable across multiple scientific domains.

3.1. Ontology development methodology

The development of the PolyLab ontology was conducted in accordance with the Linked Open Terms (LOT) methodology, a systematic approach that supports an agile, iterative refinement of semantic models [17]. Given LOT’s strong orientation toward ontology reuse, it was selected as the foundational methodology because PolyLab relies heavily on the integration of other ontologies as well as established semantic standards and their extensions. The extended development timeline and the need to produce a model that works across several closely related scientific domains required an approach capable of continual generalisation and evaluation – core aspects of the LOT framework. PolyLab also incorporates the previously published PolyMat domain ontology which already encodes a substantial body of laboratory-specific concepts relevant to a wide range of experimental disciplines [3]. The requirement specifications formulated for PolyMat proved instrumental in guiding the extension and formalisation processes underlying the construction of PolyLab.

Expert consultation and direct observation of routine laboratory activities were integral to the requirement specification phase, providing detailed insight into typical planning and execution workflows. The requirements were defined through close collaboration among ontology developers, domain experts, and future users. Ontology developers first immersed themselves in multiple research domains connected with laboratory activities, joining domain experts for an on-site work stay that let them follow daily laboratory tasks. To capture the range of laboratory practice, we carried out on-site investigations in a number of internal lab environments – spanning chemistry laboratories, mechanical-testing rooms, optical-characterisation areas, additive-manufacturing sites, and modest in-institution production workshops. These environments exhibited a wide spectrum of Research Data Management (RDM) practices and were hosted by a few independent scientific institutions. After the on-site visits, remote collaboration continued, leading to the definition of use cases, research applications, and competency questions following the LOT methodology. The resulting ontology, PolyLab, reflects this hybrid, iterative development process.

3.1.1. Ontology purpose and scope

Ontology development began by identifying the concepts and terminology needed to semantically describe laboratory processes. To guide this effort, we examined three complementary aspects. First, an organisational analysis identified the key roles of agents in the ontology (e.g. client, clerk, author, maintainer). Second, a process analysis defined the target hardware and laboratory equipment to be represented. Third, a data analysis addressed relevant data formats and structures, clarifying storage requirements for both digital and physical data. From this foundation, we determined that the ontology should be compatible with ELNs and LIMS, enabling process descriptions to be layered on top of domain knowledge. Finally, the model must strike a balance between simplicity and comprehensiveness. It needs to be straightforward enough to facilitate adoption, support standardisation and digitalisation, and minimise resistance to implementation. At the same time, it has to be detailed enough to contribute to the implementation of the FAIR principles [18] for laboratory artefacts and data management. This allows different user groups to describe processes at the granularity most relevant to them.

3.1.2. Use cases

PolyLab was designed to address the following use cases:

- #1 The PolyLab ontology aims to support the standardisation of templates and predefined procedures, serving as a foundation for consistent documentation and workflow design.
- #2 The ontology is designed to enable cross-referencing of protocol components within ELNs and to support the tracking of laboratory process components and their executions, thereby enhancing protocol integrity and traceability.
- #3 PolyLab addresses the reproducibility of laboratory processes and their constituent elements while supporting the formalisation of records from past runs, enabling the identification of outliers and facilitating predictive analysis.

3.1.3. Competency questions

To establish the functional requirements for the ontology, Competency Question (CQs) were developed jointly by ontology engineers and domain experts. The list was then verified with potential users spanning multiple scientific groups whose research involves substantial laboratory work. Following several rounds of review, a final set of 24 CQs was compiled. A selection of these questions appears below; the complete list is available in the ontology's repository²:

- 2 How long does it take to perform the experiment?
- 4 What characteristics of the process need to be documented during the experiment?
- 6 Who was involved or planned to be involved in a given experiment to perform it?
- 7 What equipment is involved in the experiment?
- 9 What kind of data is stored for later?
- 13 What is the next step for the result of the experiment?
- 14 What substances are used in the experiment?
- 18 Who was the customer of a given process?
- 22 How was the result made available to the customer?
- 23 What publication did the resulting object or data end up in?

The CQs serve to conceptualise the ontology structure. Subsequently, a set of evaluation examples was formulated for each question, which were later transformed into an example dataset – an evaluation KG that can also be found in the repository³.

²https://gitlab.com/dlr-dw/poly-ontologies/polylab-ontology/-/blob/main/doc/polylab_competency_questions_and_evaluation_queries.md

³https://gitlab.com/dlr-dw/poly-ontologies/polylab-ontology/-/blob/main/data/polylab_example_data.ttl

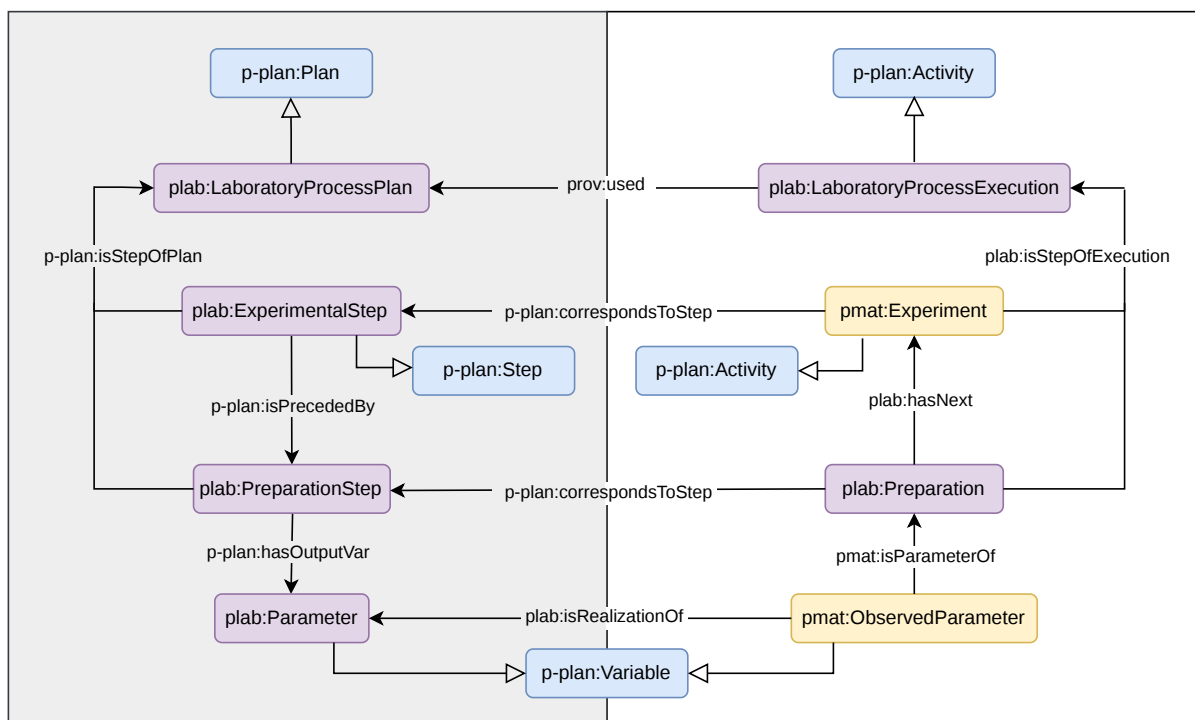


Figure 1: Summary of the laboratory process provenance model for plans and executions. The prospective plan component (grey rectangle, left) is shown alongside the retrospective execution component (right). P-Plan classes are blue, PolyMat classes are yellow. Subclasses are indicated by white-headed arrows.

3.1.4. Ontology reuse

A key requirement was interoperability with existing provenance standards. Since PolyLab is designed to capture both the plans and executions of laboratory processes and their components, we leveraged two complementary ontologies: PROV-O [1] and its extension P-Plan [2]. To strengthen interoperability with the PolyMat ontology and with other ontologies relevant to biological, chemical, and materials science and engineering domains – where the BFO is a widely adopted top-level ontology – we integrated BFO and the RO [19]. Alignment with BFO was further needed to support reuse of chemistry-specific ontologies, which led to the inclusion of selected BFO classes. Key object properties from the RO were also incorporated to ensure adequate relational coverage. Deeper alignment with the abstract BFO classes is not a priority; the current status suffices for reusing other ontologies and for compatibility with PolyMat. To enable formalisation of measurements and their results, we integrated OM [11]. Additionally, a limited number of concepts from other ontologies were used where appropriate. A full list of reused ontologies can be found in the repository of the PolyLab ontology. We adopted a soft reuse strategy for the aforementioned ontologies. Full integration would have introduced unnecessary complexity and instability, and the source ontologies were often too large for our purposes. Since only partial alignment was required, we chose to reference the IRIs of the reused ontologies directly rather than importing them in full.

3.1.5. Conceptualisation and implementation

The ontology was built in OWL with the Protégé editor [20]. Based on domain knowledge, we specified the core classes and relations and arranged them into a coherent taxonomy. We then defined the relevant properties and attributes that describe the key features of each class. To broaden coverage and improve interoperability, we incorporated terms from existing ontologies where appropriate.

We validated the ontology throughout its development and upon completion using two complementary approaches. During critical phases, we examined the model together with domain experts to

confirm its semantic adequacy. In parallel, we generated synthetic test data to assess whether the ontology faithfully captured the intended use cases and scope. Due to data-privacy and intellectual-property restrictions, we could not employ genuine laboratory records. Instead, we designed test cases that mirror real laboratory conditions. These artificial examples served as the foundation for manually constructing a KG that addressed all CQs. The graph was created solely for evaluation purposes. To keep the effort tractable, we did not materialise every conceivable relationship for each individual. At the end, we created SPARQL queries based on the CQs and ran them against the graph. This step verified that the required capabilities were satisfied.

3.1.6. Documentation and publication

To complement the documentation embedded directly within the ontology, a human-readable description was generated using WIDOCO [21]. The PolyLab namespace (<https://w3id.org/polylab/>) is resolved via the w3id.org service, ensuring the persistence of IRIs. The project repository at <https://gitlab.com/dlr-dw/poly-ontologies/polylab-ontology> hosts the CQs, the SPARQL evaluation queries, and the final ontology release under a CC-BY 4.0 license. We welcome comments, feedback, and contributions via the issue tracker and remain committed to the sustained maintenance and evolution of the ontology.

4. The PolyLab ontology

PolyLab allows to describe laboratory workflows in experimental research. As highlighted in Figure 1, it reflects the dualism between (prospective) planning and (retrospective) execution: Many classes used in the documentation of a carried-out experiment have a counterpart on the planning side. Where possible, PolyLab reuses PROV-O and P-Plan to keep the connection between entities on both sides:

```
[a plab:LaboratoryProcessExecution]
  prov:used [a plab:LaboratoryProcessPlan] .
[a pmat:Experiment]
  p-plan:correspondToStep [a plab:ExperimentalStep] .
[a pmat:ObservedParameter]
  plab:isRealizationOf [a plab:Parameter] .
```

4.1. Overview

In the following, we will focus on the execution part of a laboratory process, while noting that corresponding structures for laboratory process planning are also available. PolyLab allows to describe the different activities conducted in a laboratory process. To further classify these activities, various subclasses of `p-plan:Activity` can be used like `plab:Preparation`, `plab:Measurement`, or `plab:Evaluation`. Some of these activity classes are themselves subdivided to represent the entire range of laboratory tasks from preparation via data analysis to documentation. For an overview, kindly refer to [Figure 2](#).

For each activity, further details can be provided using `prov:used`. This allows, e.g., to link the activity to the method applied, the devices employed, or the substances used. Similarly, an activity may have certain outputs. Outputs can be of various types including `plab:ExperimentalData`, `plab:Substance`, or `plab:Consumable`. Maintaining the conventions of the PolyMat ontology, laboratory outputs that are produced during experimental activities (e.g., the product of a reaction) are linked via `prov:generated`, while any other outputs (e.g., data, measurement results, etc.) are linked via `ro:has_output`.

```
[a p-plan:Activity] prov:used      [a pmat:Method] ,
                                   [a pmat:Device] ,
                                   [a pmat:Substance] ,
                                   [a plab:Consumable] ;
ro:has_output                      [a plab:ExperimentalData] ;
prov:generated                     [a pmat:Substance] .
```

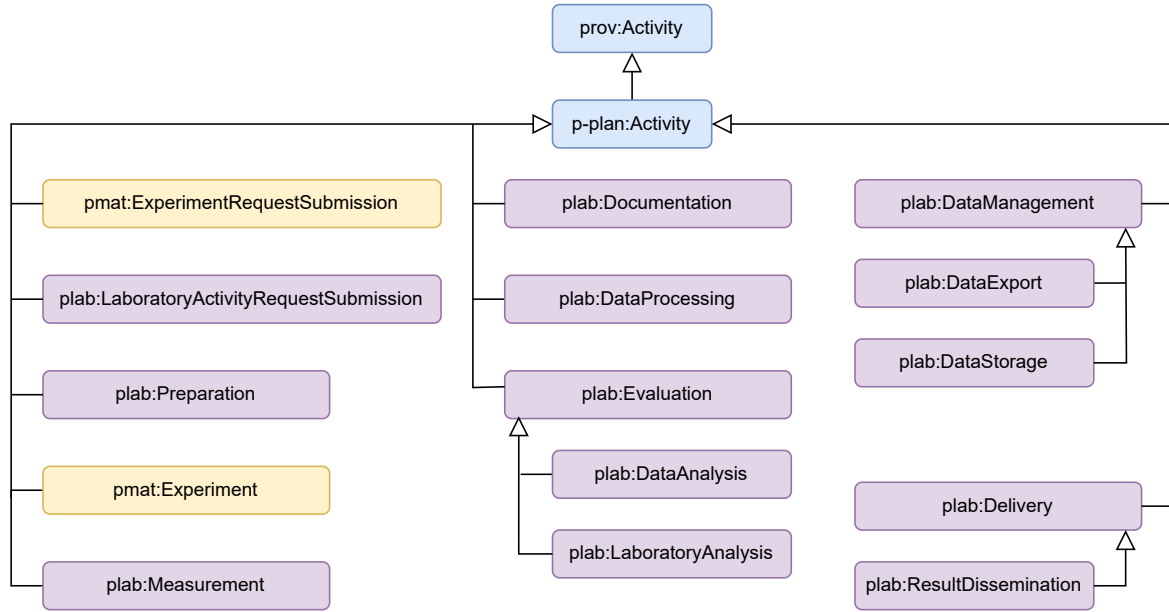


Figure 2: Overview of specialisations of `p-plan:Activity`. Hierarchy via `rdfs:subClassOf`.

Individual activities in a laboratory process execution can then be chained using `plab:hasNext`. This allows to document entire sequences of activities associated to a specific execution given by an instance of `plab:LaboratoryProcessExecution` (see Figure 3).

Laboratory processes, their constituent components, and the plans that organise them may be characterised by a set of variables that describe the conditions under which the processes occur or that serve as fundamental building blocks for the planning and execution stages. These variables include, e.g., entities such as `plab:Consumable` and `pmat:Method`, as well as any other artefacts that define the experimental context. A subset of these variables is explicitly designated as parameters – variables that directly influence the conditions, components, or behaviour of a process or of its individual elements.

In the ontology, we make a clear distinction between two kinds of parameters. On the one hand, we have `plab:Parameter` – the intended (prescribed) parameter for a given process or component, as defined in the plan. On the other hand, there is `plab:ObservedParameter` – the parameter that was actually used during the execution phase. A common example is setting the temperature to a specific value and then observing the actual temperature which may be influenced by other factors. PolyLab allows to clearly distinguish between both aspects and can connect planned and actual value via `plab:isRealizationOf` (also indicated in Figure 1).

```

:plannedActivity1 a p-plan:Step .
:activity1 a p-plan:Activity ;
              p-plan:CorrespondsToStep :plannedActivity1 .
:param1 a plab:Parameter ;
          pmat:isParameterOf :plannedActivity1 .
:obsParam1 a pmat:ObservedParameter ;
            pmat:isParameterOf :activity1 ;
            plab:isRealizationOf :param1 .
  
```

This should not be confused with the relation between `pmat:NominalMeasure` and `pmat:Measure`. While `pmat:NominalMeasure` describes a designated characteristic of an object (e.g., the size of an item as purchased), `pmat:Measure` is the result of a measurement to determine the actual value of that characteristic (e.g., the item's size after measuring it). A quantity corresponds to a parameter, and has measure as a value:

```

[a om:Quantity] plab:correspondsToObservedVariable [a pmat:ObservedParameter] ;
               om:hasValue [a pmat:Measure] .
  
```

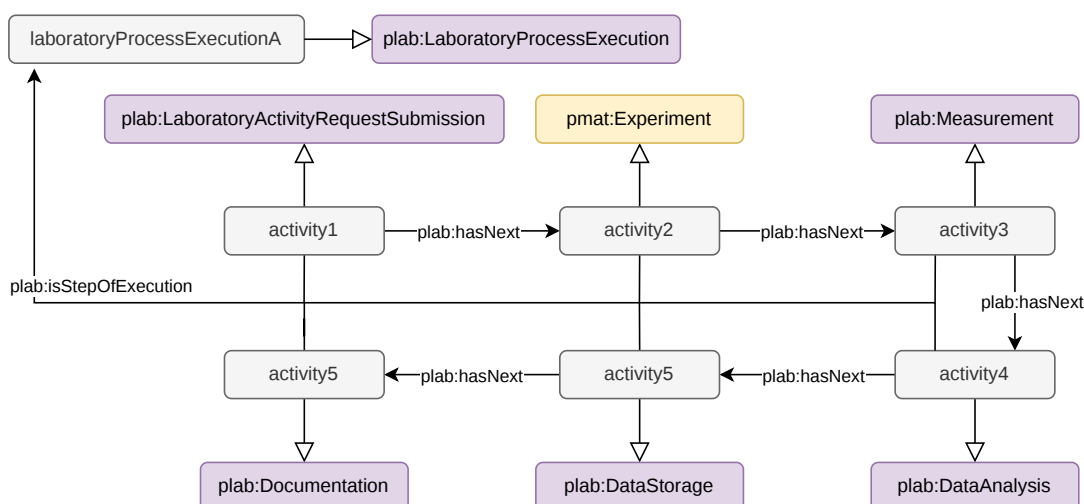


Figure 3: Individual activities within a `plab:LaboratoryProcessExecution` are linked by the object property `plab:hasNext`, forming a chain that documents the exact sequence of activities performed. Subclasses are indicated by white-headed arrows.

Generated chemical entities can further be described by the roles they play in the activity. This could, e.g., be `pmat:Solvent` or other subclasses of `pmat:Role` from PolyMat.

```
:substance1      a          chebi:Chemical_entity ;
                  plab:hasRole [a pmat:Role] .
[a p-plan:Activity] prov:used :substance1 .
```

PolyLab provides a framework for the processing–structure–properties paradigm as defined by Bayerlein et al. [6], in which processing and properties are directly represented, while material structure can be incorporated through the reuse of respective ontologies. The classes defined by PolyLab allow connecting to different aspects described by other ontologies. This ranges from generic ones (e.g., OM for describing quantities) to various levels described by domain-specific ontologies (e.g., ChEBI for substances or PolyMat for individual parts of an experiment). PolyLab relies on these ontologies for the modelling of details and combines them into a comprehensive picture of experiments. By softly reusing BFO concepts, it can be extended with other BFO-compliant ontologies of the chemistry domain and beyond. Taken together, this provides the means to create exhaustive KGs for knowledge extraction and discovery.

The documentation of a process (`plab:LaboratoryProcessExecution`) allows to capture all involved, relevant activities (`p-plan:Activity`), equipment (`pmat:Device`) or consumables used (`plab:Consumable`), planned parameters (`plab:Parameter`) as well as their observed realisations (`plab:ObservedParameter`), and other components. Although only exemplified here by a process execution, similar structures exist for its planning (`plab:LaboratoryProcessPlan`).

```
:characteristic1 a          om:Quantity .
:substance1      a          plab:Consumable, plab:Substance ;
                  ro:has_characteristic :characteristic1 .
:sample1         a          plab:Consumable, pmat:Sample ;
                  prov:wasDerivedFrom :substance1 .
:method1         a          pmat:Method .
:device1         a          pmat:Device .

:experiment1     a          p-plan:Activity, pmat:Experiment ;
                  prov:generated :substance1 ;
                  prov:used      :method1 ;
                  prov:used      :device1 .
```


the ontology's flexibility, allowing users to adopt the representation that best suits their laboratory documentation practices.

5. Discussion

5.1. Ontology development – lessons learned

Reusing existing ontologies emerged as a principal challenge throughout the ontology development process. We first identified a collection of reference models aligned with the conceptual scope of our domain. We then evaluated alternative reuse strategies and decided against a direct import of entire ontologies (hard reuse), even though this approach offers advantages such as automated propagation of updates when the source ontology changes. Combining all of the chosen resources would have produced a model that was overly complex and unnecessarily large. Consequently, to preserve ontological simplicity – thereby facilitating adoption and minimizing implementation resistance – we refrained from wholesale imports. Moreover, some of the source ontologies caused recurring incompatibilities when we attempted to import them into the primary development environment (Protégé). Because these external artefacts are beyond our control, we adopted a soft-reuse approach, integrating only the specific classes, properties, and axioms required for our purposes. We note that this approach requires a systematic, periodic review of the reused fragments so that any updates in the original ontologies can be reflected in our model when needed.

When reusing multiple ontologies, we must align overlapping concepts as well as entities that carry similar or identical labels but are not the same across the source ontologies. In our final model we encounter different naming conventions and definition standards – for some classes the definition is implicit, expressed through a different axiom (e.g., they are intended to be used with), whereas other ontologies provide explicit textual definitions. To navigate this heterogeneous landscape we found it crucial to document the intended use of each concept, to record the alignment decisions we made.

5.2. PolyLab interoperability and reusability

The ontology originated from a HIDA⁵ collaboration among the ontology authors. The initial goal was to create a laboratory process model compatible with the PolyMat ontology, intended to describe processes for a specific domain (polymer membrane research) and to support ELN-based documentation. Since then, parallel initiatives in chemistry, materials science, engineering, physics, and related fields have been advancing the semantic description of laboratory work. In these collaborations we acted as ontology consultants, learning the specifics of process description from domain experts while also strengthening our ties to the ELN community. Through this involvement we observed that, despite domain specific variations, laboratory processes can be expressed using a common structural framework. Accordingly, we refined the original model to maximise interoperability and reusability, ensuring it is usable beyond polymer membrane research and can be adopted across a broader range of scientific domains. As a next step, following the ontology's publication, its feasibility will be evaluated through external real-world use cases – discussions with a couple research groups are already underway.

One of the primary objectives of extending the initial model to a broader domain was to promote its uptake by the scientific community. We identified two plausible adoption scenarios. The first involves groups adopting the existing model “as-is.” To assess reusability at a high level, we applied the model to the description of real data (a KG is currently being prepared as part of an ongoing project and will be published shortly). The lessons learned from this exercise informed improvements to PolyLab prior to its release. The second scenario concerns ontology adoption coupled with extension. In this case we are collaborating closely with scientific institutions that are planning to implement PolyLab to describe their workflows. Our role is to support requirement analyses, cross-check whether PolyLab, in its current form, can accommodate the users' needs, and identify where further refinements or extensions are required.

⁵Helmholtz Information & Data Science Academy

We also considered the aspect of varying granularity requirements for describing laboratory processes. The model was designed to allow users to decide whether to treat an entire laboratory activity as a single unit (e.g., a single `pmat:Experiment`) or to decompose the process into distinct stages (e.g., reusing more PolyLab subclasses of `p-plan:Activity`). The current ontology accommodates both approaches, leaving it to users to determine the appropriate level of granularity and which specific activities should be recorded. Consequently, we regard the development of the ontology as an ongoing process, continuously shaped by adoption experiences and iterative refinements.

6. Conclusion

We presented PolyLab, an ontology that formalises both the planning and execution of laboratory processes. This fills a gap in the existing landscape of ontologies, lacking the details of experimental provenance and workflow. Building on PROV-O and its P-Plan extension, it introduces laboratory-specific classes while remaining domain-agnostic. The ontology was developed iteratively and collaboratively, combining expert workshops, on-site lab observations, and a set of competency questions to ensure relevance and completeness. Reuse of established standards ensures the integration of and interoperability with relevant BFO-aligned domain ontologies, while keeping the model lightweight enough for easy adoption in ELNs and LIMS. The provision of example data and queries offers a self-guided adoption by other researchers. Validation with a synthetic KG and SPARQL queries confirms that PolyLab satisfies its core use cases – standardising protocol templates, tracking execution provenance, and supporting reproducibility and FAIR-compliant documentation – across chemistry, materials science and related experimental domains.

Acknowledgments

MD acknowledges the Helmholtz Information & Data Science Academy (HIDA) for their financial support enabling a short-term research stay at the Institute of Membrane Research of the Helmholtz-Zentrum Hereon in Geesthacht to get familiar with their laboratory environment as a crucial part of requirement analysis for PolyLab development.

Acknowledgements are also due to over a dozen domain experts whose work MD could observe and who served as points of contact for laboratory work and other processes in various scientific institutions. The results presented in the paper could be achieved due to the support of the Impuls- und Vernetzungsfonds of the Helmholtz Gemeinschaft Deutscher Forschungszentren e.V. within the 2025 project cohort of the Helmholtz Metadata Collaboration (HMC, Project number: ZT-I-PF-3-101, LabFriend).

Declaration on Generative AI

During the preparation of this work, the author(s) used GPT-OSS 120B in order to grammar- and spell-check, paraphrase, and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication’s content.

References

- [1] T. Lebo, S. Sahoo, D. McGuinness, K. Belhajjame, J. Cheney, D. Corsar, D. Garijo, S. Soiland-Reyes, S. Zednik, J. Zhao, PROV-O: The PROV ontology, 2013. URL: <https://www.w3.org/TR/prov-o/>.
- [2] D. Garijo, Y. Gil, Augmenting PROV with Plans in P-PLAN: Scientific Processes as Linked Data, in: T. Kauppinen, L. C. Pouchard, C. Keßler (Eds.), Second International Workshop on Linked Science 2012 - Tackling Big Data, LISC@ISWC2012, 2012. URL: <https://ceur-ws.org/Vol-951/paper6.pdf>.
- [3] M. Dembska, M. Held, S. Schindler, PolyMat - bringing semantics to polymer membrane research, in: A. Valdestilhas, H. Li, P. Lambrix, H. Sack (Eds.), 1st International Workshop on Semantic Materials

- Science: Harnessing the Power of Semantic Web Technologies in Materials Science, SeMatS 2024, CEUR Workshop Proceedings, CEUR-WS, 2024. URL: <https://ceur-ws.org/Vol-3760/paper6.pdf>.
- [4] R. Arp, B. Smith, A. D. Spear, Building ontologies with basic formal ontology, The MIT Press, Cambridge, Massachusetts, 2015.
 - [5] O. Bruns, T. Tietz, J. Waitelonis, E. Posthumus, H. Sack, NFDIcore 2.0: A BFO-Compliant Ontology for Multi-Domain Research Infrastructures, 2024. doi:10.48550/ARXIV.2410.01821.
 - [6] B. Bayerlein, M. Schilling, H. Birkholz, M. Jung, J. Waitelonis, L. Mädler, H. Sack, PMD Core Ontology: Achieving semantic interoperability in materials science, Materials & Design 237 (2024) 112603. doi:10.1016/j.matdes.2023.112603.
 - [7] H. Beygi Nasrabadi, E. Norouzi, K. Hubaiev, J. Waitelonis, H. Sack, NFDI MatWerk Ontology (MWO): A BFO-Compliant Ontology for Research Data Management in Materials Science and Engineering, Advanced Engineering Materials (2025). doi:10.1002/adem.202502331.
 - [8] W. Ceusters, An information artifact ontology perspective on data collections and associated representational artifacts, Studies in Health Technology and Informatics (2012). doi:10.3233/978-1-61499-101-4-68.
 - [9] W. Ceusters, B. Smith, Aboutness: Towards foundations for the information artifact ontology, in: Proceedings of the Sixth International Conference on Biomedical Ontology (ICBO), CEUR vol. 1515, 2015, pp. 1–5. URL: <https://ceur-ws.org/Vol-1515/regular10.pdf>.
 - [10] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.-A. Sansone, R. H. Scheuermann, N. Shah, P. L. Whetzel, S. Lewis, The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, Nature biotechnology 25 (2007) 1251–1255. doi:10.1038/nbt1346.
 - [11] H. Rijgersberg, M. van Assem, J. Top, Ontology of units of measure and related concepts, Semantic Web 4 (2013) 3–13. doi:10.3233/sw-2012-0069.
 - [12] D. Iglezakis, D. Terzijska, S. Arndt, S. Leimer, J. Hickmann, M. Fuhrmans, G. Lanza, Modelling Scientific Processes With the m4i Ontology, Proceedings of the Conference on Research Data Infrastructure 1 (2023). doi:10.52825/cordi.v1i.271.
 - [13] G. Lanza, D. Iglezakis, M. Fuhrmans, M. Jordan, B. Farnbacher, A. A. Sosa Rodriguez, S. Leimer, S. Hachinger, S. Arndt, D. Terzijska, A. Wellmann, J. Theissen-Lipp, J. Munke, Metadata4ing: An ontology for describing the generation of research data within a scientific activity., 2025. URL: <http://w3id.org/nfdi4ing/metadata4ing/>. doi:10.5281/zenodo.5957103.
 - [14] M. T. Horsch, Mereosemantics: Parts and signs, in: T. P. Sales, E. M. Sanfilippo (Eds.), FOUST 2021: 5th Workshop on Foundational Ontology, held at JOWO 2021, 2021. URL: <https://ceur-ws.org/Vol-2969/paper3-FOUST.pdf>.
 - [15] R. Aversa, A. Boubnov, D. De Angelis, C. Eschke, M. Kabbe, F. Kraus, M. Panighel, R. Thelen, D. Valentinis, The prima thesaurus for materials science and engineering, Advanced Engineering Materials (2026) e202503159. doi:<https://doi.org/10.1002/adem.202503159>.
 - [16] Simon Stier, General-process-ontology/ontology: v0.2.0, 2023. URL: <https://gpo.ontology.link>. doi:10.5281/zenodo.7799526.
 - [17] M. Poveda-Villalón, A. Fernández-Izquierdo, M. Fernández-López, R. García-Castro, LOT: An industrial oriented ontology engineering framework, Engineering Applications of Artificial Intelligence 111 (2022) 104755. doi:10.1016/j.engappai.2022.104755.
 - [18] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3 (2016). doi:10.1038/sdata.2016.18.

- [19] C. Mungall, J. A. Overton, D. Osumi-Sutherland, M. Haendel, Mbrush, RO, 2015. URL: <http://obofoundry.org/ontology/ro.html>. doi:10.5281/zenodo.32899.
- [20] M. A. Musen, The protégé project: a look back and a look forward, *AI Matters* 1 (2015) 4–12. doi:10.1145/2757001.2757003.
- [21] D. Garijo, WIDOCO: a wizard for documenting ontologies, in: *The Semantic Web – ISWC 2017*, Springer, Cham, 2017, pp. 94–102. doi:10.1007/978-3-319-68204-4_9.