

Implementing semantic technologies in materials science and engineering

Marta Dembska^{1,*†}, Oliver Helle^{2,*†}, Itisha Yadav^{1,*†} and Diana Peters^{1,*†}

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

²German Aerospace Center (DLR), Institute of Materials Research, Cologne, Germany

Abstract

The Materials Science and Engineering (MSE) field is an interdisciplinary branch of engineering characterized by high volumes of heterogeneous data, which also serve as inputs for other fields reliant on materials. While semantic technologies are already utilized in various research areas to address data management challenges, their adoption in MSE is still in its early stages. This paper provides an overview of data management issues in MSE, existing semantic technologies, and the potential application of these technologies to address those issues. This leads to a roadmap for the implementation of semantic technologies in MSE.

Keywords

Materials Science and Engineering, Semantic Technologies, Data Management, Interoperability, Ontologies, Large Language Models

1. Introduction

Science generates vast amounts of data. Before investing effort and money in the generation of new datasets (e.g. by conducting experiments), it would be useful to find existing datasets that already address the scientific question at hand. However, the sheer volume of data makes this a daunting task, akin to searching for a needle in a haystack. Furthermore, the use of different schemas to encode and describe data complicates the search process.

To address this, the FAIR principles (Findable, Accessible, Interoperable, Reusable) were proposed to enhance data sustainability [1]. Semantic Web technologies are the primary means of implementing these principles, although their application varies significantly across scientific domains. In this paper, we focus on the use of semantic technologies in the MSE domain, explore the reasons why adoption should increase, and how this can be achieved effectively.

MSE operates within a data-centric sphere, generating diverse datasets aimed at advancing manufacturing technologies and detailing material structures along with their relevant

SeMatS 2024: The 1st International Workshop on Semantic Materials Science co-located with the 20th International Conference on Semantic Systems (SEMANTiCS), September 17-19, Amsterdam, The Netherlands.

*Corresponding author.

†These authors contributed equally.

✉ marta.dembska@dlr.de (M. Dembska); oliver.helle@dlr.de (O. Helle); itisha.yadav@dlr.de (I. Yadav); diana.peters@dlr.de (D. Peters)

ORCID 0000-0002-8180-1525 (M. Dembska); 0000-0001-9198-3900 (O. Helle); 0009-0002-7957-9600 (I. Yadav); 0000-0002-5855-2989 (D. Peters)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

parameters. The sharing of this data, however, remains a challenge due to a lack of documentation, of standardized usage of Research Data Management (RDM) practices in the field, and others. Despite the rapid evolution of technology, the process of collecting new data in MSE still heavily relies on loosely defined analogue processes, even if the data products are, in the end, digital. Moreover, the absence of proper infrastructure, such as centralized databases for both data products and metadata, hampers preservation of this data. Scientific findings are often disseminated without establishing a clear link to the underlying data or software utilized in their generation, and there is a notable lack of integration with well-defined licenses or policies governing their usage. The absence of formalized processes and detailed descriptions poses significant challenges in conducting follow-up research, reproducing results, identifying potential sources of errors and workflow bottlenecks, and validating research findings.

Descriptive, process- and domain-specific metadata are typically encapsulated either in only human-readable resources or digital artifacts that are not machine-actionable. The multidisciplinary nature of MSE compounds the issue of a lack of common and shared representations for material knowledge, e.g., by ontologies, which are still insufficient in this field.

Laboratory metadata in MSE manifests in various forms, ranging from handwritten notes in laboratory notebooks or paper protocols to semi-digital lists and fully digital, yet not necessarily machine-readable or -actionable, data files. The adoption of electronic laboratory notebooks remains minimal. These factors collectively make it exceedingly challenging to analyze scientific laboratory workflows in MSE or employ process records in automated data analysis processes.

In this paper, we will outline how the majority of the MSE community's data management challenges can be addressed through the incorporation of existing semantic technologies.

2. Related work on semantic technologies in MSE

Currently, numerous national and international initiatives and consortia in MSE aim to establish semantic technologies, such as shared vocabularies, metadata schemas, and ontologies. Notable examples include the European Materials Modelling Council (EMMC) [2], National Research Data Infrastructure (NFDI) [3] represented by its consortia MatWerk [4] and FAIRmat [5], Platform MaterialDigital [6], Materials Genome Initiative (MGI) [7], DICE Materials Data Platform [8], Diadem materials exploratory [9], materplat [10], Materials Commons [11], building the Prototype Open Knowledge Network Program [12], and re3data [13].

The first comprehensive review of ontologies in MSE, focusing on Domain-Level Ontologies (DLOs), was recently published by De Baas et al. [14]. Among the most widely used Top-Level Ontologies (TLOs) in MSE are the Basic Formal Ontology (BFO) [15], the Elementary Multiperspective Material Ontology (EMMO) [16], the Semantiscience Integrated Ontology (SIO) [17], and the Suggest Upper Merged Ontology (SUMO) [18]. These TLOs provide a framework for developing semantically consistent DLOs by offering general concepts for their development. Of the 43 analyzed DLOs, 21 reuse EMMO, 10 BFO, 2 SIO, and 1 SUMO, respectively. The remaining 10 do not reuse any of the four mentioned TLOs. Therefore, most DLOs in MSE (48%) are currently aligned with EMMO [14].

Currently, there is only one repository focused on MSE ontologies called MatPortal [19], which covers 30 MSE ontologies. IndustryPortal also includes MSE ontologies, among others

[20]. Additionally, NFDI offers the NFDI4Ing Terminology Service, which provides an ontology repository for industrially relevant ontologies, but contains only a few MSE ontologies [21].

Ontologies are often used as the schema (T-box) component of a Knowledge Graph (KG), which is then populated with domain-specific data, such as from experiments. Documents can also contain relevant data or metadata from experiments. One way to populate KG with information extracted from documents, are Large Language Models (LLMs). Ligabue et al. address open information extraction on textual documents by creating sub-graphs from documents and merging them into a larger KG [22]. Kesri et al. construct a KG using the T-box to define the entities and relationships to extract from text [23]. A sub-category of LLM that introduce attention mechanisms are transformers [24]. These can be used for text generation, translation, classification, etc. Most of the work in this area, however, focuses on text consisting of sentences, while in engineering domains like MSE documents often contain semi-structured information like tables and bullet points.

3. Challenges adopting semantic technologies in MSE

3.1. Limited ontology development expertise

Domain experts in MSE often lack the knowledge and experience required for developing usable and semantically correct ontologies. Since properly educated ontology developers are not readily available, the domain experts in MSE are often solely responsible for developing, implementing, and testing semantic technologies for their everyday work. The missing knowledge itself might already result in sub optimal semantic artifacts with issues such as ambiguity, semantic conflicts, and low interoperability. Furthermore, taking simultaneously a domain expert and an ontology developer role might cause conflicts of interest between the roles. In these cases the domain expert role is often dominant, due to the higher amount of experience in it, which might lead to decisions that are sound from the domain perspective, but semantically incorrect. Furthermore, having little experience in the ontology developer role, the domain expert is often not aware of these consequences.

This raises the question why the required experience and knowledge is lacking. First, the adoption of ontologies in MSE is still in its early stages, explaining the lack of experience. Second, acquiring the required knowledge is not straightforward. Directly learning from examples of other domains that have progressed further in adopting ontologies, e.g. biomedical domains, requires a good understanding of the respective domains. Third, domain-agnostic ontology development methodologies are not well standardized, which poses a challenge for the inexperienced ontology developer. While there are multiple ontology development methodologies available and in use, they all share common roots and core processes. However, most common frameworks (such as METHONTOLOGY [25], On-To-Knowledge [26], DILIGENT [27], or SAMOD [28], just to name a few), even if often straightforward and lightweight, do not adequately address modern challenges related to ontology reuse, community development, and maintenance. Some of the methodologies have not been updated for a long while, have never been adopted for practical use for ontology development by domain experts, or their practical applicability has never been published and is thus unknown to the public. Due to these difficulties and the lack of expertise in ontology development by the domain experts, the developed

ontologies often lack in quality. This is especially problematic when more complicated issues arise due to development activities such as ontology reuse.

3.2. Ontology reuse

There is an understanding of the necessity for and usefulness of interoperability of ontologies in the MSE domain. However, finding suitable reusable ontologies, evaluating their quality, and assessing their applicability are tasks typically within the expertise of semantic technology specialists such as ontology developers. Despite the abundance of existing ontologies, locating a specific one that fulfills the requirements to be reused for a given use case can be challenging. In heterogeneous fields like MSE, domain ontologies, even when developed as part of an initiative or consortium, are frequently created independently of other initiatives. This lack of coordination between the initiatives makes these ontologies challenging to discover or harmonize, which is especially important in case of intersecting domains. While ontology repositories that aid in finding MSE ontologies exist, they are not well connected with each other and provide little information about vital aspects like scope and addressed topics of the ontologies, which are important to consider when reusing ontologies. Consequently, the concepts of ontologies are findable, but it requires effort and expertise to assess how these concepts are related to the topics one wants to address with ontology reuse. Moreover, owing to the often necessarily pragmatic approach used in ontology development in MSE, the developed ontologies are often not rich or explicit in their semantics, offering only taxonomic relationships, and lack documentation of the developed concepts [14]. This further compromises the findability of ontologies due to missing concept definitions, only allowing ambiguous interpretations of the respective semantic artifacts [14], thus complicating identification of relevant concepts.

Finding reusable ontologies is often further complicated, because it may not be possible to find a single ontology that meets all given requirements. One option is to reuse multiple ontologies to cover the requirements while maintaining high interoperability. However, this leads to multiple challenges regarding their semantic compatibility, which possibly results in the need to repeatedly align the ontologies during their integration. TLOs offer a semantic framework for harmonization and integration of DLOs [29]. However, DLOs in MSE are sometimes not based on a TLO, or based on multiple TLOs [14]. The latter case can cause complex conflicts during DLO integration since concepts from different TLOs are often based on foundationally different assumptions. As a result, implications for the semantic compatibility of DLO concepts based on different TLOs can be hard to grasp, even for experienced ontology developers.

Furthermore, it might not be possible to find required subdomains for a particular use case, or the available ontologies of the subdomain are not interoperable with other required subdomains. While the former aspect may be attributed to generally incomplete and heterogeneous coverage of MSE subdomains, the latter is a consequence of the lack of harmonization endeavors between subdomains [14].

The last group of challenges regarding the compatibility of semantic artefacts between different DLOs is caused by the inconsistent use of terminology and semantics. For example, two concepts from different DLOs might have the same label, but actually refer to different semantic concepts - or vice versa. This example becomes especially challenging, when the semantic concepts also lack an elucidating definition and descriptions, which may lead to

ambiguous semantic concepts. While the structure of the DLOs can help decrease ambiguity of these concepts, the pragmatically structured DLOs in MSE often do not contain the required semantic richness to do so.

3.3. Standardizing laboratory documentation

Laboratory work documentation is essential to ensure accurate replication, reproducibility, and traceability of experiments. At the same time, its preparation is a time-consuming and tedious task. Meticulous records of experimental procedures facilitate efficient data analysis and collaboration by enabling researchers to build on existing work. It is also necessary to meet regulatory requirements in the laboratory environment. However, diversity in documentation formats and limited use of Electronic Laboratory Notebooks (ELNs) complicate standardization efforts and the digitalization of laboratory (meta)data in MSE. Non-unified documentation practices and often manual (meta)data acquisition produce digital artifacts that have limited reusability, even for their original authors. Since metadata is not always connected to their related data, the establishment of meaningful relationships between datasets is hindered as well. This especially applies to process-specific records, where, without a semantic model of laboratory processes in place, the result is only artifacts of workflow runs lacking data provenance or data lineage for specific data products. The systematic use of Persistent Identifiers (PIDs) for samples, equipment, and other elements of laboratory processes is also insufficient, itself not preventing poor metadata quality associated with these elements. Moreover, if MSE input data does not align with FAIR principles, it becomes difficult to produce FAIR data outputs. Manual, unstructured documentation is time-consuming and inefficient, yet researchers and laboratory technicians often resist adopting new technologies due to perceived complexity and the significant time required for implementation. The high investment in time, money, and expertise needed to implement semantic technologies presents a considerable barrier for many laboratories. Furthermore, the complexity of MSE data, which is often highly specialized, complicates mapping to existing semantic schemas. Establishing and maintaining relevant ontologies is resource-intensive, posing additional challenges, as noted in Sections 3.1 and 3.2.

4. Addressing the challenges

4.1. Standardization and interoperability

Adhering to a well-defined methodology in ontology development ensures high quality and reusability of the final product by providing structured guidelines for consistent design, rigorous validation to catch errors early, and providing systematic documentation to support maintenance and interoperability with other ontologies. This aspect is particularly relevant for domain experts who are new to ontology development, as predicting potential issues in ontology development can be challenging for them. The Linked Open Terms (LOT) methodology [30], to our best knowledge, is the first ontology development methodology specifically oriented towards ontology publication using semantic web and FAIR principles. It provides users with practical guidance and tools, which are particularly beneficial for domain experts transitioning into the role of ontology developer, navigating the complexities associated with lack of experience in

ontology development. This methodology was developed using established practices and aligns ontology development with agile software practices by using sprints and iterations. It divides the entire iterative process into four main steps: requirements specification, implementation, publication, and maintenance, making task division more accessible for domain experts acting as ontology developers. While domain knowledge is essential for the first two steps, domain experts not advanced in version control best practices can share the latter two steps with an ontology expert or IT specialist without domain knowledge. This approach mitigates challenges arising from potential role conflicts. The methodology improves the structure of the development process, enhancing task granularity and accessibility throughout the steps. It offers practical advice on formulating competency questions, defines core ontology elements during the conceptualization phase, and helps to decide between self-modeling and reusing concepts from other ontologies. These aspects address common risks of potentially low quality outputs of the ontology development process in MSE, resulting from lack of extensive experience in ontology development among domain experts taking this role.

LOT has been proven effective across a selection of diverse projects, covering a range of applications and domains [30], positioning it as a robust candidate for a domain-agnostic standard approach to ontology development, with high potential for adoption in such diverse domains as MSE. It introduces activities such as ontology publication, along with practical tips and recommendations, which are critical as other methodologies may not fully align with modern standards. Moreover, the methodology significantly enhances ontology reuse and tooling, particularly beneficial and relevant for heterogeneous domains like MSE. Establishing a dedicated framework for ontology development based on iterative processes, harmonization, and ontology reuse within this framework proves more efficient as it reduces redundancy in development efforts, and ensures consistency across ontology development projects. Adopting LOT widely in MSE has the potential to establish a standardized workflow for ontology design, bridging the gap between domain experts and ontology developers through structured guidelines and processes. By providing specific modeling guidelines based on established ontologies, LOT can serve as a model for other ontology development initiatives, fostering alignment and integration across different projects. This approach not only promotes consistency but also enhances the overall quality of ontologies developed within these diverse initiatives.

Meanwhile, to enhance the interoperability of DLOs in MSE, concepts from a common TLO should be reused to provide a base harmonisation of the DLO concepts. If DLOs use different TLOs, it is suggested to engineer bridge concepts [14]. Interoperability between hard to align ontologies can also be established on data level. For example, the I-ADOPT [31] Framework ontology provides concepts to establish interoperability on a level of existing variable description models. This can provide standardized interoperability on a variable level while mitigating semantic conflicts on other levels. Variables are often important concepts in MSE ontologies such as material properties, process parameters, and measurement results.

Besides manual ontology development, there are approaches to enhance ontologies (semi-)automatically. Ontology Learning (OL) is a field of research in Artificial Intelligence (AI) and knowledge engineering. LLM-based OL techniques have shown to perform better than traditional ones on unstructured text to extract significant and domain relevant concepts and their inter-dependencies [32]. Work by da Silva et al. further explores how prompting with LLMs affect the ontology generation and found the results to be promising, thereby generating

ontologies which are error-free [33]. LLMs in general have shown to perform better than traditional supervised learning-based techniques across various domains in an out-of-the-box manner, i.e., without having the need for finetuning to attain automation. However, for performing domain-specific tasks within scientific domains like MSE, domain knowledge has to be integrated within LLMs to get optimal results and better domain relevancy.

KGs are, unlike ontologies, meant to contain also entity-level data. Modelling KGs is even more time-consuming than modeling ontologies. Information Extraction (IE) focuses on extracting information from documents, e.g., from experiment protocols or from data sheets. KGs can subsequently be enriched by this extracted information to reduce manual effort and increase scalability. To enrich KGs with IE, triples must be generated and incorporated into the KG, a process known as KG fusion or KG population. A typical IE task, comprising of sub-tasks like entity extraction and relation extraction, has shown to achieve better performance using out-of-the-box LLMs with correct prompting techniques than with model finetuning [34]. In addition to IE, LLMs can also be used for KG population by creating KG embeddings [35]. However, to enable LLMs to identify domain specific entities and concepts, domain ontologies are required. By using domain ontologies as the ground-truth, LLMs can be prompted to do entity extraction. As they are probabilistic models, LLMs inherently lack factual reliability. Constraint decoding is therefore applied to control the output of the LLM. It is a way of integrating ontologies with LLMs to make them "speak" a domain specific language. Luo et al. use ontology-based constraint decoding within LLMs to generate comparatively less noisy training data for the task of named entity recognition in resource-scarce domains [36]. Prompt engineering represents another approach, integrating ontologies into the LLM prompts to instruct LLMs to perform a particular task. Mihindukulasooriya et al. propose a prompting framework for constructing KGs using ontologies and LLMs [37]. The ontology is supplied to the prompt, which is used to instruct the LLM. Therefore, LLMs can offer a boost to IE and KG enrichment for the domain of MSE using domain-ontologies to prevent hallucinations.

4.2. Technological advancement in laboratory practices

Despite the significant variations among subdomains within MSE, there are noticeable parallels in laboratory processes across the domain. Therefore, establishing unified frameworks for collecting records of these processes can be done with a relatively low effort. Introducing a process ontology, with a particular focus on gathering process provenance records in a uniform format, can formalize these procedures. Utilizing PROV-O [38] and incorporating its extension, the P-Plan Ontology [39], can effectively address this need, ensuring standardized documentation of processes across the domain. Such unified records would greatly facilitate the analysis of process bottlenecks, identification of synergies, and areas for improvement across MSE subdomains and the field as a whole. Maintaining provenance records of workflow runs also enables predictions for similar runs in the future, such as expected runtime or resources required. Workflow provenance can also be used to detect outliers in workflow runs. A standardized framework for process description, based on a well recognized model, makes it easier to incorporate domain knowledge with the use of DLOs. Such a unified process ontology can not only address the issue of diversity in laboratory documentation practices but also ensure the FAIRification of output data products. It would be achieved by connections

between provenance information and the elements of laboratory workflows (e.g. inputs, outputs, metadata of samples or equipment).

The semantic description of laboratory processes aligns well with the utilization of ELNs. To overcome resistance to developing and implementing new technologies in the community, one should be considerate when choosing a suitable ELN. First, it should reduce the effort necessary for efficient and easy manual forms population while reducing the number of errors. This can be achieved by field restrictions, such as drop-down menus and other forms of assistance in free text entries, such as semantic annotation of the content. Second, the implementation of ontologies is necessary to semantically annotate the entries or cross-reference sub-elements of a given form, which can be especially useful for automating the ELN use. ELN Finder [40] is a tool designed to help identify a suitable ELN for a particular (sub)domain and use case. It supports a broad range of selection criteria, including the ability to define custom templates or device connections, customize the user interface, choose the type of license, and utilize controlled vocabularies. A limited number of ELNs integrate ontologies for data annotation (e.g., Chemotion ELN [41]) or to generate semantically annotated forms (e.g., Herbie [42]). These and other ELNs suitable for use in MSE are part of the ELN Consortium [43], which aims to establish common specifications for data and metadata exchange between different ELNs. Additionally, ontology use for form creation can also enable the generation of digital laboratory work protocols and other documents automatically. Once semantically annotated with an ontology, ELN records can be directly stored in a KG without requiring additional post-processing. It is important to include PIDs utilization while automating the ELN use. Even without full FAIRification of laboratory data, PIDs can serve as a minimum form of identification of workflow artifacts, including physical data, when implementing a process ontology. Reuse of provenance ontologies also addresses crucial aspects such as the reproducibility of a particular process run, comparison between different workflow runs, and visualization of process records.

5. Conclusion and outlook

We have given an overview, which and how semantic technologies can be used to tackle data management challenges in MSE, including the domain specific difficulties in doing so. To show actual results based on this theoretical roadmap, we started a project called "Ontology-based Data Integration and eXploration" (ODIX). In ODIX, we re-use and create ontologies in the MSE domain, based on the LOT methodology. These ontologies are then used to guide LLMs in information extraction from documents. At the same time, the ontologies build the T-Box of a KG, which gets enriched with the information extracted from documents. The information in the KG is consequently well described to be used as input for further applications.

A key application in ODIX is a visual exploration tool that allows domain experts to compare information across different process steps of the same material probe or between different runs of the same process — all formerly only stored in documents. The project involves collaboration between computer scientists (primarily semantic experts) and domain experts (mostly in the area of MSE). Evaluation of the results will be conducted by the domain experts with outcomes to be published as the project advances. We look forward to seeing further practical applications of the methodologies and processes outlined in this paper.

References

- [1] 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.
- [2] The European Materials Modelling Council, 2024. URL: <https://emmc.eu/>.
- [3] The German National Research Data Infrastructure (NFDI) e.V., 2024. URL: <https://www.nfdi.de/association/?lang=en>.
- [4] NFDI-MatWerk, Nationale Forschungsdateninfrastruktur für Materialwissenschaft & Werkstofftechnik, 2024. URL: <https://nfdi-matwerk.de/>.
- [5] FAIRmat, Consortium FAIRmat, 2024. URL: <https://www.fairmat-nfdi.eu/fairmat/about-fairmat/consortium-fairmat>.
- [6] Material Digital - the material digitalization platform, 2019. URL: <https://materialdigital.de/>.
- [7] Materials Genome Initiative, 2014. URL: <https://www.mgi.gov/>.
- [8] NIMS, DICE Materials Data Platform with Greatly Improved Data Collection and Accumulation Capabilities, 2023. URL: <https://www.nims.go.jp/eng/news/press/2023/01/202301170.html>.
- [9] The 'Diadem' (materials) exploratory PEPR, 2023. URL: <https://www.cnrs.fr/en/pepr/pepr-exploratoire-diademe-materiaux>.
- [10] Plataforma Tecnológica Española de Materiales Avanzados y Nanomateriales, 2024. URL: <https://materplat.org/>.
- [11] Materials Commons, 2020. URL: <https://materialscommons.org/>.
- [12] U.S. National Science Foundation, NSF invests \$26.7 million in building the first-ever prototype open knowledge network, 2023. URL: <https://new.nsf.gov/tip/updates/nsf-invests-first-ever-prototype-open-knowledge-network>.
- [13] Registry of Research Data Repositories, 2020. URL: <https://www.re3data.org/>.
- [14] 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.
- [15] R. Arp, B. Smith, A. D. Spear, Building Ontologies with Basic Formal Ontology, MIT Press, Cambridge, MA, 2015. doi:10.7551/mitpress/9780262527811.001.0001.
- [16] The Elementary Multiperspective Material Ontology, 2024. URL: <https://emmo-repo.github.io/>.
- [17] 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). doi:10.1186/2041-1480-5-14.
- [18] I. Niles, A. Pease, Towards a standard upper ontology, in: Proceedings of the international conference on Formal Ontology in Information Systems - Volume 2001, FOIS01, ACM, 2001, pp. 2–9. doi:10.1145/505168.505170.
 - [19] Material Open Laboratory MatPortal, 2021. URL: <https://matportal.org/>.
 - [20] E. Amdouni, A. Sarkar, C. Jonquet, M. H. Karray, IndustryPortal: a common repository for FAIR ontologies in industry 4.0, in: ISWC 2023 - 22nd International Semantic Web Conference, demo & poster session, Athens, Greece, 2023. URL: <https://hal.science/hal-04207343>.
 - [21] NFDI4Ing Terminology Service, 2021. URL: <https://terminology.nfdi4ing.de/ts/>.
 - [22] P. d. M. Ligabue, A. A. F. Brandão, S. M. Peres, F. G. Cozman, P. Pirozelli, Applying a Context-based Method to Build a Knowledge Graph for the Blue Amazon, Data Intelligence 6 (2024) 64–103. doi:10.1162/dint_a_00223.
 - [23] V. Kesri, A. Nayak, K. Ponnalagu, AutoKG - An Automotive Domain Knowledge Graph for Software Testing: A position paper, in: 2021 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW), IEEE, 2021, pp. 234–238. doi:10.1109/icstw52544.2021.00047.
 - [24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, I. Polosukhin, Attention is All you Need, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, volume 30, Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
 - [25] M. Fernández-López, A. Gómez-Pérez, N. Juristo Juzgado, METHONTOLOGY: From Ontological Art Towards Ontological Engineering (1997). URL: <https://oa.upm.es/5484/>, ontology Engineering Group - OEG.
 - [26] S. Staab, R. Studer, H.-P. Schnurr, Y. Sure, Knowledge processes and ontologies, IEEE Intelligent systems 16 (2001) 26–34. doi:10.1109/5254.912382.
 - [27] H. S. Pinto, S. Staab, C. Tempich, DILIGENT: Towards a fine-grained methodology for Distributed, Loosely-controlled and evolving Engineering of ontologies, in: ECAI, volume 16, 2004, p. 393.
 - [28] S. Peroni, A Simplified Agile Methodology for Ontology Development, in: OWL: Experiences and Directions—Reasoner Evaluation: 13th International Workshop, OWLED 2016, and 5th International Workshop, ORE 2016, Bologna, Italy, November 20, 2016, Revised Selected Papers 13, Springer, Springer International Publishing, 2017, pp. 55–69. doi:10.1007/978-3-319-54627-8_5.
 - [29] ISO/IEC, 21838-1:2021(en), Information technology — Top-level ontologies (TLO) — Part 1: Requirements, Standard, International Organization for Standardization, Geneva, CH, 2001.
 - [30] 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.
 - [31] B. Magagna, I. Rosati, M. Stoica, S. Schindler, G. Moncoiffé, A. Devaraju, J. Peterseil, R. Huber, The I-ADOPT Interoperability Framework for FAIRer data descriptions of biodiversity, CoRR abs/2107.06547 (2021). doi:10.48550/ARXIV.2107.06547. arXiv:2107.06547.
 - [32] H. Babaei Giglou, J. D’Souza, S. Auer, LLMs4OL: Large Language Models for Ontology

- Learning, in: International Semantic Web Conference, Springer, 2023, pp. 408–427. doi:10.1007/978-3-031-47240-4_22.
- [33] L. M. V. da Silva, A. Köcher, F. Gehlhoff, A. Fay, On the Use of Large Language Models to Generate Capability Ontologies, arXiv preprint arXiv:2404.17524 (2024). doi:10.48550/ARXIV.2404.17524.
- [34] L. Luo, Y.-F. Li, G. Haffari, S. Pan, Reasoning on Graphs: Faithful and Interpretable Large Language Model Reasoning, arXiv preprint arXiv:2310.01061 (2023). doi:10.48550/ARXIV.2310.01061.
- [35] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, X. Wu, Unifying Large Language Models and Knowledge Graphs: A Roadmap, IEEE Transactions on Knowledge and Data Engineering 36 (2024) 3580–3599. doi:10.1109/tkde.2024.3352100.
- [36] Z. Luo, Y. Wang, W. Ke, R. Qi, Y. Guo, P. Wang, Boosting LLMs with Ontology-Aware Prompt for Ner Data Augmentation, in: ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, IEEE, 2024, pp. 12361–12365. doi:10.1109/icassp48485.2024.10446860.
- [37] N. Mihindukulasooriya, S. Tiwari, C. F. Enguix, K. Lata, Text2KGBench: A Benchmark for Ontology-Driven Knowledge Graph Generation from Text, in: International Semantic Web Conference, Springer, Springer Nature Switzerland, 2023, pp. 247–265. doi:10.1007/978-3-031-47243-5_14.
- [38] The PROV Ontology, 2013. URL: <https://www.w3.org/TR/prov-o>.
- [39] The P-PLAN Ontology, 2013. URL: <https://www.opmw.org/model/p-plan>.
- [40] ELN Finder, 2024. URL: <https://eln-finder.ulb.tu-darmstadt.de>.
- [41] Chemotion ELN, 2024. URL: <https://www.chemotion.net/docs/eln>.
- [42] 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.
- [43] ELN-Consortium: A consortium of ELN vendors, 2024. URL: <https://github.com/TheELNConsortium>.