

The State of Uncertainty Representation in Copolymer Chemistry Knowledge Artifacts

Sarah T. Tüzün^{1,*}, Birgitta König-Ries¹

¹Friedrich Schiller University Jena, Inselplatz 5, 07743 Jena, Germany

Abstract

Uncertainty is fundamental to copolymer chemistry and arises from different sources such as the stochastic nature of polymerization. Capturing it requires more than storing a value, which is what ontologies and knowledge graphs are promising to offer over plain databases. The knowledge-representation literature treats uncertainty representation as an open problem, but no one has evaluated it for copolymer chemistry specifically. Therefore, we evaluate whether and how existing copolymer chemistry knowledge artifacts (knowledge graphs, ontologies, and databases) represent it. We search four bibliographic databases, four ontology portals, and five GenAI tools to identify every knowledge artifact that models the copolymer chemistry domain. The same screening also separates out the publications on uncertainty in chemistry, from which we build a domain-specific uncertainty vocabulary by extracting keywords algorithmically. We then search all identified artifacts for this vocabulary and classify each genuine hit: whether the artifact only mentions the uncertainty, records it as a value, or models it as a concept of its own. Of the 14 artifacts, only five record uncertainty as a value and none model it as a concept of its own; the rest merely mention it or ignore it. And what they record is mostly prediction uncertainty or the statistical properties of measured values, not the experimental uncertainty itself. The plain databases engage with uncertainty more often than the ontologies and knowledge graphs do, and the only semantic handling we found is inherited from upper ontologies rather than native to the domain.

Keywords

Knowledge Graphs, Copolymer Chemistry, Interdisciplinary Research, Knowledge Artifacts, Uncertainty

1. Introduction

Materials science adopts semantic technologies on a promise: that an ontology or knowledge graph represents more than a table can, capturing not only values but what those values mean. The domain is well supplied with such artifacts; surveys of the materials-science ontology landscape already count dozens [1, 2].

Uncertainty tests that promise, and copolymer chemistry provides a good testing ground. Copolymers are versatile and central to materials and plastics research, and they resist representation: unlike small molecules they have no single structure, but distributions of molecular weight, composition, and architecture [3]. Uncertainty is therefore intrinsic to the domain rather than incidental to it, and it is exactly what a single value cannot represent. A reactivity ratio is not a single number, but is determined with an explicit error and reported as a confidence region [4, 5]. That region is part of the result. It tells a practitioner, for example, whether two measurements agree, and it matters as much to whoever reuses the data as to the chemist who produced it.

The knowledge-representation literature treats uncertainty as an open problem in general [6, 7], but no one has asked it for copolymer chemistry. Therefore, we investigate three research questions:

1. Which knowledge artifacts (ontologies, knowledge graphs, databases, data schemas) model the copolymer chemistry domain?
2. Which domain vocabulary describes uncertainty in (copolymer) chemistry?
3. Do the artifacts from RQ1 model the uncertainty expressed by the vocabulary from RQ2?

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.

✉ sarah.tuzun@uni-jena.de (S. T. Tüzün); birgitta.koenig-ries@uni-jena.de (B. König-Ries)

id 0009-0005-5422-2164 (S. T. Tüzün); 0000-0002-2382-9722 (B. König-Ries)



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

The paper contributes an inventory of the knowledge artifacts that model this domain, a reproducible procedure for building a domain uncertainty vocabulary and searching artifacts against it, and an assessment of where, and at what level, each artifact engages with the uncertainty of its domain.

The remainder of this paper is structured as follows: Section 2 covers related work, Section 3 the methodology, Section 4 the results, and Section 5 the discussion; Sections 6 and 7 state the limitations and conclude with future work.

2. Related Work

Our work is related to three lines of research: 1. copolymer chemistry data and the role of uncertainty, 2. uncertainty in knowledge artifacts, and 3. the analysis and surveying of ontologies. To our knowledge, no previous work examined whether the knowledge artifacts of copolymer chemistry represent the uncertainty of the domain. This work is the continuation of a doctoral proposal [8].

2.1. Copolymer chemistry data and the role of uncertainty

A recurring obstacle in polymer informatics is that polymers have no single structure: they are stochastic molecules, which the deterministic line notations developed for small molecules cannot capture. BigSMILES addresses this with a structurally-based line notation for macromolecules [3], and PolyDAT builds on it as a generic data schema for storing and sharing polymer characterization data [9]. Reviewing these efforts, Kumar warns that unless scientific journals enforce such schemas, the wide dissemination of polymer characterization data on which data mining depends will not succeed [10]. For copolymers (polymers consisting of more than one monomer) the gap is wider still: OpenPoly, a curated database built for benchmarking property prediction, includes only homopolymers with well-defined repeat units, excluding copolymers for their “structural ambiguity and lack of standardized representations” [11]. Structural ambiguity is not the only uncertainty in play. The measurements themselves carry it too, and in copolymer chemistry it is both real and quantified: van Herk and Liu show how knowledge of measurement errors affects the determination of reactivity ratios [4], and Reischke determines confidence contours for reactivity ratios through a Bayesian approach [5].

2.2. Uncertainty representation in knowledge artifacts

Uncertainty representation is an active concern in the knowledge-graph and ontology literature, but it is treated in general terms rather than for a particular experimental domain. Uncertainty management in knowledge-graph construction is the subject of a recent survey [6], a comprehensive review catalogs how those domain ontologies that model uncertainty go about it [7], and Anand and Kumar analyze the kinds of uncertainty that arise in ontology-based knowledge representation [12]. None of these consider copolymer chemistry as an application domain.

2.3. Analyzing and surveying ontologies

Our method compares a domain vocabulary against knowledge artifacts, which places it closest to corpus-based ontology evaluation. Rospocher et al. extract key-concepts from a domain corpus and score how well an ontology’s terminology covers them [13]. We compare a corpus-derived vocabulary against an artifact as well, but our aim differs: instead of scoring an ontology’s overall terminological coverage, we look for a single theme, uncertainty, across heterogeneous artifacts in the source files and publications that describe them, and we classify how each one represents it. Keyword-based access to ontologies has been studied from the opposite direction, using keywords as a query interface to an ontology [14].

A separate line of work surveys the ontology landscape our artifacts belong to, though without asking about uncertainty. Norouzi et al. catalog 94 knowledge artifacts in materials science [1]; De Baas et al. review 43 materials science domain-level ontologies and propose a methodology for aligning them

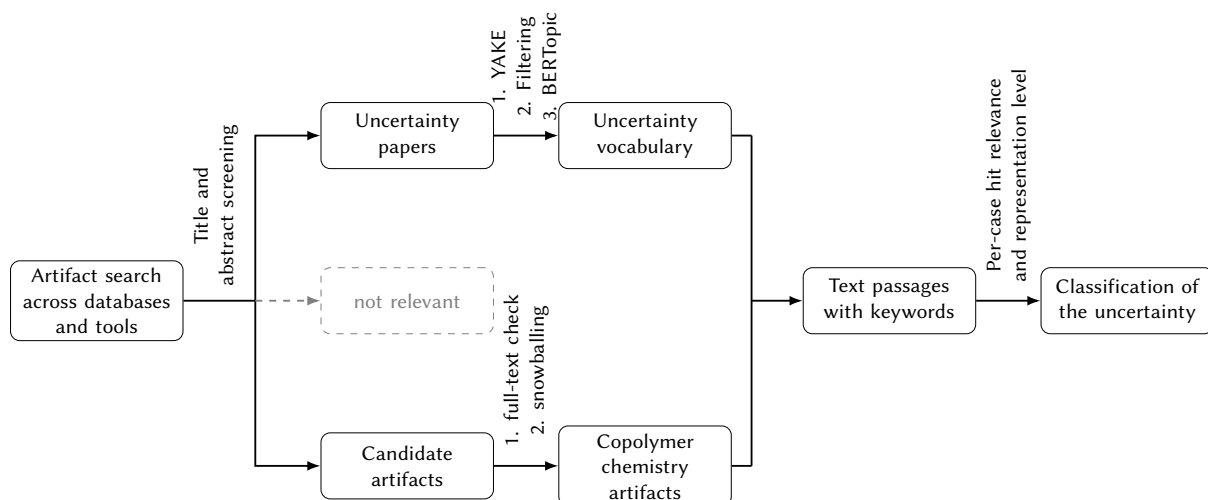


Figure 1: Overview of the methodology followed in this paper. The search across multiple portals feeds both tracks: screening splits its results into papers that discuss uncertainty in chemistry, from which we build the vocabulary, and papers that describe a candidate artifact. The tracks rejoin when we search the artifacts for that vocabulary.

to top-level ontologies, demonstrated on a proof of concept [2]; Beygi Nasrabadi et al. evaluate which upper ontologies best support materials science ontology development, using a Brinell-hardness case study [15]; Strömert et al. survey the landscape of chemistry ontologies with a view to FAIR data [16]. To the best of our knowledge, there is no survey of the knowledge artifacts of copolymer chemistry in particular.

3. Methodology

Figure 1 shows our workflow. A broad search across the bibliographic and ontology landscape of the domain provides the basis for everything that follows. Screening splits three ways: results that describe a candidate artifact, results that discuss uncertainty in chemistry, and results that do neither. The first track answers RQ1 (lower arrow), the second builds the vocabulary for RQ2 (upper arrow), and the two rejoin when we search the artifacts for that vocabulary to answer RQ3.

3.1. Finding the artifacts

To find the artifacts that model copolymer chemistry (RQ1), we use three different sources. We searched four bibliographic databases (Scopus, IEEE Xplore, the ACM Digital Library, and SciFinder) with a platform-dependent search string (given in full in the appendix folder of the accompanying repository [17]), choosing the first three for their broad academic coverage and SciFinder for its chemistry focus. We searched the ontology portals BioPortal, MatPortal, OBO Foundry, and HIVE4MAT for the term “copolymer”. And we prompted five GenAI tools (Claude¹, ChatGPT², Elicit³, Asta⁴ and Consensus⁵) with two prompts, given in the same folder, extracted the citations they returned, and verified each one against Google Scholar, keeping only sources that exist. By putting the results of those three together and removing duplicates we arrived at the corpus we screened.

We screened all titles and abstracts against inclusion and exclusion criteria with ASReview LAB [18]. This tool provides automatic relevance sorting of the provided publications. We included a publication

¹<https://claude.ai> with Opus 4.8 and effort set to “max”, accessed 2026-06-10

²<https://chatgpt.com> with GPT 5.4, accessed 2026-06-10

³<https://elicit.com/>, accessed 2026-06-10

⁴<https://asta.allen.ai/>, accessed 2026-06-10

⁵<https://consensus.app/>, accessed 2026-06-10

if it describes, presents, or evaluates an ontology, knowledge graph, database, or data schema that explicitly represents copolymers or polymers, provided the paper is in English and publicly accessible; we included an ontology if a paper describes it and its source file is publicly available. We excluded work on general chemistry with no specific copolymer classes, and paywalled items to which we have no institutional access. Screening sorted the corpus into either candidate artifacts or uncertainty papers if it fulfilled the inclusion criteria. We then skimmed each candidate artifact paper and recorded the artifact it describes, if any. We cast the net deliberately wide, but kept an artifact only when it actually models copolymers.

From every paper we deemed relevant, we snowballed one level backward and one level forward (through its references and through the papers citing it) and screened whatever surfaced against the same criteria. The ontology portal hits refer to the source files rather than publications, so we handled them alongside this track: we deduplicated them against each other, discarded those whose domain is neither the synthesis nor modelling of copolymers, and deduplicated the remainder against the artifacts recovered from the publications. What survived are the knowledge artifacts we analyze in the rest of the paper.

3.2. Building the uncertainty vocabulary

We are not copolymer chemists, so we derived the uncertainty vocabulary (RQ2) from the literature rather than from our own expertise. Its corpus is the second track from screening: the papers whose title and abstract showed them to be about uncertainty and (copolymer) chemistry, whatever source had surfaced them. Breadth mattered here, because sub-disciplines name the same phenomenon differently and a narrow corpus would yield a narrow vocabulary; for the same reason we added two polymer-specific publications. We built the vocabulary from them in three steps: preprocessing to remove noise, keyword extraction, and clustering to isolate the uncertainty-related terms.

Each paper was parsed from its PDF and preprocessed to remove noise like publisher boilerplates and references. From the cleaned text, we extracted candidate keywords with YAKE [19]. We also tried RAKE [20] but discarded it, as it coped poorly with the residual noise.

The source papers span chemistry with a focus on measurement and uncertainty, so the extracted keywords cover a correspondingly wide range of topics. To isolate the uncertainty-related terms, we clustered the list with BERTopic [21], using the `all-MiniLM-L6-v2` sentence-embedding model, a minimum cluster size of five⁶, and a fixed random seed (42) for reproducibility. We exported the clusters as JSON and inspected them in an HTML viewer, keeping the clusters about uncertainty.

3.3. Searching the artifacts

We then searched the artifacts for this vocabulary. The unit of analysis is the artifact rather than the individual document: where an artifact has a machine-readable ontology or knowledge-graph file alongside its publication(s), we searched both and derived a single per-artifact classification from their combined evidence.

Longer matches take preference⁷, and every match is stored with the previous and following sentence as context. A matched sentence is recorded once, together with the keywords it contains, so that a sentence matching several keywords counts as a single hit. For the ontology files, we process smaller ones with `rdflib`⁸, which preserves the entity each hit belongs to, and larger ones with a text-only processor that returns the line number of the hit.

⁶We started with cluster size three but found the keyword clusters too fragmented.

⁷Take the sentence *Uncertainty measures are important*. If the keyword list includes *uncertainty*, *uncertainty measure* and *uncertainty measures*, the longest match would be taken and counted as the keyword hit.

⁸<https://github.com/RDFLib/rdflib>

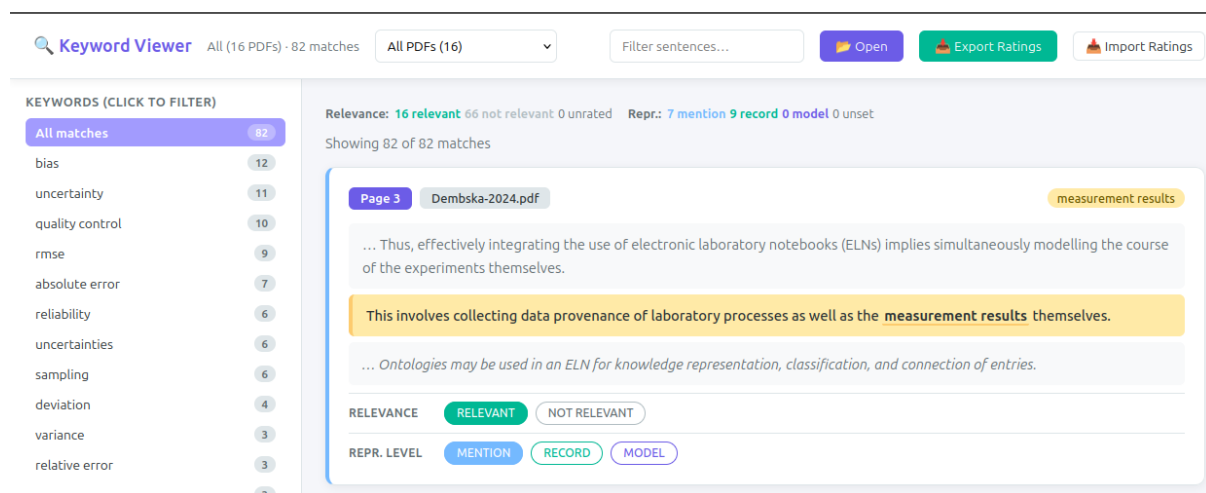


Figure 2: Screenshot of the keyword viewer with a keyword hit from the PolyMat [22] paper. On the left side, the keywords and the number of all matches are displayed. In the middle section, the keyword hit with its surrounding context is shown, which can then be classified according to relevance and representation level. These ratings can be exported and re-imported. We classify hits in both the publications and the artifacts and provide the classifications and the keyword viewers in the repository.

3.4. Classifying the hits

We classified each keyword hit manually, reading the surrounding extracted context in an HTML viewer (Figure 2):

- *Relevance:* Is the sentence actually about uncertainty in the process, experiment, data, or representation? Hits, that are not, are false positives, and therefore excluded. For example, a sentence about selection *bias* in a literature corpus is not relevant for us.
- *Representation level:* for genuine hits, how does the artifact relate to the uncertainty: does it merely *mention* it, *record* it as an explicit value (such as a standard deviation or confidence interval), or *model* it as a concept of its own with dedicated semantics?

The representation level, in particular, is read from the artifact itself rather than from the keyword match: we established what a hit refers to from the original document, and from further sources where needed.

4. Results

All code and results are publicly available on Zenodo [17]. A Jupyter notebook documents the full process.

4.1. Artifact search results

The database and GenAI searches returned 265 of the 292 documents in Table 1. After deduplication and discarding unverifiable GenAI suggestions, 168 items remained. Screening identified 70 relevant papers; the review file is available in [17]. We obtained full text for 61 of these, examined each in detail, and snowballed those we deemed relevant. Snowballing surfaced a further 111 candidate items, which we screened against the same criteria. Not every included paper yielded an artifact: we kept some in the expectation that they would point to one.

The remaining 27 documents, from the four ontology portals, are source files rather than publications and did not enter the review. Deduplicated against each other, they left 21 unique results, listed in the repository’s appendix folder [17]. We excluded those whose domain was neither the synthesis of copolymers nor a closely related topic, and deduplicated the remainder against the artifacts already

found; we dropped two otherwise interesting ontologies, BERO and CRISP, because their original papers could not be retrieved. Following up the remaining hits gained us one document indirectly: the ChemROF ontology hit led to the second, reengineered ChEBI publication [23], which adopts part of ChemROF. After deduplicating across all tracks, 14 artifacts remained. These are the artifacts that model the copolymer chemistry domain (RQ1); they are listed in Table 2.

Table 1

Initial document count of the database and tool search before deduplication

Source	Number of documents
ACM digital library	11
Scopus	56
IEEE	0
SciFinder / CAS	65
BioPortal	18
MatPortal	3
OBO Foundry	0
HIVE4MAT	6
GenAI tools (Prompt 1)	72
GenAI tools (Prompt 2)	61
Total	292

Table 2

The 14 artifacts modelling the copolymer chemistry domain and how each relates to uncertainty. Uncertainty levels: *none*: no genuine uncertainty content (keyword hits are false positives); *mentions*: acknowledged in prose only; *records*: stored as an explicit value, with *exp.* as experimental/measured dispersion and *pred.* as model-predictive; *models*: represented as a concept of its own. Each classification reflects the combined evidence of the artifact’s publication(s) and, where available, its ontology or knowledge-graph file. PoLyInfo and ChEBI are each described by two publications.

Name	Type	Source	Uncertainty
MDS-Onto	ontology	[24, 25]	records (pred.)
Copolymer Descriptor (CopDDB)	database	[26]	none
Block Copolymer Database (BCDB)	database	[27]	records (exp.)
PoLyInfo	database	[28, 29, 30]	mentions
PolyMat	ontology	[22, 31]	mentions
Polymer Li+-Electrolyte Design Database	database	[32]	mentions
NanoMine	KG	[33, 34]	records (exp.)
OPoly26 (Open Polymers 2026)	dataset	[35]	none
Devices, Experimental Scaffolds, and Biomaterials Ontology (DEB)	ontology	[36, 37]	none
Chemical Entities of Biological Interest (ChEBI)	ontology	[38, 23, 39]	none
Polymer Genome	prediction platform	[40]	records (pred.)
CRIPT	data structure	[41]	none
CoPolDB	database	[42]	records (exp.)
Materials Data Vocabulary (NMRRVOCAB)	vocabulary	[43]	none

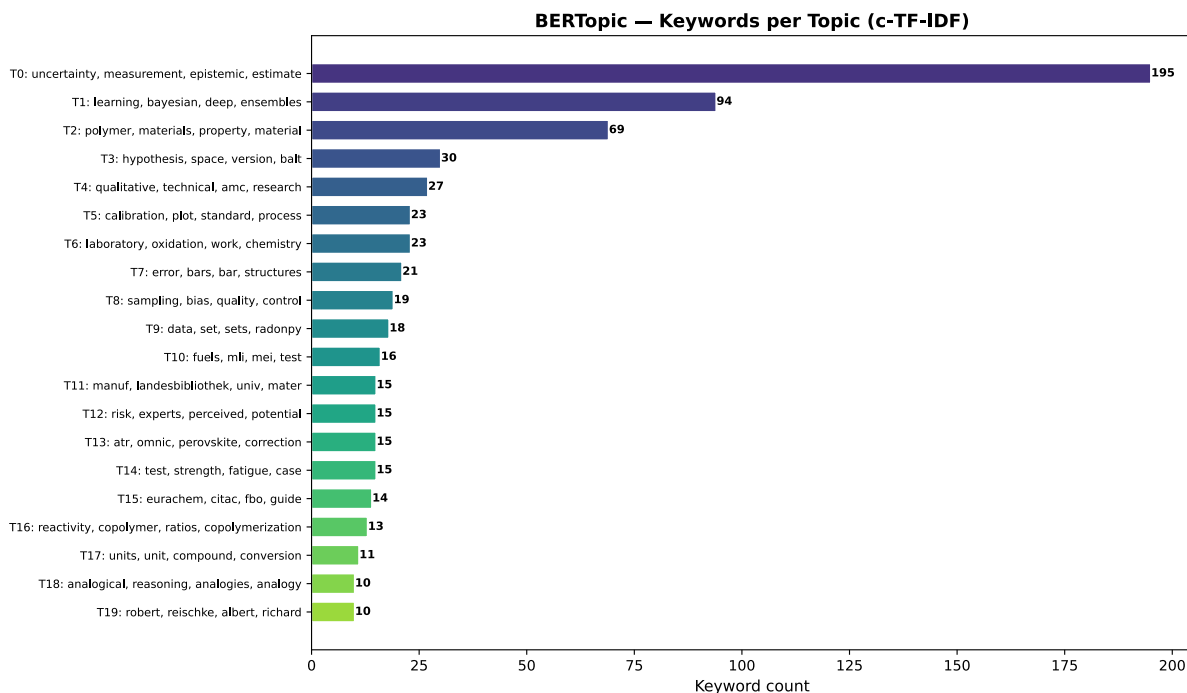


Figure 3: Number of keywords in the leading topic clusters. 20 of 59 are shown here.

4.2. Uncertainty keywords

We applied YAKE [19] to the 30 collected publications and obtained 1064 candidate keywords. Clustering them with BERTopic yielded 59 topics (see Figure 3 for the first 20 clusters). On inspection, three (namely, topics 0, 8, and 29) were about uncertainty. We combined their keywords with the relevant unclustered outliers and, after manually removing noise such as personal names, arrived at a final list of 223 keywords. This list is the domain vocabulary that describes uncertainty in (copolymer) chemistry (RQ2). The publications used for keyword extraction are provided as a separate bibliography file, and the final keyword file is available in the accompanying repository [17].

4.3. Keyword hits in the artifact corpus

The 14 artifacts of Table 2 are described across 16 documents (ChEBI and PoLyInfo each by two) which together form the publication corpus. Only 26 of the 223 keywords appeared at all: 20 in the publications, spread across 13 of the 16 documents, and six more that surfaced only in the machine-readable artifact files discussed below. Three documents gave no hit at all: ChEBI’s original description [38], the NIST materials-data vocabulary [43], and CRIPT [41]. Figure 4 shows the frequency of each matched keyword, and Table 3 how the matches distribute across the documents. For six of the fourteen artifacts, we were able to find an ontology or knowledge-graph file (.owl / .ttl); we searched these in addition to their publications, recording each download URL and date in [17]. The two searches proved complementary rather than redundant: five of the six files produced at least one keyword hit (NanoMine’s did not), and six keywords surfaced only in the files (*uncertainty quantification*, *quantification*, *propagation*, *internal standard*, *measurement information*, and *measurement data*). Conversely, MDS-Onto’s record-level uncertainty appears only in its ontology file, while its publication yields no genuine hit.

4.4. Levels of uncertainty representation

Most keyword hits fail the relevance test when applying the classification scheme consisting of relevance and representation level. False positives include *bias* as selection bias in literature corpora, *sampling* as polymer sampling or *rms/rmse* and *variance* as model-fit or PCA metrics. After removing these,

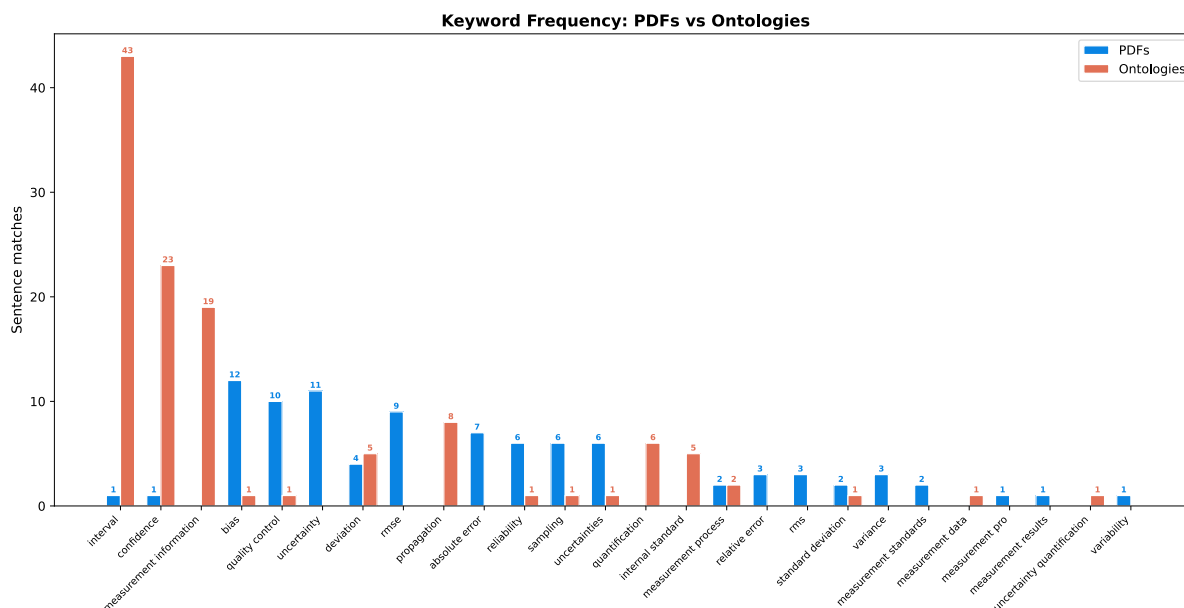


Figure 4: Frequency of each matched keyword across the copolymer chemistry artifacts.

genuine uncertainty content remains in eight of the fourteen artifacts, at two levels of representation, plus one edge case discussed below.

Mentions uncertainty only. Three artifacts acknowledge uncertainty without capturing it. PolyMat [22] notes that it includes measurements but does not follow up on their uncertainty; the same happens in its ontology file, in the role `polymat:TargetConcentration`. This subclass of `polymat:Role` defines “A certain, aspired concentration, in contrast to the real, achieved concentration, due to experimental uncertainties when measuring the volume or weight of a liquid or solid substance” [31]. But there is no `ActualConcentration` class to complement it, and no modelling of that commented uncertainty. PoLyInfo [28, 29] records measurement methods, standards, and conditions, but whether the uncertainty of that process is itself represented remains unclear. In the Polymer Li+-Electrolyte Design Database [32], the uncertainty of experimental parameters is discussed only as something to minimize.

Records uncertainty as a value. Five artifacts store an uncertainty quantity explicitly. In NanoMine [33], this is inherited from the upper ontology SIO [44]: `nanomine:Property` is a subclass of `sio:quantity`, and a property with an uncertainty range carries its own attribute of `sio: 'standard deviation'`. The Block Copolymer Database [27] allows a standard deviation, standard error, or another double value for every quantitative column, and provides a further column in which additional uncertainty metrics can be described. CoPolDB records the 95% confidence limits of reactivity ratios, but only where the polymer handbook it draws on provides them [42]. MDS-Onto [25] provides, in its ontology file, several confidence interval bounds (e.g. `ConfidenceIntervalOfPredictedResponseLowerBound95`) for recording statistics. Polymer Genome [40] surfaces a Bayesian uncertainty, with confidence intervals, for each property prediction. These last two differ from the first three, however: the values they record are model-predictive rather than experimental uncertainty.

Edge Case. MDS-Onto [25] reuses the Common Core Ontologies (CCO)⁹ class “Veracity Measurement Information Content Entity” as a “correctness-of-information” class, which works as a triad with two other classes to describe the relationship between description and reality. This points toward modelling

⁹<https://commoncoreontology.github.io/cco-webpage/>

uncertainty. But the modelling comes from CCO rather than MDS-Onto itself, and only one of the three classes is reused, so we do not count it as modelling.

Models uncertainty as a concept. No artifact models uncertainty as a concept of its own.

Table 3

Keyword hits in artifacts, separated by paper and ontology files.

Keyword	Papers	Ontologies	Sources
quality control	4/16	1/6	Papers: [23], [33], [24], [27] Ontologies: ChEBI
deviation	3/16	1/6	Papers: [40], [33], [27] Ontologies: MDS-Onto
uncertainty	4/16	0/6	Papers: [40], [33], [27], [32]
interval	1/16	3/6	Papers: [35] Ontologies: MDS-Onto, ChEBI, PolyMat
bias	2/16	1/6	Papers: [36], [35] Ontologies: DEB
measurement process	1/16	2/6	Papers: [28] Ontologies: MDS-Onto, PoLyInfo
uncertainties	2/16	1/6	Papers: [40], [32] Ontologies: PolyMat
absolute error	3/16	0/6	Papers: [40], [35], [32]
reliability	2/16	1/6	Papers: [23], [24] Ontologies: MDS-Onto
standard deviation	2/16	1/6	Papers: [33], [27] Ontologies: MDS-Onto
quantification	0/16	3/6	Ontologies: DEB, MDS-Onto, ChEBI
propagation	0/16	3/6	Ontologies: MDS-Onto, ChEBI, PolyMat
variance	2/16	0/6	Papers: [40], [24]
relative error	2/16	0/6	Papers: [40], [35]
sampling	1/16	1/6	Papers: [35] Ontologies: MDS-Onto
rms	2/16	0/6	Papers: [35], [26]
confidence	1/16	1/6	Papers: [42] Ontologies: MDS-Onto
internal standard	0/16	2/6	Ontologies: MDS-Onto, ChEBI
measurement information	0/16	2/6	Ontologies: MDS-Onto, PoLyInfo
measurement results	1/16	0/6	Papers: [22]
measurement standards	1/16	0/6	Papers: [29]
measurement pro	1/16	0/6	Papers: [28]
variability	1/16	0/6	Papers: [40]
rmse	1/16	0/6	Papers: [40]
measurement data	0/16	1/6	Ontologies: MDS-Onto
uncertainty quantification	0/16	1/6	Ontologies: MDS-Onto

5. Discussion

Our results answer RQ3 in the negative: no artifact models uncertainty as a concept of its own, and only five of the fourteen record it even as a value. More telling than the count is which artifacts these are, and what kind of uncertainty they hold.

Uncertainty is handled more often by the plain databases than by the semantic artifacts: four of the five databases engage with it in some form, compared to three of the five ontologies and knowledge graphs. And where the ontologies and knowledge graphs reach for a semantic construct, they borrow it from an upper ontology rather than define their own: NanoMine records uncertainty through SIO’s standard-deviation construct, and MDS-Onto’s single move toward modelling it, a “Veracity Measurement” entity, is reused from the Common Core Ontologies. No domain artifact models uncertainty as a concept of its own; the nearest thing is inherited one level up, consistent with how these resources are built, through reuse.

Polymer Genome stands apart. It is a prediction platform rather than a database, and it records the uncertainty its own application calls for: the Bayesian uncertainty of a property prediction, not the experimental uncertainty behind the measurement. Its developers are explicit about that boundary, citing the glass transition temperature, often reported as a range that depends on the measurement technique, and describing a framework to quantify such experimental uncertainties as still under development [40]. The gap that motivates our work is thus drawn by the developers of a well-established platform themselves.

For semantic materials science, two directions follow. The databases show that the domain’s uncertainty is recordable, so how they do it is worth learning from; and the only semantic handling we found already lives in upper ontologies such as SIO and CCO, so wiring domain artifacts to those (which most do not yet do) is a more direct route than building new machinery. Whether these representations capture the full range of uncertainty practitioners encounter remains an open question. Our interactions with practitioners suggest that they do not, and we leave a systematic investigation of this issue to future work.

6. Limitations

Our assessment rests on two kinds of evidence that are not evenly available across the corpus. For the six artifacts distributed as an ontology or knowledge-graph file, we searched the artifact itself as well; for the remaining eight, available only as a publication, we rely on how the authors describe them. Reading the files directly matters: MDS-Onto’s recorded uncertainty is present only in its ontology file and absent from its publication, so a publication-only study would have missed it. For the eight publication-only artifacts, conversely, a developer may have handled uncertainty without describing it, which only inspecting a machine-readable artifact, or contacting the developers, would reveal.

Our search instrument is lexical, so uncertainty can escape it in two ways. An artifact may encode uncertainty structurally, without a matching lexical label; for the ontology files, we mitigated this by inspecting the structure surrounding each hit, but an unlabeled encoding, there or in the artifacts we could only read about, may have gone unnoticed. Or it may name uncertainty in words our vocabulary does not contain, which produces no hit at all. Neither failure is recoverable: the relevance check discards false positives, but it cannot surface a miss.

Finally, a single rater classified every hit. Therefore, we made every judgement open to inspection: each hit is published with its surrounding context and its assigned level [17], and each representation level is traceable to a passage in the artifact or the sources describing it, so any judgement can be re-examined against the evidence behind it.

7. Conclusion and future work

Materials science adopts semantic technologies to represent more than a table can. For the uncertainty of copolymer chemistry, we are not there yet: none of the fourteen artifacts we examined represent uncertainty as a concept of its own, and only five record it even as a value, three of those for the uncertainty of a measurement rather than of a prediction. The plain databases engage with uncertainty more often than the ontologies and knowledge graphs do, and where a semantic artifact handles it at all, the construct is inherited from an upper ontology, SIO in NanoMine and the Common Core Ontologies in MDS-Onto, rather than native to the domain.

What a practitioner treats as uncertainty is not necessarily what these artifacts record, and a vocabulary drawn from the literature rather than from the laboratory cannot settle the difference. Existing measurement ontologies (e.g. [45, 46]) and the upper ontologies, from which the domain artifacts already borrow, give us machinery to evaluate and extend. But the starting point needs to be the laboratory: we plan to observe copolymer chemists at work and to build the representation from what they call uncertainty.

Acknowledgments

This work is funded by a doctoral scholarship by the State of Thuringia and supported by the Friedrich Schiller University of Jena and the DFG research training group (RTG) 3040 “Copolymer Informatics (COIN): How digital technologies shape copolymer chemistry – From design to application” (DFG project number 527537972). STT thanks Alican Tüzün for the fruitful discussions and comments.

CRedit Statement

STT: Conceptualization, Methodology, Software, Validation, Visualization, Writing - Original Draft; BKR: Supervision, Writing - Review & Editing;

Declaration on Generative AI

During the preparation of this work, the authors used Claude Sonnet, Opus, and Fable in order to: Grammar and spelling check, Improve writing style, Peer review simulation, Citation and Consistency Checking, Drafting Content from notes. Further, the authors used Claude for rendering Figure 1 to TikZ from an original draw.io document. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] E. Norouzi, J. Waitelonis, H. Sack, The landscape of ontologies in materials science and engineering: a survey and evaluation, in: Proceedings of the First International Workshop on Semantic Materials Science: Harnessing the Power of Semantic Web Technologies in Materials Science co-located with the 20th International Conference on Semantic Systems (SEMANTiCS 2024), volume 3760, CEUR-WS.org, 2024. URL: <https://ceur-ws.org/Vol-3760/paper3.pdf>.
- [2] 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.
- [3] T.-S. Lin, C. W. Coley, H. Mochigase, H. K. Beech, W. Wang, Z. Wang, E. Woods, S. L. Craig, J. A. Johnson, J. A. Kalow, K. F. Jensen, B. D. Olsen, BigSMILES: A Structurally-Based Line Notation for Describing Macromolecules, *ACS Central Science* 5 (2019) 1523–1531. doi:10.1021/acscentsci.9b00476.
- [4] A. M. van Herk, Q. Liu, The Importance of the Knowledge of Errors in the Measurements in the Determination of Copolymer Reactivity Ratios from Composition Data, *Macromolecular Theory and Simulations* 33 (2024) 2400043. doi:10.1002/mats.202400043.
- [5] R. Reischke, Copolymerization Reactivity Ratio Inference: Determining Confidence Contours in Parameter Space via a Bayesian Hierarchical Approach, *Macromolecular Theory and Simulations* 32 (2023) 2200063. doi:10.1002/mats.202200063.
- [6] L. Jarnac, Y. Chabot, M. Couceiro, Uncertainty management in the construction of knowledge graphs: A survey, *Transactions on Graph Data and Knowledge* 3 (2025) 3:1–3:48. doi:10.4230/tgdk.3.1.3.
- [7] D. Alomair, R. Khedri, W. MacCaull, A comprehensive review of information uncertainty modelling in domain ontologies, *Acm Computing Surveys* 58 (2026). doi:10.1145/3794841.
- [8] S. T. Bachinger, Examining the Representation of Uncertainty in Knowledge Graphs: A Case Study in Copolymer Chemistry, in: I. Celino, O. Hassanzadeh, A. Bernstein, N. Noy, G. Cheng, S. Wang, S. Ferrada, T. Soulard, K. Kozaki, H. Takeda, A. L. Gentile (Eds.), Joint Proceedings of Industry, Doctoral Consortium, Posters and Demos of the 24th International Semantic Web Conference (ISWC-C 2025), volume 4085 of *CEUR Workshop Proceedings*, CEUR, Nara, Japan, 2025, pp. 167–175. URL: <https://ceur-ws.org/Vol-4085/#paper26>.

- [9] T.-S. Lin, N. J. Rebello, H. K. Beech, Z. Wang, B. El-Zaatari, D. J. Lundberg, J. A. Johnson, J. A. Kalow, S. L. Craig, B. D. Olsen, PolyDAT: A Generic Data Schema for Polymer Characterization, *Journal of Chemical Information and Modeling* 61 (2021) 1150–1163. doi:10.1021/acs.jcim.1c00028.
- [10] R. Kumar, Materiometrically Designed Polymeric Vehicles for Nucleic Acids: Quo Vadis?, *ACS Applied Bio Materials* 5 (2022) 2507–2535. doi:10.1021/acsabm.2c00346.
- [11] J.-F. Wang, Y.-B. Sun, Q.-T. Chen, F.-F. Ji, Y.-Y. Song, M.-Y. Ruan, Y. Wang, OpenPoly: A Polymer Database Empowering Benchmarking and Multi-property Predictions, *Chinese Journal of Polymer Science (English Edition)* 43 (2025) 1749–1760. doi:10.1007/s10118-025-3402-y.
- [12] S. K. Anand, S. Kumar, Uncertainty Analysis in Ontology-Based Knowledge Representation, *New Generation Computing* 40 (2022) 339–376. doi:10.1007/s00354-022-00162-6.
- [13] M. Rospocher, S. Tonelli, L. Serafini, E. Pianta, Corpus-based terminological evaluation of ontologies, *Applied Ontology* 7 (2012) 429–448. doi:10.3233/AO-2012-0114.
- [14] C. Gavankar, Y.-F. Li, G. Ramakrishnan, Explicit Query Interpretation and Diversification for Context-Driven Concept Search Across Ontologies, in: P. Groth, E. Simperl, A. Gray, M. Sabou, M. Krötzsch, F. Lecue, F. Flöck, Y. Gil (Eds.), *The Semantic Web – ISWC 2016*, Springer International Publishing, Cham, 2016, pp. 271–288. doi:10.1007/978-3-319-46523-4_17.
- [15] H. Beygi Nasrabadi, E. Norouzi, H. Sack, B. Skrotzki, Performance Evaluation of Upper-Level Ontologies in Developing Materials Science Ontologies and Knowledge Graphs, *Advanced Engineering Materials* 27 (2025) 2401534. doi:10.1002/adem.202401534.
- [16] 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.
- [17] S. T. Tüzün, Code for SeMatS 2026 paper: The State of Uncertainty Representation in Copolymer Chemistry Knowledge Artifacts, 2026. doi:10.5281/zenodo.21517739.
- [18] ASReview LAB developers, ASReview LAB - A tool for AI-assisted systematic reviews, 2026. doi:10.5281/zenodo.19663843.
- [19] R. Campos, V. Mangaravite, A. Pasquali, A. M. Jorge, C. Nunes, A. Jatowt, YAKE! Collection-Independent Automatic Keyword Extractor, in: G. Pasi, B. Piwowarski, L. Azzopardi, A. Hanbury (Eds.), *Advances in Information Retrieval*, Springer International Publishing, Cham, 2018, pp. 806–810. doi:10.1007/978-3-319-76941-7_80.
- [20] S. Rose, D. Engel, N. Cramer, W. Cowley, Automatic Keyword Extraction from Individual Documents, John Wiley & Sons, Ltd, 2010, pp. 1–20. doi:https://doi.org/10.1002/9780470689646.ch1.
- [21] M. Grootendorst, Bertopic: Neural topic modeling with a class-based tf-idf procedure, *arXiv preprint arXiv:2203.05794* (2022).
- [22] 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://elib.dlr.de/206770/>.
- [23] A. Malik, M. Arsalan, C. Moreno, J. Mosquera, E. Félix, T. Kizilören, V. Muthukrishnan, B. Zdrazil, A. R. Leach, N. M. O’Boyle, ChEBI: re-engineered for a sustainable future, *Nucleic Acids Research* 54 (2026) D1768–D1778. doi:10.1093/nar/gkaf1271.
- [24] B. P. Rajamohan, A. C. H. Bradley, V. D. Tran, J. E. Gordon, H. W. Caldwell, R. Mehdi, G. Ponon, Q. D. Tran, O. Dernek, J. Kaltenbaugh, B. G. Pierce, R. Wieser, W. Yue, K. Lin, J. Kambo, C. Lopez, A. Nihar, D. J. Savage, D. W. Brown, H. Sharma, B. Giera, P. K. Tripathi, Y. Wu, M. Li, K. O. Davis, L. S. Bruckman, E. I. Barcelos, R. H. French, Materials Data Science Ontology(MDS-Onto): Unifying Domain Knowledge in Materials and Applied Data Science, *Scientific Data* 12 (2025) 628. doi:10.1038/s41597-025-04938-5.
- [25] B. P. Rajamohan, A. C. H. Bradley, V. D. Tran, J. E. Gordon, H. W. Caldwell, R. Mehdi, G. Ponon, Q. D. Tran, O. Dernek, J. Kaltenbaugh, B. G. Pierce, R. Wieser, W. Yue, K. Lin, J. Kambo, C. Lopez, A. Nihar, D. J. Savage, D. W. Brown, H. Sharma, B. Giera, P. K. Tripathi, Y. Wu, M. Li, K. O. Davis, L. S. Bruckman, E. I. Barcelos, R. H. French, Materials Data Science Ontology, 2025. URL:

https://cwrusdle.bitbucket.io/files/MDS_Onto/index-en.html, last accessed: 2026-07-29.

- [26] T. Yoshimura, H. Kato, S. Oikawa, T. Inagaki, S. Asano, T. Sugawara, T. Miyao, T. Matsubara, H. Ajiro, M. Fujii, Y.-y. Ohnishi, M. Hatanaka, CopDDB: a descriptor database for copolymers and its applications to machine learning, *Digital Discovery* 4 (2025) 195–203. doi:10.1039/d4dd00266k.
- [27] N. J. Rebello, A. Arora, H. Mochigase, T.-S. Lin, J. Shi, D. J. Audus, E. S. Muckley, A. Osmani, B. D. Olsen, The Block Copolymer Phase Behavior Database, *Journal of Chemical Information and Modeling* 64 (2024) 6464–6476. doi:10.1021/acs.jcim.4c00242.
- [28] M. Ishii, T. Ito, K. Sakamoto, Nims polymer database polyinfo (ii): machine-readable standardization of polymer knowledge expression, *Science and Technology of Advanced Materials: Methods* 4 (2024) 2354651. doi:10.1080/27660400.2024.2354651.
- [29] M. Ishii, T. Ito, H. Sado, I. Kuwajima, NIMS polymer database PoLyInfo (I): an overarching view of half a million data points, *Science and Technology of Advanced Materials: Methods* 4 (2024) 2354649. doi:10.1080/27660400.2024.2354649.
- [30] M. Ishii, K. Sakamoto, PoLyInfo Ontology Version 1.2, 2025. URL: <https://mdr.nims.go.jp/datasets/b5fb3e53-cc31-4775-8e9a-33ccc7329dca>, last accessed: 2026-07-29.
- [31] M. Dembska, M. Held, S. Schindler, PolyMat ontology, 2024. doi:10.5281/ZENODO.10286389, last accessed: 2026-07-29.
- [32] N. S. Schausser, G. A. Kliegle, P. Cooke, R. A. Segalman, R. Seshadri, Database Creation, Visualization, and Statistical Learning for Polymer Li+-Electrolyte Design, *Chemistry of Materials* 33 (2021) 4863–4876. doi:10.1021/acs.chemmater.0c04767.
- [33] J. P. McCusker, N. Keshan, S. Rashid, M. Deagen, C. Brinson, D. L. McGuinness, NanoMine: A Knowledge Graph for Nanocomposite Materials Science, in: J. Z. Pan, V. Tamma, C. d’Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne, L. Kagal (Eds.), *The Semantic Web – ISWC 2020*, Springer International Publishing, Cham, 2020, pp. 144–159. doi:10.1007/978-3-030-62466-8_10.
- [34] tetherless-world, Nanomine-ontology/nanomine.ttl at master · tetherless-world/nanomine-ontology, 2018. URL: <https://github.com/tetherless-world/nanomine-ontology/blob/master/nanomine.ttl>, last accessed: 2026-07-29.
- [35] D. S. Levine, N. Liesen, L. Chua, J. Diffenderfer, H. Ingolfsson, M. P. Kroonblawd, N. Kumar, A. Maiti, S. S. Mohottalalage, M. Shuaibi, B. V. Essen, B. M. Wood, C. L. Zitnick, S. M. Blau, E. R. Antoniuk, The Open Polymers 2026 (OPoly26) Dataset and Evaluations, 2026. doi:10.48550/arXiv.2512.23117, arXiv:2512.23117 [physics.chem-ph].
- [36] O. Hakimi, J. L. Gelpi, M. Krallinger, F. Curi, D. Repchevsky, M.-P. Ginebra, The Devices, Experimental Scaffolds, and Biomaterials Ontology (DEB): A Tool for Mapping, Annotation, and Analysis of Biomaterials Data, *Advanced Functional Materials* 30 (2020) 1909910. doi:10.1002/adfm.201909910.
- [37] ProjectDebbie, Ontology_DEB/DEB_ont.owl at master · ProjectDebbie/Ontology_DEB, 2021. URL: https://github.com/ProjectDebbie/Ontology_DEB/blob/master/DEB_ont.owl, last accessed: 2026-07-29.
- [38] K. Degtyarenko, P. de Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcántara, M. Darsow, M. Guedj, M. Ashburner, ChEBI: a database and ontology for chemical entities of biological interest, *Nucleic Acids Research* 36 (2008) D344–D350. doi:10.1093/nar/gkm791.
- [39] Index of /pub/databases/chebi/ontology, 2026. URL: <https://ftp.ebi.ac.uk/pub/databases/chebi/ontology/>, last accessed: 2026-07-29.
- [40] C. Kim, A. Chandrasekaran, T. D. Huan, D. Das, R. Ramprasad, Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions, *The Journal of Physical Chemistry C* 122 (2018) 17575–17585. doi:10.1021/acs.jpcc.8b02913.
- [41] D. J. Walsh, W. Zou, L. Schneider, R. Mello, M. E. Deagen, J. Mysona, T.-S. Lin, J. J. de Pablo, K. F. Jensen, D. J. Audus, B. D. Olsen, Community Resource for Innovation in Polymer Technology (CRIPT): A Scalable Polymer Material Data Structure, *ACS Central Science* 9 (2023) 330–338. doi:10.1021/acscentsci.3c00011.
- [42] K.-i. Takahashi, H. Mamitsuka, M. Tosaka, N. Zhu, S. Yamago, CoPolDB: a copolymeriza-

tion database for radical polymerization, *Polymer Chemistry* 15 (2024) 965–971. doi:10.1039/D3PY01372C.

- [43] A. Medina-Smith, C. A. Becker, R. L. Plante, L. M. Bartolo, A. Dima, J. A. Warren, R. J. Hanisch, A Controlled Vocabulary and Metadata Schema for Materials Science Data Discovery, *Data Science Journal* 20 (2021) 18. doi:10.5334/dsj-2021-018.
- [44] M. Dumontier, C. J. Baker, J. Baran, A. Callahan, L. Chepelev, J. Cruz-Toledo, N. R. Del Rio, G. Duck, L. I. Furlong, N. Keath, D. Klassen, J. P. McCusker, N. Queralt-Rosinach, M. Samwald, N. Villanueva-Rosales, M. D. Wilkinson, R. Hoehndorf, The SemanticScience Integrated Ontology (SIO) for biomedical research and knowledge discovery, *Journal of Biomedical Semantics* 5 (2014) 14. doi:10.1186/2041-1480-5-14.
- [45] G. V. Gkoutos, P. N. Schofield, R. Hoehndorf, The Units Ontology: a tool for integrating units of measurement in science, *Database* 2012 (2012) bas033. doi:10.1093/database/bas033.
- [46] 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.