

AI4DiTraRe: Building the BFO-Compliant Chemotion Knowledge Graph

Ebrahim Norouzi^{1,2,*}, Nicole Jung³, Anna M. Jacyszyn¹, Jörg Waitelonis¹ and Harald Sack^{1,2}

¹FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

²Karlsruhe Institute of Technology, Institute of Applied Informatics and Formal Description Methods, Kaiserstr. 89, 76133 Karlsruhe

³Karlsruhe Institute of Technology, Institute of Biological and Chemical Systems, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

Abstract

Chemistry is an example of a discipline where the advancements of technology have led to multi-level and often tangled and tricky processes ongoing in the lab. The repeatedly complex workflows are combined with information from chemical structures, which are essential to understand the scientific process. An important tool for many chemists is Chemotion, which consists of an electronic lab notebook and a repository. This paper introduces a semantic pipeline for constructing the BFO-compliant Chemotion Knowledge Graph, providing an integrated, ontology-driven representation of chemical research data. The Chemotion-KG has been developed to adhere to the FAIR (Findable, Accessible, Interoperable, Reusable) principles and to support AI-driven discovery and reasoning in chemistry. Experimental metadata were harvested from the Chemotion API in JSON-LD format, converted into RDF, and subsequently transformed into a Basic Formal Ontology-aligned graph through SPARQL CONSTRUCT queries. The source code and datasets are publicly available via GitHub. The Chemotion Knowledge Graph is hosted by FIZ Karlsruhe Information Service Engineering. Outcomes presented in this work were achieved within the Leibniz Science Campus “Digital Transformation of Research” (DiTraRe) and are part of an ongoing interdisciplinary collaboration.

Keywords

digitalisation, chemistry, ontology, knowledge graph

1. Introduction

The generation of FAIR data relies on the ability to easily apply standards and to produce well-structured, well-annotated datasets. Electronic Lab Notebooks (ELNs) are essential tools in promoting the digitalization of scientific research, as they help incorporate standards and structure into the workflows of experimental scientists [1].

In the field of chemistry, the development of an ELN that supports the creation of findable, accessible, interoperable, and reusable (FAIR) data is particularly challenging. This complexity arises from the nature of chemical research, which involves intricate and highly diverse experimental workflows, the use of various measurement devices that generate large volumes of data, and a wide range of data formats. Additionally, the handling of chemical structures which must be drawn, interpreted, and processed is indispensable for documenting and understanding chemical experiments. These requirements among others make chemistry data management especially demanding compared to other disciplines.

In recent years, a team of software developers and scientists at the Karlsruhe Institute of Technology (KIT) has been working on a research data management (RDM) environment specifically designed for chemistry. This environment, called Chemotion, comprises an ELN [2, 3] tailored to the specific needs of chemists, along with a research data repository [4] that interoperates seamlessly with the ELN. The

5th International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment November 2025 - Nara, Japan; co-located with The 24th International Semantic Web Conference, ISWC 2025

*Corresponding author.

✉ Ebrahim.Norouzi@fiz-Karlsruhe.de (E. Norouzi); nicole.jung@kit.edu (N. Jung); Anna.Jacyszyn@fiz-Karlsruhe.de (A. M. Jacyszyn); joerg.waitelonis@fiz-karlsruhe.de (J. Waitelonis); Harald.Sack@fiz-Karlsruhe.de (H. Sack)

ORCID: 0000-0002-2691-6995 (E. Norouzi); 0000-0001-9513-2468 (N. Jung); 0000-0002-5649-536X (A. M. Jacyszyn); 0000-0001-7192-7143 (J. Waitelonis); 0000-0001-7069-9804 (H. Sack)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

repository allows researchers to publish their recorded and analyzed data in accordance with FAIR principles.

The combination of the Chemotion ELN and the interoperable repository Chemotion provides scientists with a comprehensive digital infrastructure that addresses their domain-specific requirements. It enables them to manage, organize, analyze, and publish their data in a way that enhances transparency and ensures the reusability of their research outcomes [5, 6, 7].

Driven by the needs of scientists for improved documentation and data management, particularly in response to the unique challenges posed by chemistry data, the Chemotion systems were initially developed without a systematic approach for integrating ontologies and semantic meaning. This limitation is now being addressed incrementally, through the implementation of a semantically coherent framework capable of representing research lifecycle stages, agent roles, provenance information, and domain-specific chemical entities in a machine-interpretable manner.

To support and enhance the semantic enrichment of Chemotion, an ontology-driven modeling approach grounds the Chemotion Knowledge Graph (Chemotion-KG) in upper-level semantics provided by the Basic Formal Ontology (BFO) using the Chemotion repository as a source as it includes well curated data assigned to publication metadata. Ontology Design Patterns (ODPs) are adopted as a systematic method to encode these structures, serving as reusable solutions to recurring ontology engineering problems [8]. Their use ensures not only the semantic consistency of datasets, creators, studies, and chemical substances in Chemotion-KG but also the reuse of established modeling practices that facilitate alignment with broader scientific knowledge graphs and ontology standards.

The Chemotion-KG, a synergy of Chemotion and semantics, is being created within the Leibniz Science Campus “Digital Transformation of Research” (DiTraRe)¹ [9, 10]. DiTraRe studies effects of digitalisation of research in a multilevel, interdisciplinary way. In this multidimensional exchange, one of the use cases is Chemotion which represents novel methods of data acquisition. The DiTraRe use cases are being analysed from different aspects by teams called *dimensions*. “Exploration and Knowledge Organisation” (AI4DiTraRe) is the dimension responsible for applying AI and studying its effects [11]. Collaboration of the aforementioned dimension and the Chemotion use case has by now provided outcomes described in this paper.

In brief, this work presents the construction of the BFO-compliant Chemotion-KG for semantically integrating experimental chemistry data. Section 2 describes the workflow for data harvesting, semantic transformation, and ontology alignment. Section 3 presents the resulting knowledge graph structures and instantiated entities generated using ontology design patterns. Section 4 discusses the broader impact of the Chemotion-KG and future directions. Finally, section 5 provides a summary of the achieved contributions.

The source code and datasets are publicly available at <https://github.com/ISE-FIZKarlsruhe/chemotion-kg>. The Chemotion-KG is hosted at <https://ditrare.ise.fiz-karlsruhe.de/chemotion-kg/>.

2. Knowledge Graph Construction Approach

This section describes the end-to-end pipeline for constructing the BFO-compliant Chemotion-KG, covering metadata harvesting, semantic enrichment, ontology alignment, reasoning, and materialization. The workflow is depicted in Figure 1, illustrating the transformation from schema-based metadata to a semantically aligned knowledge graph.

The initial metadata were harvested from the Chemotion repository², which provides structured schema.org-based descriptions of chemical research data. The repository’s data covered (in)organic reactions, analytical measurements such as nuclear magnetic resonance (NMR) spectroscopy, mass spectrometry (MS), IR-, and Raman spectroscopy data. The metadata were expressed primarily with `schema:Dataset`, `schema:ChemicalSubstance`, and `schema:Study`, which provide only lightweight descriptive semantics. Consequently, while adequate for basic metadata exchange, the schema.org

¹DiTraRe web page, <https://www.ditrare.de/en>

²<https://www.chemotion-repository.net>

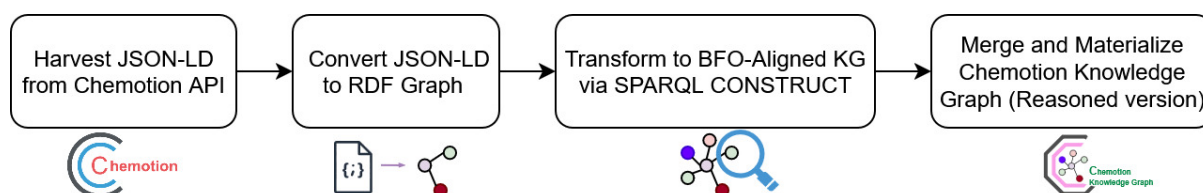


Figure 1: Schematic workflow of the Chemotion Knowledge Graph construction.

representation did not capture the rich semantic structure required for ontology-driven reasoning, provenance modeling, and alignment with upper-level ontologies such as the BFO³[12].

To address this, the harvested JSON-LD was converted into RDF graphs with canonical URIs and semantically enriched through a SPARQL CONSTRUCT transformation. The mapping aligned the schema-based metadata to the BFO, leveraging the NFDICore ontology⁴[13] for research data structures, lifecycle modeling, and provenance, and reusing ChEBI⁵[14] for chemical entities. The transformation queries were authored and documented using shmarql⁶, a Linked Data publishing platform supporting SPARQL-based data pipelines.

The overall SPARQL CONSTRUCT query implementing the semantic transformation is openly available in our GitHub repository⁷. An example SPARQL UPDATE used to semantically enrich a dataset is shown below. It replaces a plain `schema:Dataset` description with NFDICore-aligned metadata, preserving original values while adding explicit BFO-compliant types. One of the essential steps in this transformation is the use of `BIND` to create new IRIs for elements such as descriptions, identifiers, titles, and URLs. Instead of representing these values only as plain literals, they are modeled as separate instances (e.g., description node, identifier node) with their own URIs. This approach follows BFO principles, where information content entities are treated as individual resources that can carry provenance and lifecycle information.

Creating explicit IRIs for these nodes allows the knowledge graph to keep stable references to metadata entities even when their literal values change. For instance, if a dataset’s description or identifier is updated, the corresponding instance remains the same, making it possible to track changes over time and preserve provenance.

Listing 1: SPARQL CONSTRUCT for dataset transformation.

```
PREFIX schema: <http://schema.org/>
PREFIX nfdicore: <https://nfdi.fiz-karlsruhe.de/ontology/>
PREFIX obo: <http://purl.obolibrary.org/obo/>

CONSTRUCT {
  # Recast the dataset to BFO-aligned NFDICore class
  ?dataset a nfdicore:NFDI_0000009 ; # Dataset
    nfdicore:NFDI_0001027 ?creator ; # has creator (Person)
    nfdicore:NFDI_0000191 ?publisher ; # has publisher (Organization)
    obo:IAO_0000235 ?descriptionNode ; # description node
    nfdicore:NFDI_0001006 ?identifierNode ; # identifier node
    nfdicore:NFDI_0000142 ?license ; # license information
    nfdicore:NFDI_0000216 ?technique ; # associated measurement technique
    obo:IAO_0000235 ?nameNode ; # title node
    obo:IAO_0000235 ?urlNode ; # landing page URL node
```

³<https://basic-formal-ontology.org/>

⁴<https://ise-fizkarlsruhe.github.io/nfdicore/>

⁵<https://www.ebi.ac.uk/chebi/>

⁶<https://github.com/epoz/shmarql>

⁷<https://github.com/ISE-FIZKarlsruhe/chemotion-knowledge-graph/blob/main/processing/all-nfdicore.py>

```

    nfdicore:NFDI_0001023 ?study ; # is output of a study
    obo:BFO_0000178 ?catalog . # is part of a registered catalog
}
WHERE {
  # Original Chemotion metadata described using schema.org
  ?dataset a schema:Dataset ;
    schema:creator ?creator ;
    schema:publisher ?publisher ;
    schema:description ?description ;
    schema:identifier ?identifier ;
    schema:license ?license ;
    schema:measurementTechnique ?technique ;
    schema:name ?name ;
    schema:url ?url ;
    schema:includedInDataCatalog ?catalog ;
    schema:isPartOf ?study .

  # Generate canonical URIs for literal-based nodes to ensure global identification
  BIND(IRI(CONCAT("https://ditrare.ise.fiz-karlsruhe.de/chemotion-kg/nodes/",
    ENCODE_FOR_URI(?description))) AS ?descriptionNode)
  BIND(IRI(CONCAT("https://ditrare.ise.fiz-karlsruhe.de/chemotion-kg/nodes/",
    ENCODE_FOR_URI(?identifier))) AS ?identifierNode)
  BIND(IRI(CONCAT("https://ditrare.ise.fiz-karlsruhe.de/chemotion-kg/nodes/",
    ENCODE_FOR_URI(?name))) AS ?nameNode)
  BIND(IRI(CONCAT("https://ditrare.ise.fiz-karlsruhe.de/chemotion-kg/nodes/",
    ENCODE_FOR_URI(?url))) AS ?urlNode)
}

```

To ensure that the resulting ontology design patterns and their interconnections are transparent and reusable, a standardized graphical representation was adopted. The Chemotion-KG modules and their links were visualized using Graffoo, a formal graphical notation for OWL ontologies [15].

3. Chemotion-KG Structure and Ontological Representation

The Chemotion-KG was constructed by transforming schema.org-based metadata into a BFO-aligned representation. The resulting graph provides a semantically enriched, interoperable representation of experimental chemistry data, integrating structural, analytical, and provenance information. The semantic transformation leveraged the NFDICore ontology for research data lifecycle modeling and ChEBI for chemical entities, ensuring alignment with established FAIR data standards.

Figure 2 illustrates the dataset representation, where schema.org Dataset instances were mapped to nfdicore:NFDI_0000009 and enriched with identifiers, licensing information, catalog registration, and links to associated studies. The class nfdicore:NFDI_0000009 represents a dataset, modeled as an information content entity that denotes a structured collection of data, typically organized for a defined purpose such as analysis, research, or reference. In this context, a dataset constitutes structured information about a resource curated or provided by an individual researcher. The alignment introduces explicit semantics for provenance and lifecycle stages through NFDICore and BFO relations. Provenance is explicitly captured via relations such as nfdicore:NFDI_0001027 (*has creator*) and nfdicore:NFDI_0000191 (*published by*), while lifecycle stages are represented using BFO continuant–occurrent relations like obo:BFO_0000178 (*has continuant part*).

The creator model (Figure 4) demonstrates the transformation of schema:Person metadata into a semantically rich structure. Individual names, ORCID identifiers, and organizational affiliations were represented as distinct resources, using nfdicore:NFDI_0000004 for persons and nfdicore:NFDI_0000003

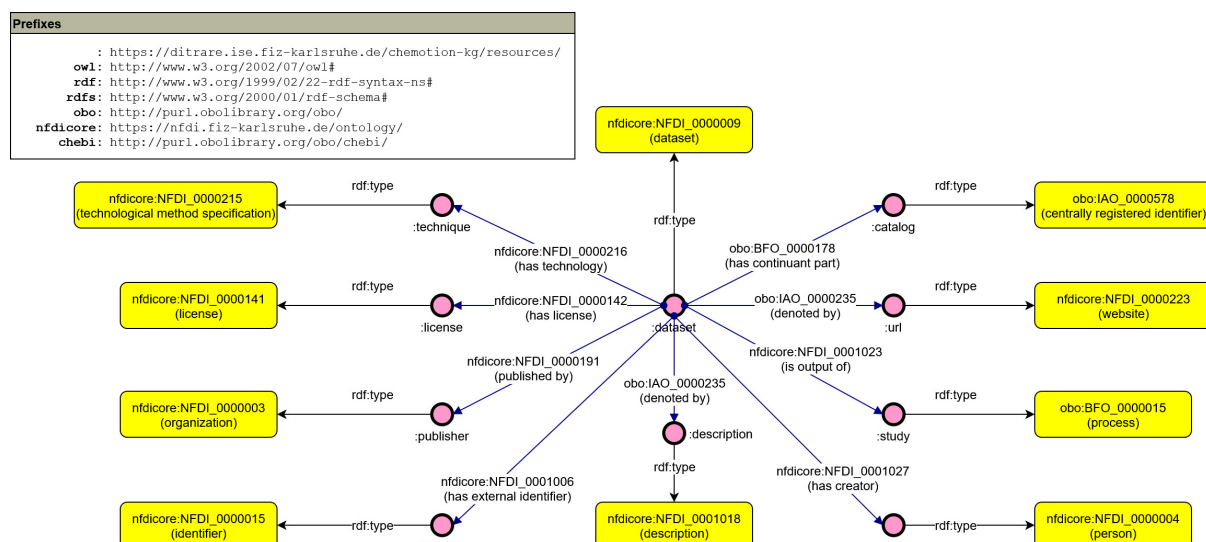


Figure 2: Dataset representation in the Chemotion-KG.

for organizations. Roles and affiliations were explicitly modeled using the Process–Agent–Role ODP⁸, aligning with the NFDICore and BFO modeling principles. This pattern establishes the relationship between *bfo:Process* (Occurrent), *nfdicore:Agent* (Independent Continuant), and *bfo:Role* (Specifically Dependent Continuant). Within the Chemotion-KG, *nfdicore:Agent* instances represent both organizations and persons involved in experimental chemistry workflows, including research institutions, and individual researchers. The pattern uses *bfo:has_participant*, *bfo:realizes*, and *bfo:bearer_of* to connect processes, agents, and their assigned roles, enabling explicit provenance tracking and role-based attribution for research data.

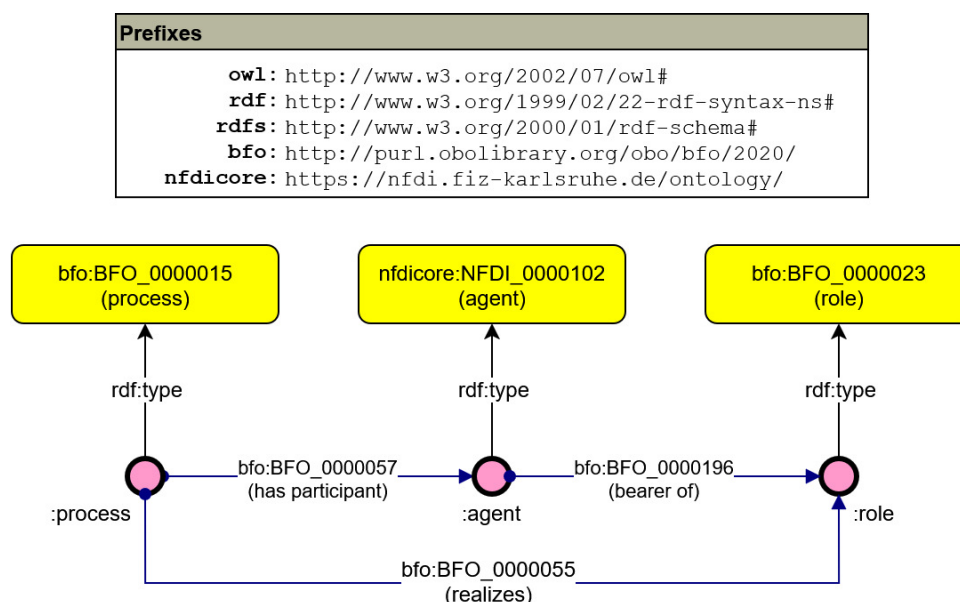


Figure 3: Process–Agent–Role Ontology Design Pattern.

Studies were represented as *obo:BFO_0000015* processes, connecting datasets, chemical substances, and associated publishing events (Figure 5). The publishing activities were explicitly modeled using *nfdicore:NFDI_0000014*, which is defined as a *process of making information available to the public*

⁸<https://ise-fizkarlsruhe.github.io/mwo/docs/patterns/#pattern-1-process-agent-role>

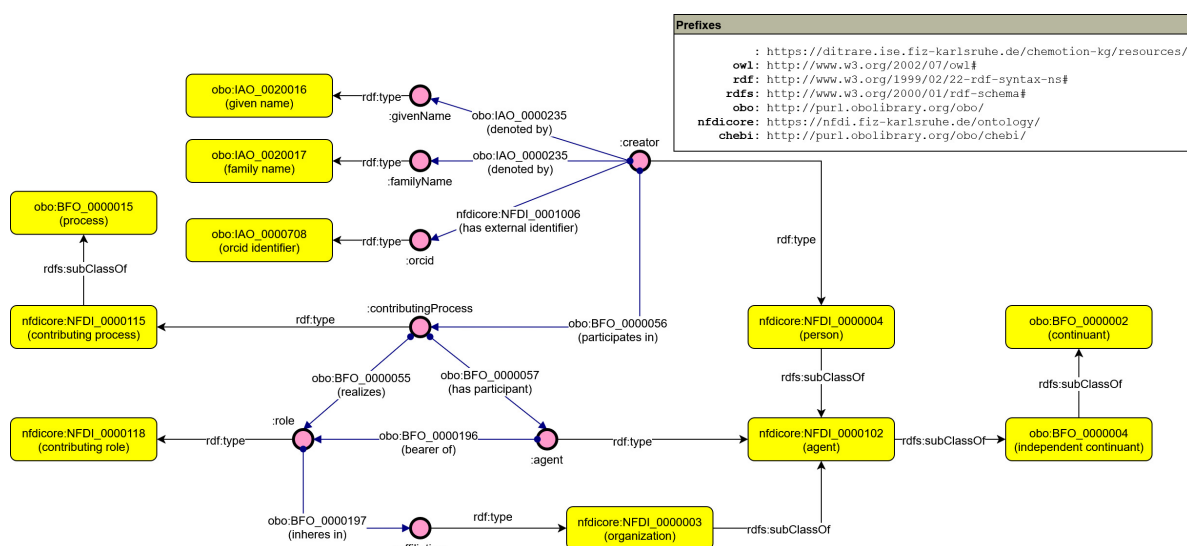


Figure 4: Creator description aligned to NFDICore and BFO patterns, enabling explicit representation of roles and affiliations.

either for sale or free access⁹. Within the Chemotion-KG, this class was used to capture the temporal regions of publication via obo:BFO_0000199 (*occupies temporal region*) and to link standard profiles with each study through nfdicore:NFDI_0000207 (*has standard*).

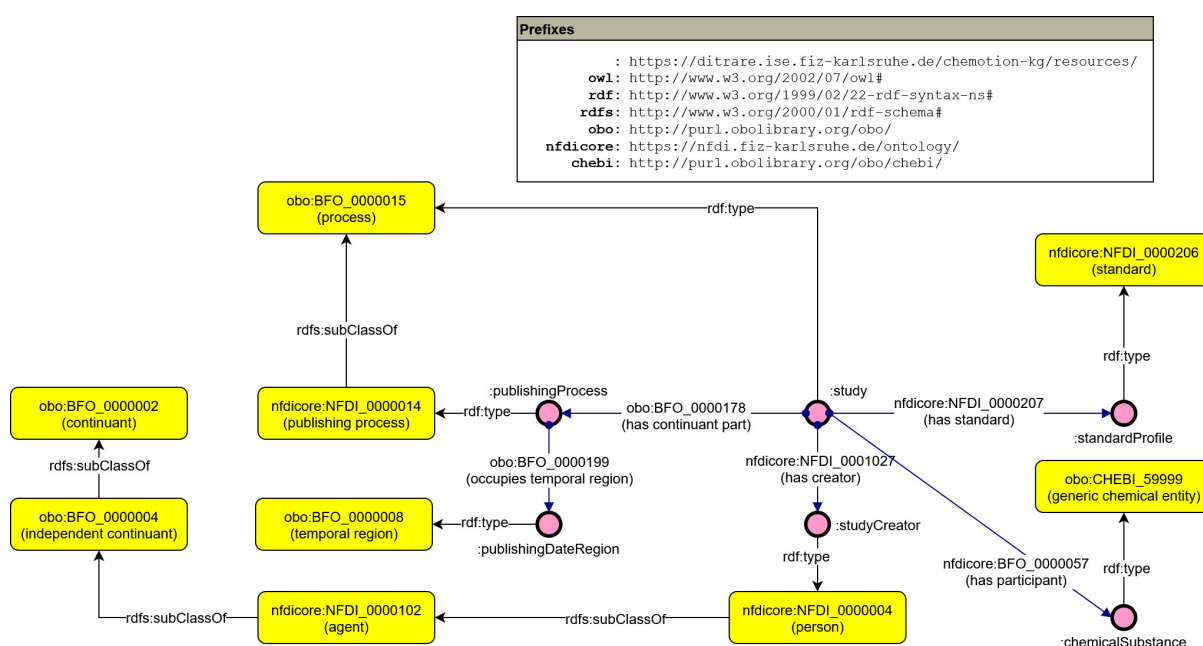


Figure 5: Study representation with explicit modeling of publishing processes, temporal regions, and standard profiles.

Chemical entities (Figure 6) were aligned to ChEBI classes to ensure standardized representation of chemical knowledge. Generic chemical substances were modeled as obo:CHEBI_59999, capturing the abstract notion of a chemical entity within a study, while specific molecular structures were represented as obo:CHEBI_23367 molecular entities. To guarantee unambiguous identification and interoperability, the alignment included explicit mapping of InChI identifiers, InChIKeys, SMILES strings, and molecular formulas to structured nodes. Molecular weights were represented using the obo:BFO_0000019 *quality* pattern combined with obo:IAO_0000109 *measurement datum*, linking to

⁹https://nfdi.fiz-karlsruhe.de/ontology/NFDI_0000014

obo:IAO_0000003 measurement units following the BFO measurement modeling principles. Structural images and external references were attached via nfdicore:NFDI_0000223 URL nodes to maintain provenance and facilitate linking to external chemical databases.

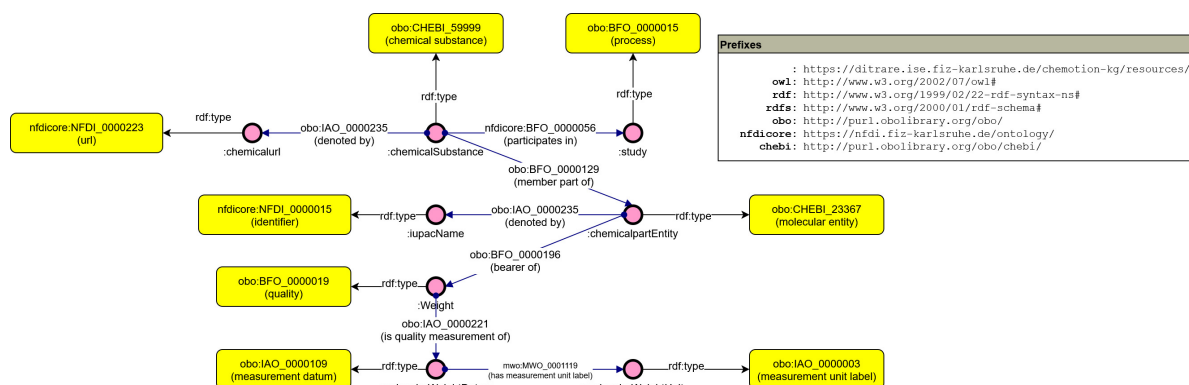


Figure 6: Chemical substance representation aligned to ChEBI and NFDICore, with explicit molecular entity and measurement modeling.

The resulting Chemotion-KG provides a semantically enriched integration of experimental chemistry data, employing formal ODPs for each core entity type, including datasets, creators, studies, and chemical substances. The transformation pipeline applies a SPARQL CONSTRUCT-based approach that preserves the original schema.org metadata while aligning it with BFO semantics and enriching it through NFDICore and ChEBI ontologies. This results in a BFO-compliant graph that supports semantic interoperability, logical reasoning, and linkage to external knowledge resources. The overall Chemotion-KG integrates all ODPs into a coherent and interconnected model, capturing datasets, creators, studies, and chemical substances within a BFO-compliant framework. As illustrated in Figure 7, each ODP is instantiated as a modular pattern and linked via well-defined object properties to form a semantically consistent network. This approach ensures that dataset descriptions, experimental studies, agents with roles, and chemical entities are harmonized through reusable modeling practices. The figure highlights how these ODPs are composed together to create a unified, semantically enriched representation of the Chemotion experimental data.

The Chemotion-KG is continuously updated, with new data from the Chemotion repository ingested into the graph on a daily basis. Instances are managed using a named graph that organizes resources according to their submission date. Each resource IRI encodes the year and month of submission alongside the original Chemotion identifier, ensuring temporal provenance and stable referencing. For example, the dataset representing Raman spectroscopy data is available at <https://ditrare.ise.fiz-karlsruhe.de/chemotion-kg/resources/2014/05/10.14272/VRYPQVRFMNXTJS-UHFFFAOYSA-N/Raman>, where the path segments 2014/05 capture the submission date and the suffix encodes the Chemotion-specific identifier. The corresponding source entry in the Chemotion repository is <https://www.chemotion-repository.net/inchikey/VRYPQVRFMNXTJS-UHFFFAOYSA-N/Raman>, ensuring traceability between the knowledge graph and its originating records in the research data repository.

Access to the graph is provided through a public SPARQL endpoint at <https://ditrare.ise.fiz-karlsruhe.de/chemotion-kg/sparql>. This architecture combines automated metadata harvesting, ontology-based semantic enrichment, and fine-grained provenance modeling to create a high-quality, AI-ready knowledge graph for experimental chemistry. The constructed Chemotion-KG as of July 2025 comprises a total of 1,462,187 RDF triples, reflecting the semantic integration of experimental chemistry data and metadata. The graph contains 87,782 instantiated entities, including 20,701 datasets, 20,563 studies, and 3,746 molecular entities. The knowledge graph models 250 individual creators with explicit provenance, and integrates chemical information through 4,923 instances of obo:CHEBI_59999 for generic chemical substances.

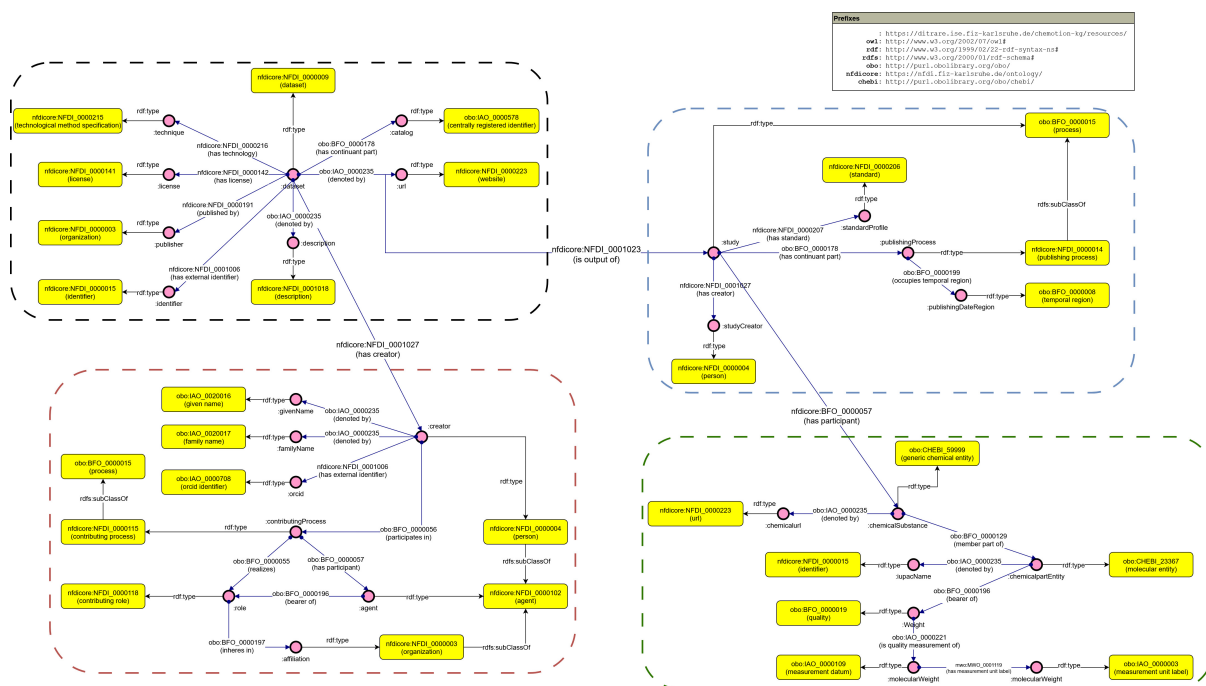


Figure 7: Integrated view of the Chemotion Knowledge Graph.

4. Impact and Future Work

The Chemotion-KG establishes a semantically enriched, BFO-compliant representation of experimental chemistry data derived from the Chemotion repository. By combining schema.org-based metadata with NFDICore and ChEBI ontologies through a SPARQL CONSTRUCT-driven pipeline, the approach ensures semantic interoperability, provenance preservation, and reasoning support. The daily ingestion workflow and ontology design patterns for datasets, creators, studies, and chemical substances provide a scalable foundation for AI-driven discovery and integration with external resources. The public SPARQL endpoint further facilitates community access, reuse, and advancement of FAIR chemical research data.

Future work will focus on the inclusion of additional information from the Chemotion repository model (e.g. dataset details and details about the chemical substance), linking the Chemotion-KG with external chemical knowledge graphs and databases, including PubChem RDF¹⁰ [16], ChemSpider¹¹ [17], and the NFDI4Chem [18] Knowledge Graph¹² to enhance data interoperability and cross-resource reasoning. Future work also includes the development of SHACL shapes to systematically validate the RDF graphs generated in the Chemotion-KG pipeline. Additional efforts will explore leveraging the Chemotion-KG for automated knowledge extraction and enrichment by integrating it with heterogeneous data sources and applying machine learning techniques for semantic alignment and knowledge discovery. Furthermore, the integration of large language models (LLMs) with structured scientific data will be investigated to support advanced query answering.

The overarching goal is to prepare the Chemotion system to fully leverage the potential of AI-based support and to enable its integration into the broader ecosystem of semantically enriched RDM systems. As part of this effort, artificial intelligence methods, including natural language processing, are being explored to automate data curation, particularly in the context of reaction descriptions and analysis modules within the Chemotion repository [19, 20]. The development of fully automated curation workflows is expected to significantly enhance data quality and streamline system performance.

¹⁰<https://pubchem.ncbi.nlm.nih.gov/docs/rdf-federated-query>

¹¹<https://www.chemspider.com/>

¹²https://knowledgebase.nfdi4chem.de/knowledge_base/

Ultimately, these advancements will contribute to the realization of AI-assisted chemistry and support the emergence of self-driving laboratories [21], which rely heavily on the automated analysis of chemical data.

Merging chemistry with AI has an immense potential for research. Studying this process and its effects is one of the foci of the Leibniz Science Campus DiTraRe. On a larger scale, DiTraRe brings an interdisciplinary perspective by investigating the influence of digitalisation of research. It sheds light on potentials, challenges, and risks of the digital transformation through an interactive exchange between numerous research areas. Activities of the dimension “Exploration and Knowledge Organisation” in the Chemotion use case are strikingly similar to efforts with which researchers deal in other disciplines. More than the others, material science could serve here as a prime example, as it is already advancing the application of ontologies and knowledge graphs for experimental research [22, 23]. The corresponding German National Research Data Infrastructure, NFDIMatWerk, has already created a rich publication database concerning many topics which can be related to Chemotion-KG [24]. DiTraRe plans to serve here as a platform connecting researchers from varying disciplines with an aim to exchange ideas, identify common problems, and create generalised solutions. This way, by raising above the level of individual research disciplines, we will study influence and effects of applied AI as an additional dimension to the digital transformation of research.

5. Summary

In this paper, a comprehensive pipeline for constructing the BFO-compliant Chemotion Knowledge Graph has been presented, integrating experimental chemistry data from the Chemotion Repository into a semantically enriched, ontology-driven representation. By leveraging ontology design patterns and aligning schema-based metadata with NFDICore and ChEBI ontologies, the approach establishes a reusable and interoperable framework for chemical research data. The generated knowledge graph comprises over 1.4 million triples and more than 78,000 instances.

Acknowledgments

The Leibniz Science Campus “Digital Transformation of Research” (DiTraRe) is funded by the Leibniz Association (W74/2022). The authors would like to thank Pei-Chi Huang and Chia-Lin Lin (KIT IBCS) for their work on JSON-LD in Chemotion.

Declaration on Generative AI

During the preparation of this work, the author(s) used GPT-4 in order to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication’s content.

References

- [1] C. Steinbeck, O. Koepler, F. Bach, et al., Nfdi4chem-towards a national research data infrastructure for chemistry in germany, Research ideas and outcomes 6 (2020) e55852.
- [2] P. Tremouilhac, A. Nguyen, Y.-C. Huang, et al., Chemotion eln: an open source electronic lab notebook for chemists in academia, Journal of Cheminformatics 9 (2017) 54.
- [3] S. Kotov, P. Tremouilhac, N. Jung, et al., Chemotion-eln part 2: adaption of an embedded ketcher editor to advanced research applications., Journal of Cheminformatics 10 (2018). URL: <https://doi.org/10.1186/s13321-018-0292-9>. doi:10.1186/s13321-018-0292-9.
- [4] P. Tremouilhac, C.-L. Lin, P.-C. a. a. Huang, The repository chemotion: Infrastructure for sustainable research in chemistry, Angewandte Chemie In-

- ternational Edition 59 (2020) 22771–22778. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.202007702>. doi:<https://doi.org/10.1002/anie.202007702>. arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.202007702>.
- [5] P. Tremouilhac, P.-C. Huang, C.-L. Lin, Y.-C. Huang, A. Nguyen, N. Jung, F. Bach, S. Bräse, Chemotion repository, a curated repository for reaction information and analytical data, *Chemistry methods* 1 (2021) 8–11. doi:10.1002/cmt.d.202000034.
 - [6] P. Huang, C. Lin, P. Tremouilhac, et al., Using the chemotion repository to deposit and access fair research data for chemistry experiments, *Nature Protocols* 20 (2025). URL: <https://doi.org/10.1038/s41596-024-01074-z>. doi:10.1038/s41596-024-01074-z.
 - [7] Y. Huang, P. Tremouilhac, A. Nguyen, et al., hemspectra: a web-based spectra editor for analytical data, *Journal of Cheminformatics* 18 (2021). URL: <https://doi.org/10.1186/s13321-020-00481-0>. doi:10.1186/s13321-020-00481-0.
 - [8] A. Gangemi, V. Presutti, Ontology design patterns, in: *Handbook on ontologies*, Springer, 2009, pp. 221–243.
 - [9] M. Razum, F. Bach, S. Brünger-Weilandt, et al., Proposal for a Leibniz ScienceCampus – Digital Transformation of Research (DiTraRe), 2023. doi:10.5281/zenodo.11109406, project proposal.
 - [10] A. M. Jacyszyn, F. Bach, H. Sack, et al., Interim report of the leibniz science campus "digital transformation of research" (ditrare), 2025. URL: <https://doi.org/10.5281/zenodo.14941635>. doi:10.5281/zenodo.14941635.
 - [11] A. M. Jacyszyn, H. Sack, D.-S. Group, et al., Ditrare: Ai on a spider's web. interweaving disciplines for digitalisation, in: *4th International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment*, volume Vol-3780, 2024. URL: <https://ceur-ws.org/Vol-3780/paper5.pdf>. doi:10.5281/zenodo.13862017.
 - [12] R. Arp, B. Smith, A. D. Spear, *Building ontologies with Basic Formal Ontology*, MIT Press, 2015.
 - [13] O. Bruns, T. Tietz, J. Waitelonis, et al., Nfdicore 2.0: A bfo-compliant ontology for multi-domain research infrastructures, *arXiv preprint arXiv:2410.01821* (2024).
 - [14] P. Del Nostro, G. Goldbeck, D. Toti, Chameo: An ontology for the harmonisation of materials characterisation methodologies, *Applied Ontology* (2022) 1–21.
 - [15] R. Falco, A. Gangemi, S. Peroni, et al., Modelling owl ontologies with graffoo, in: *European Semantic Web Conference*, Springer, 2014, pp. 320–325.
 - [16] S. Kim, J. Chen, T. Cheng, et al., Pubchem 2025 update, *Nucleic Acids Research* 53 (2024) D1516–D1525. URL: <https://doi.org/10.1093/nar/gkae1059>. doi:10.1093/nar/gkae1059. arXiv:<https://academic.oup.com/nar/article-pdf/53/D1/D1516/60743708/gkae1059.pdf>.
 - [17] H. E. Pence, A. Williams, Chemspider: An online chemical information resource, *Journal of Chemical Education* 87 (2010) 1123–1124. URL: <https://doi.org/10.1021/ed100697w>. doi:10.1021/ed100697w. arXiv:<https://doi.org/10.1021/ed100697w>.
 - [18] S. Neumann, A.-C. Andres, F. Bach, et al., Interim report nfdi4chem 2023, *Research Ideas and Outcomes* 10 (2024) e124977. URL: <https://doi.org/10.3897/rio.10.e124977>. doi:10.3897/rio.10.e124977. arXiv:<https://doi.org/10.3897/rio.10.e124977>.
 - [19] D. Punjabi, Y. Huang, L. Holzhauser, et al., Infrared spectrum analysis of organic molecules with neural networks using standard reference data sets in combination with real-world data, *Journal of Cheminformatics* 17 (2025). URL: <https://doi.org/10.1186/s13321-025-00960-2>. doi:10.1186/s13321-025-00960-2.
 - [20] Y.-C. Huang, P. Tremouilhac, S. Kuhn, et al., (semi-) automatic review process for common compound characterization data in organic synthesis, *ChemRxiv* (2024). doi:10.26434/chemrxiv-2024-1r9tb, preprint.
 - [21] P. M. Maffettone, P. Friederich, S. G. Baird, et al., What is missing in autonomous discovery: open challenges for the community, *Digital Discovery* 2 (2023) 1644–1659. URL: <http://dx.doi.org/10.1039/D3DD00143A>. doi:10.1039/D3DD00143A.
 - [22] 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* n/a (????) 2401534. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/adem.202401534>. doi:<https://doi.org/10.1002/adem.202401534>. arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/adem.202401534>.

- [23] 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 (SeMatS 2024): Harnessing the Power of Semantic Web Technologies in Materials Science*, volume 3760 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 78–100. URL: <https://ceur-ws.org/Vol-3760/paper3.pdf>.
- [24] *Proceedings NFDI-MatWerk Conference 2023*, Zenodo, 2024. URL: <https://doi.org/10.5281/zenodo.11353339>. doi:10.5281/zenodo.11353339.