PolyMat - bringing semantics to polymer membrane research

Marta Dembska[1,*,†], Martin Held[2,†] and Sirko Schindler[1,†]

[1]*German Aerospace Center (DLR), Institute of Data Science, Mälzerstraße 5, 07745 Jena, Germany*
[2]*Helmholtz-Zentrum Hereon, Institute of Membrane Research, Max-Planck-Str. 1, 21502 Geesthacht, Germany*

Abstract

Electronic laboratory notebooks (ELNs) and other data management systems are increasingly replacing more traditional means of documenting experimental processes to foster (meta)data capture and reuse. However, the usage of free-text hinders automated, large-scale data processing and invites a variety of data quality issues. Ontologies can address this issue by providing the necessary vocabulary and context — a quality that also puts them at the core of the FAIR principles. The reusability of free-text lab notes increases enormously through semantic descriptions of experimental data. Still, to date, many domains are lacking sufficiently expressive ontologies for more advanced features like consistency checks at data collection or large cross-experiment analyses.

For the field of polymer membranes, we present PolyMat, an ontology to document laboratory experiments and their results. Located at the crossroads of material science and chemistry, this ontology acts as a bridge and can enable new cross-domain discoveries. It is specifically designed to be used in electronic laboratory notebooks and applicable for standardisation of terminology there, to ease and improve FAIR-compliant data collection from the get-go.

Keywords
Ontology, Polymer Membrane, Electronic Lab Notebook

1. Introduction

Membranes play a crucial role in various applications of chemical technology including desalination of seawater, removal of fertiliser residue from drinking water, purification of carbon dioxide before storage, or separation of natural gas and hydrogen in a mixed gas grid – essential tasks in a sustainable world dealing with the effects of climate change [1, 2]. Polymeric materials, renowned for their outstanding processability, cost-effectiveness, and abundance, remain central in membrane development. In the interdisciplinary field of membrane science and technology, collaboration spans various disciplines. Polymer chemists contribute to the development of innovative membrane materials, while physical chemists and mathematicians work on models to characterise transport properties. Finally, chemical engineers design large-scale industrial separation processes. The expanse of this domain adds a considerable layer of complexity when Research Data Management (RDM) is concerned.

Similar to many areas of research, membrane science and technology is undergoing a shift towards an increasingly digitalised environment. With controlled laboratory experiments at its core, electronic laboratory notebooks (ELNs) become an indispensable building block in this transition by providing the interface to document each step in an experiment and sometimes even (semi)automatically import results from measurement devices. This enhances the quality of such records and improves reproducibility of experiments. A strong influence are the FAIR principles [3] which focus on the findability, accessibility, interoperability, and reusability of research artefacts. Being at the core of modern RDM, they are commonly addressed by using suitable ontologies in metadata descriptions. Semantic concepts complement free-text fields by providing contextual frameworks for measurements, research software, and other research outputs. The domain of polymer membranes lacks a formal semantic description, impeding a common understanding of terminology by the various disciplines with their local vocabularies. An unanimous membrane ontology would constitute a machine-actionable, structured knowledge base, paralleling literature and enabling cognitive reprocessing by researchers or machine learning models.

In this paper, we introduce PolyMat, an ontology aimed at knowledge representation within the domain of polymer membrane research. PolyMat serves as a framework for capturing and organising information relevant to laboratory experiments, their results, and the modelling of laboratory processes. The ontology is specifically designed to support the semantic annotation of experiments using ELNs, thus fostering low-effort FAIRification of experimental results. Being at the intersection of material science and chemistry, it provides a bridge between these domains and can enable cross-domain discoveries.

1.1. Online resources

Resource type Ontology
Licence CC BY 4.0 International
URL https://w3id.org/polymat/
GitLab https://gitlab.com/dlr-dw/poly-ontologies/polymat-ontology
DOI 10.5281/zenodo.10286389 **TiB terminology service** https://terminology.tib.eu/ts/ontologies/pmat

1.2. Terminology

The Polymat ontology is crafted for a particular domain and thus requires certain domain-specific terms. While also defined in the ontology itself, we repeat some here to ease understanding.

Monomer A chemical substance whose molecules can be joined together to form a polymer.
Polymer Material made of long, repeating chains of molecules.
Membrane A barrier separating two volumes (both can be in a similar phase states).
Module Several membranes arranged sequentially.

2. Related work

2.1. Ontologies in materials science and engineering and chemistry

Materials science and engineering (MSE) comprises a vast community with diverse subfields, presenting challenges in data interoperability and standardisation in RDM. One strategy for achieving interoperability involves the reuse of existing ontologies and adherence to best practices during the development. Further, alignment with Top Level Ontologies (TLOs) streamlines the harmonisation of different conceptual frameworks. To establish a standardised representational ontology framework rooted in current understanding of materials modelling and characterisation, the Elementary Multiperspective Material Ontology (EMMO) was developed [4]. It currently serves as the most commonly used TLO in MSE, with the Basic Formal Ontology (BFO) right after [5]. However, as polymer membrane research incorporates aspects of chemical engineering, existing MSE-specific ontologies frequently lack critical elements describing corresponding processes, inhibiting their reuse.

Considering the prominence of polymer membranes and chemical processes, specific ontologies related to chemistry become essential components for knowledge representation within the polymer membrane domain. Chemistry relies heavily on the accurate identification and categorisation of chemical substances and their reactions. Information on chemical substances is readily accessible in databases such as PubChem [6], CAS [7], or ChemSpider [8] where unique numerical identifiers or chemical structure identifiers like SMILES [9] or InChI [10] facilitate the linkage of data from diverse sources. Unfortunately, the multiple domain ontologies in chemistry are mainly used without a coordinated approach unlike, e.g., the field of biomedicine where the OBO Foundry [11] is the nucleus for most if not all relevant ontologies. Still, most of the prominent ontologies for the chemistry domain adhere to the OBO Foundry principles and are aligned to BFO [12]. Especially the BFO- and OBO-based Chemical Entities of Biological Interest (ChEBI) [13] ontology finds extensive utilisation in chemistry as it integrates seamlessly with other domain-specific ontologies and offers a comprehensive and well-documented classification of chemical entities but is also considered a domain ontology in MSE.

2.2. Electronic Laboratory Notebooks

As laboratories increasingly embrace digitalisation, ontologies become pivotal in providing a structured understanding of chemical experiments. Thus, effectively integrating the use of electronic laboratory notebooks (ELNs) implies simultaneously modelling the course of the experiments themselves. This involves collecting data provenance of laboratory processes as well as the measurement results themselves. Ontologies may be used in an ELN for knowledge representation, classification, and connection of entries. Only a limited number of ELNs are specifically designed for the chemical sciences. Chemotion [14] is a prominent web-based ELN developed for the field of synthetic and analytical chemistry. Implementing Chemical Methods Ontology (CHMO) [15] and Name Reaction Ontology (RXNO) [16] ontologies, it is one of very few free ELNs in chemistry employing ontologies. Users can annotate a method used in chemical analysis and a type of reaction utilising concepts available respectively in CHMO and RXNO. These annotations enhance the Chemotion dataset by enabling connections based on method or reaction types and facilitating ontology-based searches within the Chemotion repository [17].

Presently, the ELN facilitates data transfer to the Chemotion repository, a platform designed for storing and managing chemical data that also offers various search options, including filtering by CHMO.

Due to the diversity and interdisciplinary nature of MSE, general-purpose ELNs fall short in addressing the entire range of the domain and only few specialised ELNs are available. Kadi4Mat [18] is a generic ELN directed towards MSE experiments and simulation. Garabedian et al. present several studies concerning the integration of a vocabulary [19, 20] and an ontology [21] into Kadi4Mat, aiming at generating user interfaces and a completely FAIR treatment of laboratory data [22, 23]. Furthermore, the non-generic ELN Herbie [24] is intended to focus even more on the MSE aspects in the domain by enabling fully structured and validated entries. Notably, this ELN stands out as one of the first in the domain of chemistry and MSE that incorporates ontologies in its core. The system is designed to offer users the flexibility to select and implement ontologies during form design, allowing for automatic form creation based on the chosen ontology. This feature ensures a semantic connection between form fields and their outputs in the ELN records, enhancing data integration and usability.

3. PolyMat ontology development

The motivation behind the PolyMat ontology stems from two key objectives: the documentation of laboratory processes and knowledge representation within the domain. The domain of polymer membrane research presents a challenge for solely reusing domain-specific ontologies due to its unique combination. Despite these challenges, by establishing relationships between domain-specific terms and connecting them with existing domain ontologies, the semantic annotation of laboratory metadata becomes possible. This documentation also enhances RDM practices in the domain of polymer membrane research, aligning with the ongoing digitalisation initiatives in the field. These efforts of establishing digital processes and tools encompass both bottom-up and top-down approaches, without necessarily implying organised consortia.

In terms of knowledge representation, establishing the PolyMat ontology is designed to address future needs for modeling laboratory processes. By semantically representing a substantial portion of knowledge, which has so far been documented only in paper notebooks or resides within researchers' minds, we can capture this information and make it accessible for both human and machine use. The ultimate goal is to seamlessly integrate the ontology with ELNs, contributing to a searchable, queryable, more efficient, and streamlined research environment.

3.1. Methodology

The ontology was constructed utilising the Linked Open Terms (LOT) methodology [25], recognised as a suitable framework for crafting ontologies and vocabularies tailored for industry projects. The LOT methodology is known for its lightweight and iterative approach, which proved useful when involving domain experts without formal ontology engineering backgrounds. Built on existing methodologies, LOT thoroughly addresses various crucial aspects of PolyMat ontology development. The intentional inclusion of significant ontology reuse supports ongoing community development, which aligns well with the methodology. Given our emphasis on aligning with industrial development alongside academic, research, and software
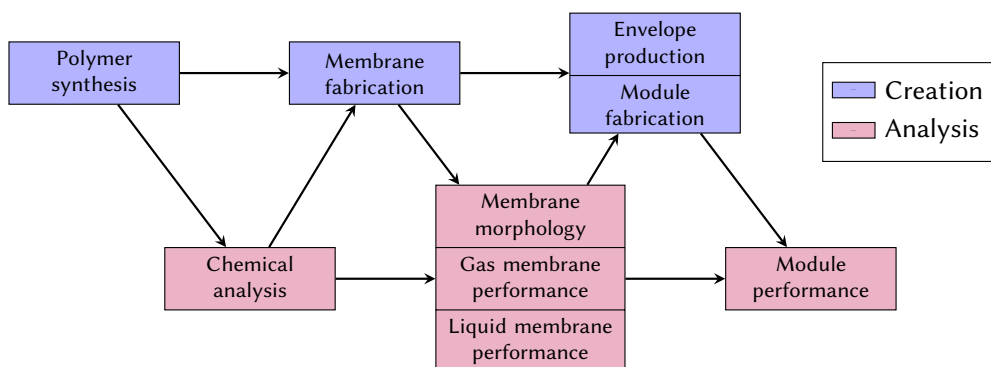
Figure 1: Laboratory experiments workflow in the polymer membrane domain. The workflow includes two main groups of methods: creation and analysis. Creation involves synthesising polymers and fabricating them into membranes used to produce envelopes and fabricate modules. Analysis involves chemical characterisation of polymers, morphological examination of membranes, and performance evaluation of membranes and modules. The arrows indicate the possible order of methods in a given workflow but the starting point depends on a specific use case.

development initiatives, adopting LOT is the most suitable approach for our context. Moreover, LOT's focus on crafting ontologies and vocabularies for Linked Data generation makes it a natural fit for our scope and ontology development process.

3.1.1. Requirements specification

The requirements formulation process relied on close collaboration among ontology developers, domain experts, and future ontology users. The primary task for ontology developers was to thoroughly familiarise themselves with the specifics of the scientific work in the field of polymer membrane research. This was achieved through hybrid collaboration between domain experts and ontology developers, including an on-site research stay where domain experts were accompanied in their daily work. Besides direct conversations, domain data sources included posters, paper laboratory notebooks, experiment protocols, and others. This collaboration provided valuable insights into the procedures, intricacies of laboratory processes, and the typical infrastructure found in such institutions. Given the nature of this field, the processing of both physical data (e.g., substrates or laboratory materials) and electronic data (e.g., measurements obtained from digital output devices) played a significant role. Particularly, aspects like data storage and access were crucial from an RDM perspective. It was necessary to gather information about processes and practices other than laboratory-related ones, including the document circulation cycle, planning and preparation of laboratory processes, and the flow of information within the organisation. The on-site stay was followed up by remote collaboration between domain experts and ontology developers. Based on the acquired domain knowledge, use cases, research applications, and competency questions were defined in accordance with the LOT methodology. The PolyMat ontology emerged as a result of this collaboration.

3.1.2. Scope

First, ontology developers organised laboratory processes within the polymer membrane domain, as shown in Figure 1, to define essential terminology and the structuring of the PolyMat ontology at an early stage. Additionally, the ontology developers identified Chemotion and Herbie as suitable ELNs for the research applications relevant in polymer membrane research due to their ability to manage complex, multi-step processes with standardised protocols enabling reproducibility and quality assurance. Herbie's modular structure and lifecycle management capabilities allow for comprehensive tracking of experiments, while Chemotion's specialised features for chemical documentation and data transfer support interdisciplinary work and facilitate easy data sharing within the scientific community. Both systems provide the flexibility and integration needed for efficient research in polymer membranes. The close collaboration between developers ensures a seamless integration of both ELNs.

3.1.3. Use cases

The PolyMat ontology is intended to fulfil the following objectives and use cases:

#1 The primary aim of the PolyMat ontology is knowledge representation within the domain of polymer membrane research.
#2 The ontology is designed to document scientific work and laboratory processes. This is instrumental in promoting good practices in data management and advancing RDM.
#3 The integration of the PolyMat ontology is planned for the Herbie ELN, a system presently undergoing development at Helmholtz-Zentrum Hereon.
#4 The PolyMat ontology is set to provide the basis for a future modelling of laboratory processes, complementing another ontology currently in development.

3.1.4. Competency Questions

The initiation of the ontology implementation process involved the collaborative creation of a set of Competency Questions (CQs). Following the visit, ontology engineers, aided by domain experts, determined the scope and purpose of the ontology. This led to the selection of groups of upper-level concepts that met the requirements. Using these concepts (e.g. experiment, method, device, characteristic, person, data), 19 CQs were formulated. Several examples of these CQs are presented below:

1 What polymers were used to fabricate a given membrane?
4 What method was used for a given polymer synthesis or membrane fabrication?
9 Who were the persons performing a given experiment?
10 What equipment was used in a given experiment?
11 What characteristics of polymers or membranes are recorded?
15 Where are the results of a given calculation stored?
18 When was an experiment request submitted?

The full list of the CQs is available in the Gitlab repository of the ontology[1]. The CQs, integral to the ontology development framework, played a pivotal role in shaping the definition of

---

[1]https://gitlab.com/dlr-dw/poly-ontologies/polymat-ontology/-/blob/main/doc/competency_questions.md

necessary classes and properties. To systematically address each CQ, examples of use were developed. These were later leveraged in the creation of an example dataset.

3.1.5. Ontology reuse

To enhance the interoperability of the PolyMat ontology with other domain-specific ontologies, we adopted a soft reuse approach. This approach, as demonstrated by Poveda Villalón et al. [26], involves referencing the IRIs of the reused ontology. The decision was made to minimise unnecessary overhead, especially when importing ontologies that contain a substantial number of concepts, not all of which may be directly applicable. This consideration was particularly relevant for the reuse of ChEBI. Given that the PolyMat ontology is intended for implementation in Herbie, which will be synchronised with Chemotion, a primary objective was to achieve a high level of compatibility with Chemotion. Additionally, the reuse of other chemistry-specific ontologies necessitated BFO compatibility, requiring the use of selected BFO classes. Further, the incorporation of selected object properties from the OBO Relation Ontology (RO) [27] in PolyMat was deemed essential. Since research data provenance documentation is a key requirement, this was achieved through the reuse of selected parts of the PROV Ontology (PROV-O) [28]. Additionally, to facilitate the formalisation of measurements and their results, the Ontology of Units of Measure (OM) [29] was incorporated.

3.2. Conceptualisation and implementation

This ontology was formulated in the OWL language utilising Protegé [30]. Based on understanding the domain, main entities and relationships were defined and subsequently organised into a structured taxonomy. Properties and attributes of these entities were identified to capture essential characteristics. Furthermore, elements from existing ontologies were seamlessly integrated, enhancing the ontology's comprehensiveness and interoperability.

Validation of intermediate states as well as the final result took two forms: Firstly, at various stages we discussed the current state of the ontology with domain experts to verify its content-related correctness and with ELN developers to align it with their current development. Secondly, we created a set of artificial test data to evaluate whether the ontology can accurately represent the domain of polymer membranes. Adherence to data privacy and intellectual property protection regulations necessitates the exclusion of authentic laboratory data. Instead, test data was created by domain experts to emulate real laboratory experiences. These examples serve as the foundation for manually constructing a knowledge graph that fulfils all aspects outlined in the competency questions. It is important to emphasise that the knowledge graph was exclusively developed for ontology evaluation purposes. To streamline the process, not every potential connection for each individual was generated. Subsequently, SPARQL queries were defined and evaluated based on the previously developed CQs. This helped not only to confirm that the posed requirements could be fulfilled but also allowed to spot regressions throughout the development similar to unit tests in software development.
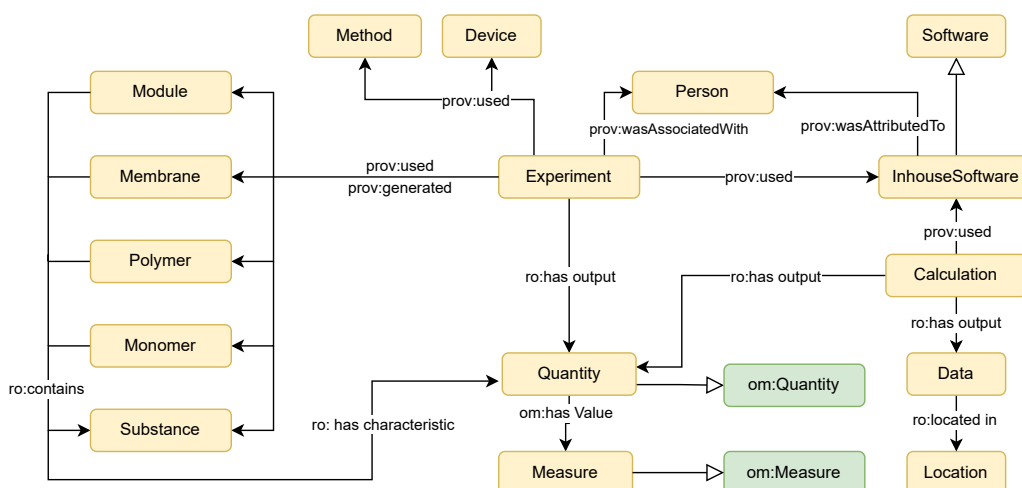
Figure 2: Summary of the PolyMat ontology. Details omitted for readability. PolyMat classes are colored in yellow and classes from re-used ontologies in green. Subclasses are indicated by a white-headed arrow.

3.2.1. Documentation and publication

Besides the documentation inherently part of the ontology itself, we created a human-readable documentation using WIDOCO [31]. The namespace of PolyMat, https://w3id.org/polymat/, is relying on the services of w3id.org persistent IRIs. All intermediate results like CQs or the aforementioned SPARQL queries as well as the final ontology are published at https://gitlab.com/dlr-dw/poly-ontologies/polymat-ontology under a CC-BY 4.0 license. We welcome contributions, comments, and other feedback via the corresponding issue tracker and are committed to further maintaining and advance the ontology.

4. The PolyMat ontology

In polymer membrane research, specialised terminology is used to describe various concepts and characteristics associated with membrane materials, fabrication, devices, and systems. The core components of terminology include structure and properties of monomers, polymers and membranes, creation techniques of polymers, modules and envelopes, and characterisation methods of all the polymer-based elements for membrane technology. Figure 2 illustrates the most significant elements of the PolyMat ontology structure (selected elements and relations have been omitted, despite their discussion in the text, for better readability)[2].

    At the core of the ontology is `pmat:Experiment` characterised by the `pmat:Persons`

---

[2]Here and in the following examples, we use the following namespace prefixes: `chebi:` <http://purl.obolibrary.org/obo/>; `om:` <http://www.ontoloqy-of-units-of-measure.org/resource/om-2/>; `pmat:` <https://w3id.org/polymat/>; `prov:` <http://www.w3.org/ns/prov#>; `rdfs:` <http://www.w3.org/2000/01/rdf-schema#>; `ro:` <http://purl.obolibrary.org/obo/>; `xsd:` <http://www.w3.org/2000/10/XMLSchema#>

involved, the tools used (e.g., pmat:Software, pmat:Device, or pmat:Method), the actual objects of interest, i.e. pmat:Polymers and pmat:Monomers, and the measured characteristics being recorded in form of om:Quantity. Further, instances of pmat:Method describe the underlying workflow of experiments as well as the om:Quantitys being involved.

```
[a pmat:Experiment] prov:used       [a pmat:Device] ;
                    prov:used       [a pmat:Method] ;
                    prov:used       [a pmat:Monomer] ;
                    prov:generated  [a pmat:Polymer] ;
                    ro:has_output   [a om:Quantity] .
```

A pmat:ExperimentRequestSubmission precedes the execution of a pmat:Experiment and encapsulates the planning phase of actual experiments. This documents a rather administrative process but provides a link to resources residing in different systems. A pmat:ComputationalModel can participate in modelling in two ways. Initially, the model can be utilised in the preparation of an experiment scenario as depicted in Figure 2. Secondly, the application of PROV-O via prov:used facilitates indicating whether the model was actively employed during the execution of a pmat:Experiment.

```
[a pmat:ExperimentRequestSubmission]
      ro:is_basis_for_realizable [a pmat:Experiment] .
[a pmat:ComputationalModel]
      ro:has_role_in_modelling   [a pmat:Experiment] .
[a pmat:Experiment]
      prov:used                  [a pmat:ComputationalModel] .
```

pmat:Polymer, pmat:Monomer, and pmat:Membrane and their respective subclasses are the main objects of interest in experiments. Their chemical relationships are represented using ro:contains. Instances are further described by possible additional physical features given by instances of om:Quantity or links to external databases, e.g., via pmat:hasCASNr.

```
[a pmat:Polymer]  ro:contains [a pmat:Monomer] .
[a pmat:Membrane] ro:contains [a pmat:Polymer] .

pmat:Copolymer              rdfs:subClassOf pmat:Polymer .
pmat:MembraneForLiquids     rdfs:subClassOf pmat:Membrane .
pmat:FlatSheetEnvelopeModule rdfs:subClassOf pmat:Module .

[a pmat:Membrane] ro:has_characteristic [a om:Quantity] .
[a pmat:Monomer]  pmat:has_CAS_nr       "XXX-XX-X"^^xsd:string .
```

Physical features (om:Quantity) can often not be measured directly but are the results of more or less complex calculations. This fact is represented by instances of pmat:Calculation and its relation to the resulting instances of om:Quantity. The software used to execute those calculations is documented in instances of pmat:Software. More details can be included via, e.g., pmat:hasVersion or pmat:usesRuntimeEnvironment. In general, results of pmat:Calculation will also include other outputs represented by instances of pmat:Data.

```
[a pmat:Calculation]  ro:has_output [a om:Quantity] ;
                      ro:has_output [a pmat:Data] ;
                      prov:used      [
                          a pmat:Software ;
                          pmat:hasVersion "1.2.3" ;
                          pmat:usesRuntimeEnvironment [
                              a pmat:RuntimeEnvironment
                          ]
                      ] .
```

For the resulting instances of pmat:Data, the location, both physical and digital, is defined via ro:located_in. This especially considers cases when results are exclusively stored on offline media or can not be accessed via generic interfaces. While such cases are slowly fading out, it is still an important use case to consider.

```
[a pmat:Data] ro:located_in [a pmat:Location] .
```

Part of the provenance record are also the people involved both within the experiments as well as the software development at least for tools maintained in-house. Their contributions are encoded using the PROV-O vocabulary using prov:Attribution and prov:Association.

```
[a pmat:InhouseSoftware] prov:wasAttributedTo   [a pmat:Person] .
[a pmat:Experiment]       prov:wasAssociatedWith [a pmat:Person] .
```

Kindly note that the overview of Figure 2 omits large parts of the details modelled in Poly-Mat. Especially class hierarchies have been omitted for readability's sake. Examples include the above-mentioned hierarchy of pmat:Polymer, pmat:Monomer, pmat:Membrane, and pmat:Module but extend to other areas like pmat:Software and pmat:Calculation.

Proxy classes (e.g., pmat:ChemicalEntity) allow to reuse ontologies like ChEBI representing substances to provide more detail on some aspects. However, this approach does not make any assumptions on which ontology is used to provide the corresponding entities.

4.1. Examples of use

We generated examples of use, which formed the foundation for producing test data[3]. This sample data serves two purposes: Firstly, it documents the intended use of the ontologies by providing examples for common scenarios. Secondly, together with SPARQL queries[4] for each CQ, it allowed us to validate the ontology at several stages. The results of each SPARQL query were assessed with respect to their expected completeness and accuracy. This process was repeated for every example of use (and their respective SPARQL queries) after completing the knowledge graph to validate the ontology. Consequently, we also provide examples for all CQs alongside the ontology[5]. An example is illustrated in Figure 3.

---

[3]https://gitlab.com/dlr-dw/poly-ontologies/polymat-ontology/-/blob/main/data/example_data.ttl
[4]https://gitlab.com/dlr-dw/poly-ontologies/polymat-ontology/-/blob/main/doc/queries.md
[5]https://gitlab.com/dlr-dw/poly-ontologies/polymat-ontology/-/blob/main/doc/competency_questions.md
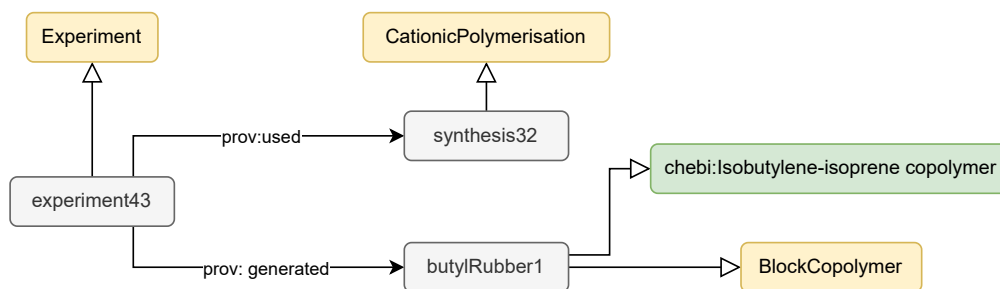
Figure 3: CQ#4. What method was used for a given polymer synthesis or membrane fabrication? Answer: The experiment involves the use of cationic polymerization as the method to synthesise a polymer instance known as butyl rubber.

5. Discussion

We briefly examine the lessons learned from the ontology development process, the impact of this work, its application, and future directions.

Ontology development process. Knowledge representation and adapting modelling for implementation in ELNs posed significant challenges. First, the quality, reusability, and potential harmonisation of existing ontologies are critical considerations. This caused a divergence from the EMMO model and leaning more towards ontologies from chemistry as they exhibit better compatibility and less complex structures. The latter was crucial to ensure the active involvement of domain experts. Here, the support for an agile, iterative approach, particularly concerning ontology reuse, of LOT fulfils these requirements. Next, existing ontologies were often very complex hindering their use in ELNs. Even the ontology implementation in Chemotion initially limits the number of concepts before allowing users to annotate their data.

Scientific impact. In the context of polymer membrane research, existing ontologies often suffer from limitations in quality, completeness, and interoperability, as revealed during the development of PolyMat. Additionally, in this domain, the description of knowledge and experimental procedures are closely intertwined. To ensure a model that is relatively user-friendly, we decided against detailed separation of concepts, as seen in EMMO or other typically modular ontologies. However, the need to integrate descriptions and other laboratory processes led to the development of a separate ontology (to be published soon) solely for modelling laboratory procedures at the project's inception. Both ontologies are designed to be mutually compatible and incorporate data provenance. After sharing our concept with other MSE laboratories, we inspired similar approaches to develop their own modules representing domain knowledge that can be aligned with laboratory process modelling. Specific use cases will be detailed in a paper currently in preparation. The adoption of PROV-O contributes to the reproducibility of experimental results. Moreover, employing the reference model description of provenance ensures interoperability across different domains.

Application and future directions. The integrated development of ELN and ontology, as exemplified by Herbie and PolyMat, enables adjustments to be made cohesively before the testing or implementation phase and serves multiple purposes. First, the ontology will be

used to generate SHACL shapes, facilitating the automated creation of forms within the ELN. Second, it involves the semantic annotation of (meta)data entries within the ELN forms. This is especially crucial for free-text entries, which are typically more challenging for machines to comprehend. With PolyMat, scientific resources are semantically correctly described. This semantic annotation is particularly valuable for future records of laboratory protocols, as it supports text mining applications. Third, irrespective of the data type, the implementation involves metadata enrichment of records from specific experiments. This enrichment aims to contextualise them within a broader framework by fostering more efficient interconnection between specific fields of ELN forms.

While this approach may be limited to a particular setup, it allows for necessary adjustments to both ELN and ontology simultaneously. As the integration of ontologies in ELNs is currently under intense development, diverse coupling mechanisms for different ELNs may arise. Still, the semantic structure provided by ontologies ensures the interoperability of (meta)data from two ELNs when adhering to the same ontology. Future plans include expanding the existing knowledge graph with data from use cases across diverse institutions within the same domain. PolyMat, as a pioneer ontology tailored for ELNs, will be reused to model knowledge of more membrane researcher groups. Therefore, fostering more efficient interconnection between specific fields of forms. It will establish semantically annotated connections between records of laboratory activities to enhance the reproducibility and queriability of results.

6. Summary and conclusion

We presented PolyMat, an ontology for polymer membrane research specifically designed for use within electronic laboratory notebooks (ELNs). It allows documenting laboratory experiments and thus represents a building block towards the further FAIRification of the domain. The development has been conducted in close collaboration with both domain experts and practitioners as well as the developers of a specific ELN, Herbie. We provided a detailed account of its development including Competency Questions as well as SPARQL queries to verify the ontology's comprehensiveness and suitability. All resources are published under a permissive license and are publicly available under both the namespace URL, https://w3id.org/polymat/, and the corresponding development repository, https://gitlab.com/dlr-dw/poly-ontologies/ polymat-ontology. The ontology is accessible through the TiB Terminology Service, https: //terminology.tib.eu/ts/ontologies/pmat. We are further committed to advance the ontology and continue to adapt it to emerging needs especially in context of the continuous spread of ELNs.

Acknowledgments

References

[1] V. Abetz, T. Brinkmann, M. Dijkstra, K. Ebert, D. Fritsch, K. Ohlrogge, D. Paul, K.-V. Peinemann, S. Pereira-Nunes, N. Scharnagl, M. Schossig, Developments in Membrane Research: from Material via Process Design to Industrial Application, Advanced Engineering Materials 8 (2006) 328–358. doi:`10.1002/adem.200600032`.

[2] V. Abetz, Isoporous block copolymer membranes, Macromolecular rapid communications 36 (2015) 10–22. doi:`10.1002/marc.201400556`.

[3] 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`.

[4] G. Goldbeck, E. Ghedini, A. Hashibon, G. Schmitz, J. Friis, A reference language and ontology for materials modelling and interoperability, in: Proceedings of the 2019 NAFEMS World Congress, 2019.

[5] A. De Baas, P. D. Nostro, J. Friis, E. Ghedini, G. Goldbeck, I. M. Paponetti, A. Pozzi, A. Sarkar, L. Yang, F. A. Zaccarini, D. Toti, Review and Alignment of Domain-Level Ontologies for Materials Science, IEEE Access 11 (2023) 120372–120401. doi:`10.1109/ACCESS.2023.3327725`.

[6] S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang, S. H. Bryant, PubChem Substance and Compound databases, Nucleic Acids Research 44 (2015) D1202–D1213. doi:`10.1093/nar/gkv951`.

[7] P. G. Dittmar, R. E. Stobaugh, C. E. Watson, The Chemical Abstracts Service Chemical Registry System. I. General Design, Journal of Chemical Information and Computer Sciences 16 (1976) 111–121. doi:`10.1021/ci60006a016`.

[8] H. E. Pence, A. Williams, ChemSpider: An Online Chemical Information Resource, Journal of Chemical Education 87 (2010) 1123–1124. doi:`10.1021/ed100697w`.

[9] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, Journal of Chemical Information and Computer Sciences 28 (1988) 31–36. doi:`10.1021/ci00057a005`.

[10] S. Heller, A. McNaught, S. Stein, D. Tchekhovskoi, I. Pletnev, InChI - the worldwide chemical structure identifier standard, Journal of Cheminformatics 5 (2013). doi:`10.1186/1758-2946-5-7`.

[11] 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`.

[12] P. Strömert, J. Hunold, A. Castro, S. Neumann, O. Koepler, Ontologies4Chem: the landscape

of ontologies in chemistry, Pure and Applied Chemistry 94 (2022) 605–622. doi:`10.1515/pac-2021-2007`.

[13] K. Degtyarenko, P. de Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcantara, M. Darsow, M. Guedj, M. Ashburner, ChEBI: a database and ontology for chemical entities of biological interest, Nucleic Acids Research 36 (2007) D344–D350. doi:`10.1093/nar/gkm791`.

[14] P. Tremouilhac, C. Lin, P. Huang, Y. Huang, A. Nguyen, N. Jung, F. Bach, R. Ulrich, B. Neumair, A. Streit, S. Bräse, The Repository Chemotion: Infrastructure for Sustainable Research in Chemistry, Angewandte Chemie International Edition 59 (2020) 22771–22778. doi:`10.1002/anie.202007702`.

[15] C. Batchelor, CHMO – Chemical Methods Ontology, 2019. URL: https://github.com/rsc-ontologies/rsc-cmo.

[16] C. Batchelor, RXNO: reaction ontologies, 2020. URL: https://github.com/rsc-ontologies/rxno.

[17] P. Tremouilhac, P.-C. Huang, C.-L. Lin, Y.-C. Huang, A. Nguyen, N. Jung, F. Bach, S. Bräse, Chemotion Repository, a Curated Repository for Reaction Information and Analytical Data, Chemistry–Methods 1 (2021) 8–11. doi:`https://doi.org/10.1002/cmtd.202000034`.

[18] N. Brandt, L. Griem, C. Herrmann, E. Schoof, G. Tosato, Y. Zhao, P. Zschumme, M. Selzer, Kadi4Mat: A Research Data Infrastructure for Materials Science, Data Science Journal 20 (2021). doi:`10.5334/dsj-2021-008`.

[19] N. Garabedian, I. Bagov, K. Weber, C. Greiner, B. Klusemann, F. Bock, M. Held, F. Wieland, C. Eschke, MetaCook: FAIR Vocabularies Cookbook, 2022. doi:`10.5281/ZENODO.7125643`.

[20] I. Bagov, M. Flachmann, N. Garabedian, T. Tiezema, Y. Li, J. Rau, I. Blatter, A. Dollmann, M. Seitz, C. Greiner, Vocabulary of Materials Tribology Lab at KIT, 2023. doi:`10.5281/ZENODO.7709546`.

[21] N. Garabedian, I. Bagov, TriboDataFAIR Ontology, 2023. doi:`10.5281/ZENODO.5720197`.

[22] N. Garabedian, P. J. Schreiber, N. Brandt, P. Zschumme, I. L. Blatter, A. Dollmann, C. Haug, D. Kümmel, Y. Li, F. Meyer, C. E. Morstein, J. S. Rau, M. Weber, J. Schneider, P. Gumbsch, M. Selzer, C. Greiner, Generating FAIR research data in experimental tribology, Scientific data 9 (2022). doi:`10.1038/s41597-022-01429-9`.

[23] N. Brandt, N. Garabedian, E. Schoof, P. J. Schreiber, P. Zschumme, C. Greiner, M. Selzer, Managing FAIR Tribological Data Using Kadi4Mat, Data 7 (2022) 15. doi:`10.3390/data7020015`.

[24] F. Kirchner, C. Eschke, A.-L. Höhme, M. Meller, A. Foremny, M. Held, S. A. Sahim, R. Willumeit-Römer, Herbie - The Semantic Laboratory Notebook & Research Database., 2024. doi:`10.5281/zenodo.12205430`.

[25] 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`.

[26] M. Poveda Villalón, M. C. Suárez-Figueroa, A. Gómez-Pérez, The Landscape of Ontology Reuse in LinkedData, in: Proceedings Ontology Engineering in a Data-driven World (OEDW 2012), 2012.

[27] 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.

[28] 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.

[29] 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.

[30] M. A. Musen, The protégé project: a look back and a look forward, AI Matters 1 (2015) 4–12. doi:10.1145/2757001.2757003.

[31] D. Garijo, WIDOCO: a wizard for documenting ontologies, in: International Semantic Web Conference, Springer, Cham, 2017, pp. 94–102. doi:10.1007/978-3-319-68204-4_9.