Examining the Representation of Uncertainty in Knowledge Graphs: A Case Study in Copolymer Chemistry

Sarah T. Bachinger[1]

[1]*Heinz Nixdorf Chair for Distributed Information Systems, Friedrich Schiller University Jena, Leutragraben 1, 07743 Jena, Germany*

Abstract

In copolymer chemistry, researchers aim to create molecules with desirable properties. The experimental process towards a successful synthesis inherently contains uncertainties in different forms and can be described as a structured process. Using ontologies to describe this process could help in predicting the synthesis setup and conserve resources. Therefore in this thesis project, we want to examine how the representation of different uncertainties in the ontology can increase the prediction success. Guided by five research questions, we detail our approach consisting of expert interviews, mapping of uncertainties from the interviews to known representations on Knowledge Graphs (KGs) and its evaluation.

Keywords

Uncertain Knowledge Graphs, Experimental Uncertainty, Uncertainty Representation

1. Problem statement and importance

Copolymer chemistry, a subfield of material science, investigates the creation (synthesis) and properties of copolymers, macromolecules consisting of different monomers. These are widely used in the industry, for example, as packaging materials [1, 2]. Chemists try to synthesize molecules that display desirable properties such as biodegradability by consulting different knowledge sources to create a synthesis plan and improving upon the past experiments (see Figure 1).

A main reason for the integration of methods from computer science with chemistry is the prediction of reactions to conserve resources and focus on optimizing the chosen property [3]. To describe these procedures and molecules in a formal model that can be then used for prediction regarding experimental outcomes, ontologies are a suitable choice [4]. Several ontologies already exist in the field of copolymer chemistry, with some describing a range of polymers including copolymers [5, 6], while others are more concerned with the experimental setup [7].

As in any other experimental domain, chemistry deals with uncertainty and incomplete data from experiments. For reliable predictions regarding the experimental setup or desirable properties, uncertainties need to be figured into these calculations and ontologies need to be able to handle the representation of those uncertainties, which is an active area of research [8, 9]. This is the starting point for this thesis, where we want to explore how ontologies as formal models can support the scientific discovery process in the natural sciences when uncertainty and incomplete data are factored into the representation and reasoning process using Knowledge Graphs (KGs) [10] as a graph-based representation. To achieve this, we apply our research to the domain of copolymer chemistry because of the big potential impact of enhancing experiment prediction (see Figure 1).

The remainder of this proposal is structured as follows: In Section 2, we list related work on (1) the term uncertainty, how (2) uncertainty is represented in KGs and (3) ontologies and data in the copolymer chemistry domain. We then motivate our research questions and hypothesis in Section 3. Preliminary results are then reported in Section 4, while we discuss the plans for the evaluation in Section 5. Section 6 concludes this proposal with reflections and next steps.

ISWC 2025 Companion Volume, November 2–6, 2025, Nara, Japan

sarah.bachinger@uni-jena.de (S. T. Bachinger)
https://www.fmi.uni-jena.de/en/14619/sarah-t-bachinger (S. T. Bachinger)
0009-0005-5422-2164 (S. T. Bachinger)

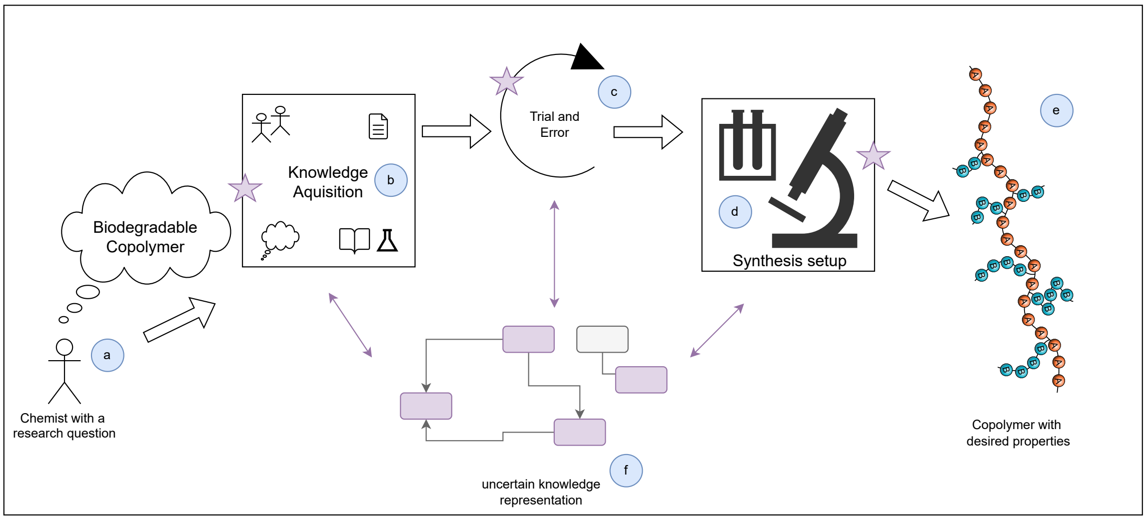

Figure 1: The workflow in the synthesis of a copolymer: the starting point of the research process is (a) the chemist with a research question, e.g., *How can I create a copolymer with the property "biodegradability"?* There are three key points in the process: Knowledge Acquisition, the Trial and Error Experimentation process, and the resulting Synthesis setup. After posing the research question, the chemist will consult (b) a plethora of different sources of knowledge like related literature, colleagues or their (electronic) laboratory notebooks to find a good starting point. Then, they use the gathered knowledge to perform an experiment in a (c) trial and error process to find a suitable combination of input materials and equipment that result in a (d) synthesis setup that produces the (e) copolymer with the desired properties. In these three steps, there are inherent uncertainties (pink stars) that influence the outcome of experiments. We propose to enhance this process by the (f) (re-)use of existing ontologies with a novel uncertainty representation approach to predict the synthesis setup and thereby conserve time and resources in the experimental process.

2. Related work

During a preliminary literature review, we focused on different definitions of the term uncertainty and its representation in Knowledge Graphs, and ontologies and data sources in the copolymer chemistry domain.

2.1. Definitions of Uncertainty

What researchers view as uncertainty varies in different scientific domains. For example, Ningrum et al. [11] discuss this and list different definitions of uncertainty in their study on scientific uncertainty in scholarly articles.

In the final report of the W3C Incubator Group for uncertainty reasoning for the World Wide Web, the authors define uncertainty as "a variety of aspects of imperfect knowledge, including incompleteness, inconclusiveness, vagueness, ambiguity, and others" [12].

As a starting point, uncertainty in this thesis refers in relation to the above mentioned W3C Incubator Group definition to everything that hinders the prediction of the experimental outcome including, among others, statistical, ethical, societal or experimental factors. We expect this definition to adapt and to be refined as we continue the research process.

2.2. Uncertainty in Knowledge Graphs

Jarnac et al. [13] classify data from different sources with conflicting information in the construction of a KG into two groups of so-called knowledge deltas and differentiate six types of conflicting data: invalidity, vagueness, fuzziness, timeliness, ambiguity, and incompleteness.

In 2008, the final report of an W3C Incubator Group [12] on uncertainty reasoning for the world wide web was released with two goals, namely the identification of world wide web challenges that would benefit from uncertainty reasoning and building on that what techniques could be used to improve and standardize the same. In the report, they describe an uncertainty ontology and list common approaches to model uncertainty in the World Wide Web. We will apply and extend their work for our experimental domain.

Recent work on the combination of KGs with uncertainty includes Ni et al. [14] who propose a multi-hop reasoning framework combining LLMs with KGs to address uncertainty estimation in LLM-KG systems. Deng et al. [15] explore the creation of artificial intelligence systems based on structured knowledge for systems such as generative AI and investigate the handling of uncertainty. Yang et al. [16] review KG reliability from the perspective of knowledge correctness and uncertainty. Building upon knowledge representation learning, they compare score function modification, representation vector optimization, loss function adjustment, and textual information integration. Freedman et al. [17] combine Bayesian Probabilism with existing KGs to transform them to what they call probabilistic KGs.

2.3. Copolymer Data and Ontologies

Currently available datasets for copolymer chemistry include the **polyOne** dataset [18] that contains 100 million hypothetical polymers along with several predicted parameters and **PolyInfo** [19] which was introduced in 2011 and recently examined in ”NIMS polymer database PoLyInfo (I): an overarching view of half a million data points” [20]. The purpose was the collection of different polymers and related information on their structure and methods of processing including measurements, properties, and monomers. The **PoLyInfo Knowledge Collection** [6] consists among other things of the PoLyInfo ontology and aim to systematize polymer chemistry and provide a ”machine-readable understanding of the polymer knowledge contained in “PoLyInfoRDF” [21].

In the paper ”NIMS polymer database PoLyInfo (II): machine-readable standardization of polymer knowledge expression” [22], Ishii et al. introduce the management of polymers including copolymers, where they use shape expression language and used IDs to refer to different variational parts of the copolymer, such as the repeating unit.

With **PolyNERE** [5], the authors created an ontology for polymers, related entities and relationships and a corpus of abstracts annotated by the different features.

Conjugated Polymer Process Ontology [7] is used for the design of experiments for organic field-effect transistors and aim to provide FAIR data for experiments.

NanoMine [23] aims to provide an open source data resource for polymer nanocomposites[1] serving as an open source data resource.

Other ontologies that include copolymers as concepts include the **Devices, Experimental scaffolds and Biomaterials Ontology** [24] and the **Materials Data Vocabulary** [25, 26].

PolyMAT [27] is an ontology for polymer membrane research that puts a focus on being used in electronic laboratory notebooks. We find that especially PolyNERE and PolyInfo as extensive sources for polymers including copolymers could serve as interesting applications to examine the occurrence of missing, incomplete or uncertain data.

3. Research questions and hypotheses

Using the domain of copolymer chemistry, we want to examine how uncertainty in different aspects of the experimentation process can be represented in KGs. Therefore, we pose the following research questions (RQs).

- RQ1: What does the term ”Uncertainty” refer to in the experimental sciences?
- RQ2: What factors (domain-specific uncertainties) influence the outcome of experiments in a domain such as copolymer chemistry?
- RQ3: To what extent are existing uncertainty representation approaches able to represent the specific uncertainties in the experimental sciences, in particular copolymer chemistry?

[1]https://qa.materialsmine.org/nm/

- RQ4: How can approaches that represent uncertainty in KGs be applied to predict the outcome of experiments?
- RQ5: If the exploration of the previous RQs shows that there is a need for improvement: How can a novel uncertainty representation approach with regards to the specific needs for the copolymer chemistry domain be designed?

We operate under the assumption that there are certain types of uncertainty in research that are not yet (well) represented in KGs as a medium for connecting information on different domains. However, if these uncertainties were represented, this would be helpful for domain experts because the modeled knowledge can then be utilized for predicting the outcome of experiments. As an application domain, copolymer chemistry was chosen for the big potential improvement areas. To explore the above mentioned research questions, we pose the following hypotheses with each hypothesis corresponds to the same numbered research question:

1. Hypothesis 1: The definition of uncertainty depends on the scientific domain.
    - Surveying taxonomies and definitions of uncertainties
    - Exploratory expert interviews with ten researchers from different scientific domains
2. Hypothesis 2: There are certain factors in experimental domains that hinder the prediction of experimental outcomes.
    - Domain-specific expert interviews with five experts
    - Conducting a requirement analysis for the prediction of experiment outcomes
    - Decision on metrics for the evaluation
    - Establishing a use case for the remainder of the thesis
3. Hypothesis 3: Current approaches are not yet able to represent all of the specific uncertainties in experimental sciences.
    - Mapping between the state-of-the-art and the the discovered uncertainty types from the interviews from H1 and H2
    - Literature review on the representation of uncertainty in ontologies and specifically in Knowledge Graphs
    - Literature review on Knowledge Graphs in copolymer chemistry
4. Hypothesis 4: A representation design with uncertainty in mind would enhance reasoning, and therefore have a practical impact.
    - Comparative analysis of the different representation approaches on the chosen use case
5. Hypothesis 5: A new specific representation approach and accordingly reasoning abilities improve upon the chosen use case.
    - Development of a novel representation approach
    - Evaluation and comparison of the new representation approach with the state-of-the-art representations

4. Preliminary results

From January to May 2025, 10 interviews were conducted with scientists from different domains about uncertainty in their research. This sets the foundation for further exploration and more detailed questioning in the future. We use the sociological method guideline based expert interviews [28] to answer hypothesis 1 to access domain expert knowledge to uncover areas of insufficient data uncertainty coverage. We designed a guideline encompassing the following questions to be as open as possible:

- What kind of uncertainty do you deal with in your research?
- What kind of uncertainty occurs in your data?

Domain	Career-level	participants
Ecology	PhD Candidate, Postdocs, (assistant) professors	7
Copolymer Chemistry	PhD Candidate	2
Sociology	PhD Candidate	1

Table 1
Interviewee distribution across domains and career-levels

- How do you deal with that (referring to the occurring uncertainty)?
- How is this (referring to both the uncertainty and its handling) represented for future reuse?

We experimented with including the question "How do you define uncertainty?" at the beginning, but found that it narrows the participants' answers to mostly statistical uncertainty.

For a standardized transcription process, we use the seven working steps outlined by Kuckart [29]. First, to ensure comparative transcription, we adapted the transcription rules from [29]. Second, the interviews were transcribed using the software "noScribe" [30] in version 0.6.1 and manually corrected. Currently, we are redacting personal information from the interviews and will use pseudonymized transcripts for the collection of uncertainty types.

5. Evaluation

In this section, we want to connect the research questions and hypotheses from Section 3 with concrete evaluation steps and expected outcomes. We started from a very generic definition of uncertainty as everything that hinders the prediction of the experimental outcome including, among others, statistical, ethical, societal or experimental factors. To operate under a more grounded definition, we pose the first research questions as *RQ1: What does the term "Uncertainty" refer to in the experimental sciences?* To answer it, we will survey definitions from different disciplines like mathematics, philosophy, sociology and the experimental sciences. We conduct interviews with ten researchers from different scientific domains to explore the notion of uncertainty in their domain. As outcome, we expect to obtain diverse definitions of the term uncertainty as well as examples of uncertainty in different scientific domains.

At the same time, we will look more specifically at our application domain with the second research question *RQ2: What factors influence the outcome of experiments in a domain like copolymer chemistry?* to extract uncertainties specific to copolymer chemistry. For example, incomplete or missing data, representations of copolymers, or distributions of values come to mind. We plan to conduct interviews with five domain experts about uncertainties in the experimentation process which will also help us to establish a use case for the remainder of the thesis. In consideration of the domain-specific challenges, we conduct a requirement analysis to extract the specific requirements for the prediction of the outcome of an experiment and building on that metrics for the evaluation of a knowledge representation approach with reasoning including uncertainty.

After exploring RQ1 and RQ2, we can compare existing representation methods on our specific use case and therefore ask *RQ3: To what extent are existing uncertainty representation approaches able to represent the specific uncertainties in the experimental sciences, in particular copolymer chemistry?* In a first step, we will map the uncertainties from the interviews to the definitions. This will enable us to make grounded assumptions in how to represent these uncertainties as we can use established methods for the application domain specific use case. For a nuanced comparison, we need to review existing ontologies in the copolymer domain. An extensive review on the coverage for the application area copolymer science needs to include both domain specific entities like the monomer types as well as information about the license and reusability. Second, we will further analyze the existing ontologies in different aspects. Investigations include the domain area, whether they are actively used in the chemistry domain and if the described domain includes data that are uncertain. Furthermore, with the mapping of domain-specific uncertainties (RQ2) to definitions (RQ1), we can evaluate the coverage of different types of uncertainty with existing representation and reasoning approaches for KGs. We will use these insights in the decision whether a new representation method is needed if a certain type of uncertainty is not well represented.

To report on *RQ4: How can approaches that represent uncertainty in KGs be applied to predict the outcome of experiments?*, we will compare the performance of different state-of-the-art approaches regarding our chosen use case and the evaluation criteria established in H2. With this comparison, we can (1) evaluate the strengths and drawbacks of each approach and (2) judge whether all aspects of the use case are appropriately covered.

Our last research question *RQ5: If the exploration of the previous RQs shows that there is a need for improvement: How can a novel uncertainty representation approach with regards to the specific needs for the copolymer chemistry domain be designed?* uses the insights from RQ3 and RQ4 to develop a novel representation approach. We will use the comparison from RQ4 as a baseline for the novel approach.

Continuous working packages are (1) keeping up-to-date on the current literature regarding uncertainty in KGs and ontologies for the copolymer domain and (2) thesis writing. We opt for writing and researching in parallel instead of consecutive research and writing phases. Even though this thesis is not planned as a cumulative dissertation, we plan to submit milestones to suitable conferences and use the published results in the thesis.

6. Reflection and future work

In conclusion, we laid out how we plan to investigate the representation of uncertainty in KGs and how we apply this to the copolymer chemistry domain. We will now reflect on the main steps and illustrate future work.

6.1. Expert Interviews

As reported in Section 4, we have conducted ten exploratory interviews. The next steps are pseudonymization and coding to cluster different types of uncertainty. For the copolymer chemists, only two PhD level candidates were interviewed. Further interviews with senior scientists or industry specialists are necessary to ensure a complete picture. With the start of a new graduate school focused on the intersection of copolymer chemistry and computer science (COIN)[2], a new pool of potential interviewees of different levels of expertise are available. Furthermore, on the basis of the exploratory interviews, more directed questioning is possible as we plan to use the interviews we have to develop an enhanced and more focused questionnaire for the domain of copolymer chemistry. As a guideline, we aim for five interview partners that hold a PhD in copolymer chemistry and use those interviews to further clarify the uncertainty types. Furthermore, we aim to extract at least two types of synthesis as a use case for the thesis. For RQ1 and RQ2, we are dependent on the availability of experts and their willingness to participate in multiple iterations as the use case developments and we classify uncertainties.

6.2. Literature Review

Though literature review is an ongoing process, we identified several areas like uncertainty representation in KGs and copolymer ontologies where an in-depth review will be beneficial. This includes the topic of uncertainty in different scientific fields to see the state-of-the-art in uncertainty taxonomies. We plan to map the insights from the interviews to current uncertainty representation methods in Knowledge Graphs to analyze where gaps in the current research exist.

6.3. Development of a novel representation approach

Generative AI is transforming scientific practice as well as all scientific domains, as recent examples show [31, 32]. Works such as [33, 34] explore the current abilities and drawbacks of using Large

[2]https://www.coin.uni-jena.de/

Language Models for Material Science. For example, Miret and Krishnan [33] point out several future challenges, one of them being the development of domain-specific reasoning techniques that implement the principles of the domain, namely material science. While this is still in the early stages, the combination of LLMs and ontologies for representing the uncertainty in copolymer chemistry might be an interesting research direction for this thesis.

Acknowledgments

This thesis is funded by a doctoral scholarship by the State of Thuringia and supported by the Friedrich Schiller University of Jena, the ZiF group ”Mapping Evidence to Theory in Ecology” 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). It is supervised by Prof. Dr. Birgitta König-Ries.

Declaration on Generative AI

During the preparation of this work, the author used Writefull in order to: Grammar and spelling check. After using this service, the author reviewed and edited the content as needed and takes full responsibility for the publication’s content.

References

[1] D. Feldman, Polymer History, Designed Monomers and Polymers 11 (2008) 1–15. doi:10.1163/156855508X292383.
[2] N. A. Lynd, A. J. Meuler, M. A. Hillmyer, Polydispersity and block copolymer self-assembly, Progress in Polymer Science 33 (2008) 875–893. doi:10.1016/j.progpolymsci.2008.07.003.
[3] J. L. Tyler, D. Trauner, F. Glorius, Reaction development: A student’s checklist, Chemical Society Reviews 54 (2025) 3272–3292. doi:10.1039/D4CS01046A.
[4] 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.
[5] V.-T. Phi, H. Teranishi, Y. Matsumoto, H. Oka, M. Ishii, PolyNERE: A Novel Ontology and Corpus for Named Entity Recognition and Relation Extraction in Polymer Science Domain, in: N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, N. Xue (Eds.), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), ELRA and ICCL, Torino, Italia, 2024, pp. 12856–12866.
[6] M. Ishii, K. Sakamoto, PoLyInfo Knowledge Collection, 2024. doi:10.48505/NIMS.4413.
[7] A. L. Liu, M. Lee, R. Venkatesh, J. A. Bonsu, R. Volkovinsky, J. C. Meredith, E. Reichmanis, M. A. Grover, Conjugated Polymer Process Ontology and Experimental Data Repository for Organic Field-Effect Transistors, Chemistry of Materials 35 (2023) 8816–8826. doi:10.1021/acs.chemmater.3c01842, doi: 10.1021/acs.chemmater.3c01842.
[8] L. Jarnac, Y. Chabot, M. Couceiro, Uncertainty Management in the Construction of Knowledge Graphs: A Survey, 2024. doi:10.48550/arXiv.2405.16929. arXiv:2405.16929.
[9] A. Hur, N. Janjua, M. Ahmed, A Survey on State-of-the-art Techniques for Knowledge Graphs Construction and Challenges ahead, in: 2021 IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), 2021, pp. 99–103. doi:10.1109/AIKE52691.2021.00021.
[10] A. Hogan, E. Blomqvist, M. Cochez, C. d’Amato, G. de Melo, Claudio Gutierrez, Sabrina Kirrane, J. E. Labra Gayo, Roberto Navigli, Sebastian Neumaier, A.-C. Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, Antoine Zimmermann,

Knowledge Graphs, number 22 in Synthesis Lectures on Data, Semantics, and Knowledge, Springer, 2021.
[11] P. K. Ningrum, P. Mayr, N. Smirnova, I. Atanassova, Annotating Scientific Uncertainty: A comprehensive model using linguistic patterns and comparison with existing approaches, Journal of Informetrics 19 (2025) 101661. doi:10.1016/j.joi.2025.101661. arXiv:2503.11376, comment: Paper Accepted for Publication in the Journal of Informetrics (2025)
Compares different notions of uncertainty.
[12] Kenneth J. Laskey, Kathryn B. Laskey, Paulo C. G. Costa, Mieczyslaw M. Kokar, Trevor Martin, Thomas Lukasiewicz, Uncertainty Reasoning for the World Wide Web, W3C Incubator Group Report, 2008. URL: http://www.w3.org/2005/Incubator/urw3/XGR-urw3/.
[13] 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. URL: https://drops.dagstuhl.de/storage/08tgdk/tgdk-vol003/tgdk-vol003-issue001/TGDK.3.1.3/TGDK.3.1.3.pdf. doi:10.4230/tgdk.3.1.3, publisher: Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik.
[14] B. Ni, Y. Wang, L. Cheng, E. Blasch, T. Derr, Towards Trustworthy Knowledge Graph Reasoning: An Uncertainty Aware Perspective, Proceedings of the AAAI Conference on Artificial Intelligence 39 (2025) 12417–12425. doi:10.1609/aaai.v39i12.33353.
[15] S. Deng, N. Zhang, T. Wu, From Certainty to Uncertainty in Knowledge: Exploring Modeling, Extraction, Representation, and Applications, in: Handbook on Neurosymbolic AI and Knowledge Graphs, IOS Press, 2025, pp. 319–362. doi:10.3233/FAIA250214.
[16] Y. Yang, J. Chen, Y. Xiang, A review on the reliability of knowledge graph: From a knowledge representation learning perspective, World Wide Web 28 (2024) 4. doi:10.1007/s11280-024-01316-w.
[17] H. Freedman, N. Abolhassani, J. Metzger, S. Paul, Ontology Modeling for Probabilistic Knowledge Graphs, in: 2023 IEEE 17th International Conference on Semantic Computing (ICSC), 2023, pp. 252–259. doi:10.1109/ICSC56153.2023.00049, the paper to the medium article.
[18] C. Kuenneth, R. Ramprasad, polyOne Data Set - 100 million hypothetical polymers including 29 properties, 2022. doi:10.5281/zenodo.7766806.
[19] S. Otsuka, I. Kuwajima, J. Hosoya, Y. Xu, M. Yamazaki, PoLyInfo: Polymer Database for Polymeric Materials Design, in: 2011 International Conference on Emerging Intelligent Data and Web Technologies, 2011, pp. 22–29. doi:10.1109/EIDWT.2011.13.
[20] 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.
[21] M. Ishii, T. Takemura, M. Tanifuji, PoLyInfo RDF: A Semantically Reinforced Polymer Database for Materials Informatics, in: International Workshop on the Semantic Web, 2019.
[22] 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.
[23] V. Rawte, J. P. McCusker, H. Zhao, L. C. Brinson, W. Chen, L. Schadler, D. L. McGuinness, An Ontology for a Polymer Nanocomposite Community Data Resource, in: Proceedings of the 2017 ACM on Web Science Conference, WebSci '17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 411–412. doi:10.1145/3091478.3098866.
[24] 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.
[25] 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.
[26] A. Medina-Smith, C. Becker, Simple Knowledge Organization System (SKOS) version of Materials Data Vocabulary, 2017. doi:10.18434/T4/1435037.
[27] M. Dembska, M. Held, S. Schindler, PolyMat ontology, 2024. doi:10.5281/ZENODO.10286389.

[28] C. Helfferich, Leitfaden- und Experteninterviews, in: N. Baur, J. Blasius (Eds.), Handbuch Methoden der empirischen Sozialforschung, Springer Fachmedien, Wiesbaden, 2022, pp. 875–892. doi:10.1007/978-3-658-37985-8_55.
[29] U. Kuckartz, Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung, 2018.
[30] K. Dröge, Kaixxx/noScribe, 2025. URL: https://github.com/kaixxx/noScribe.
[31] K. M. Jablonka, Q. Ai, A. Al-Feghali, S. Badhwar, J. D. Bocarsly, A. M. Bran, S. Bringuier, L. C. Brinson, K. Choudhary, D. Circi, S. Cox, W. A. de Jong, M. L. Evans, N. Gastellu, J. Genzling, M. V. Gil, A. K. Gupta, Z. Hong, A. Imran, S. Kruschwitz, A. Labarre, J. Lála, T. Liu, S. Ma, S. Majumdar, G. W. Merz, N. Moitessier, E. Moubarak, B. Mouriño, B. Pelkie, M. Pieler, M. C. Ramos, B. Ranković, S. G. Rodriques, J. N. Sanders, P. Schwaller, M. Schwarting, J. Shi, B. Smit, B. E. Smith, J. V. Herck, C. Völker, L. Ward, S. Warren, B. Weiser, S. Zhang, X. Zhang, G. A. Zia, A. Scourtas, K. J. Schmidt, I. Foster, A. D. White, B. Blaiszik, 14 examples of how LLMs can transform materials science and chemistry: A reflection on a large language model hackathon, Digital Discovery 2 (2023) 1233–1250. doi:10.1039/D3DD00113J.
[32] J. V. Herck, M. V. Gil, K. M. Jablonka, A. Abrudan, A. S. Anker, M. Asgari, B. Blaiszik, A. Buffo, L. Choudhury, C. Corminboeuf, H. Daglar, A. M. Elahi, I. T. Foster, S. Garcia, M. Garvin, G. Godin, L. L. Good, J. Gu, N. X. Hu, X. Jin, T. Junkers, S. Keskin, T. P. J. Knowles, R. Laplaza, M. Lessona, S. Majumdar, H. Mashhadimoslem, R. D. McIntosh, S. M. Moosavi, B. Mouriño, F. Nerli, C. Pevida, N. Poudineh, M. Rajabi-Kochi, K. L. Saar, F. H. Saboor, M. Sagharichiha, K. J. Schmidt, J. Shi, E. Simone, D. Svatunek, M. Taddei, I. Tetko, D. Tolnai, S. Vahdatifar, J. Whitmer, D. C. F. Wieland, R. Willumeit-Römer, A. Züttel, B. Smit, Assessment of fine-tuned large language models for real-world chemistry and material science applications, Chemical Science 16 (2025) 670–684. doi:10.1039/D4SC04401K.
[33] S. Miret, N. M. A. Krishnan, Enabling large language models for real-world materials discovery, Nature Machine Intelligence 7 (2025) 991–998. doi:10.1038/s42256-025-01058-y.
[34] E. O. Pyzer-Knapp, M. Manica, P. Staar, L. Morin, P. Ruch, T. Laino, J. R. Smith, A. Curioni, Foundation models for materials discovery – current state and future directions, npj Computational Materials 11 (2025) 61. doi:10.1038/s41524-025-01538-0.