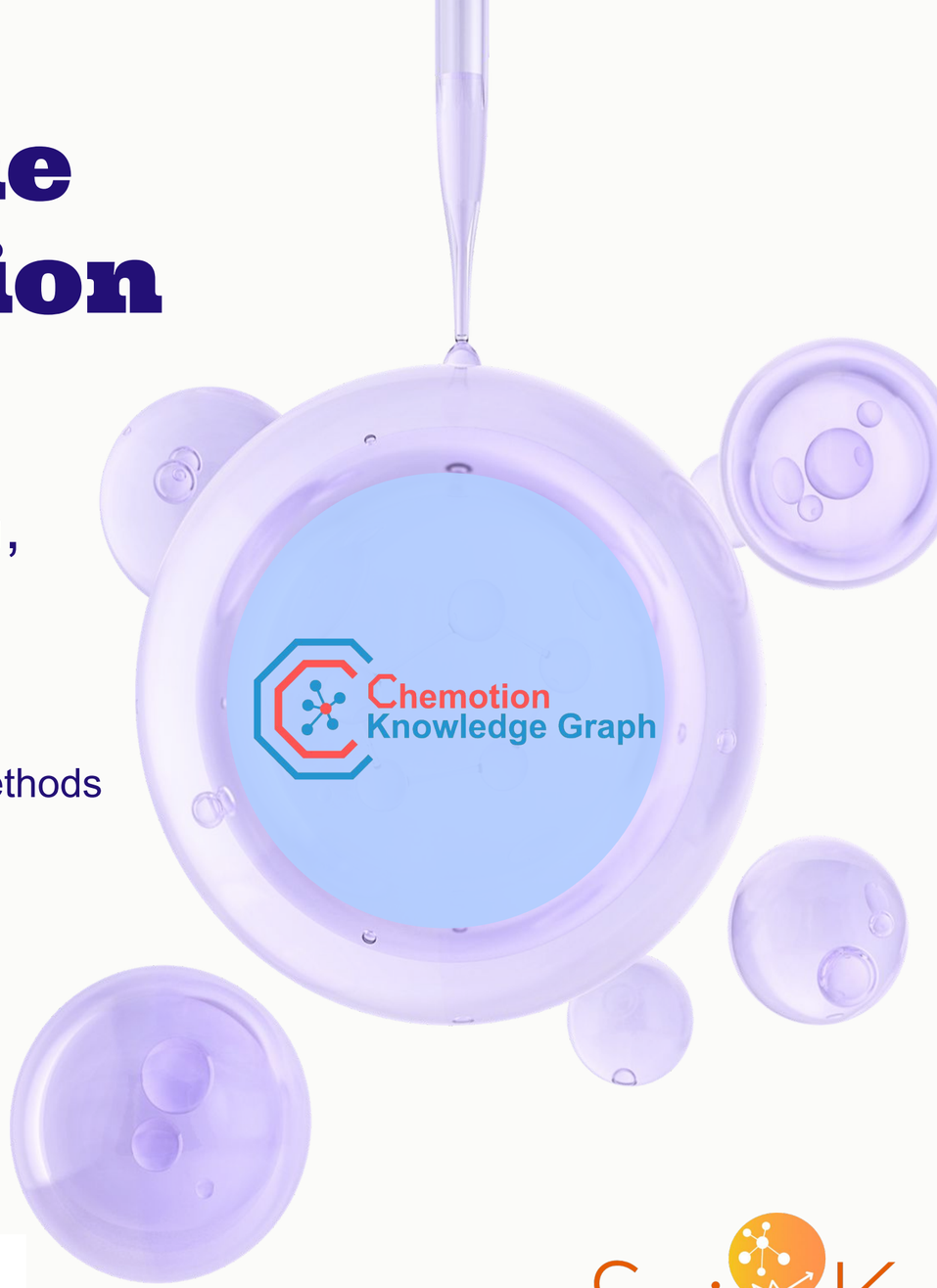
AI4DiTraRe: Building the BFO-Compliant Chemotion Knowledge Graph

Ebrahim Norouzi, Nicole Jung, Anna M. Jacyszyn, Jörg Waitelonis and Harald Sack

FIZ Karlsruhe & KIT(Institute of Applied Informatics and Formal Description Methods and Institute of Biological and Chemical Systems)

Sci-K 2025, 5th International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment @ ISWC, Nov 2nd, 2025, Nara, Japan



FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

KIT
Karlsruher Institut für Technologie

Leibniz ScienceCampus
Digital Transformation of Research

Sci-K

Agenda

- Motivation
- Knowledge Graph Construction Approach
- Ontology Integration, Why?
- Ontology Design Patterns (ODPs), Why?
- Results & Statistics
- Limitations
- Future Work & Summary

Leibniz ScienceCampus
Digital Transformation of Research

DiTraRe

Leibniz Science Campus *Digital Transformation of Research*

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

+

KIT
Karlsruhe Institute of Technology

- Growth core to **establish new research branch**.
- Consortium: **4+4 years** (start: September 2023).
- **Multilevel** approach. → We start with 4 **interdisciplinary** use cases.
- **AI4DiTraRe**: solutions based on different AI methods to solve research questions in numerous areas of research. Study overall **influence and effects of AI**.

This work (FIZ-Karlsruhe & KIT Institute of Biological and Chemical Systems):

Chemotion Electronic Lab Notebook + AI

Motivation

- Increasing complexity of chemistry workflows, multimodal data (spectra, reactions, metadata)
- **Chemotion Repository:** rich chemistry data (reactions, spectra, studies, datasets)
  - Schema.org-based structure too shallow for reasoning or linking
  - No semantic alignment with other domain ontologies
  - Hard to **connect/query** across external resources

Chemotion Repository

Goal

- Transform Chemotion repository into a machine-understandable Knowledge Graph (KG)
- Enable:
  - Reasoning: infer hidden relations (e.g., link identical chemicals across studies)
  - Provenance tracking: trace how data evolved or who contributed
  - Cross-KG integration: connect Chemotion with NFDI4Chem, PubChem, MatWerk, etc.
  - Provide semantically enriched chemistry data for discovery and reuse

Introduction: Why we need top-level ontologies?

Common foundation

Enable integration across different domain ontologies

Consistency

Avoid ambiguity by defining core categories (e.g., object, process, quality)

Facilitate reasoning

Logical structure helps automated reasoning and validation

Interoperability

Allow systems and datasets to “speak the same language”

Reusability

Save time and effort by building on established upper-level concepts

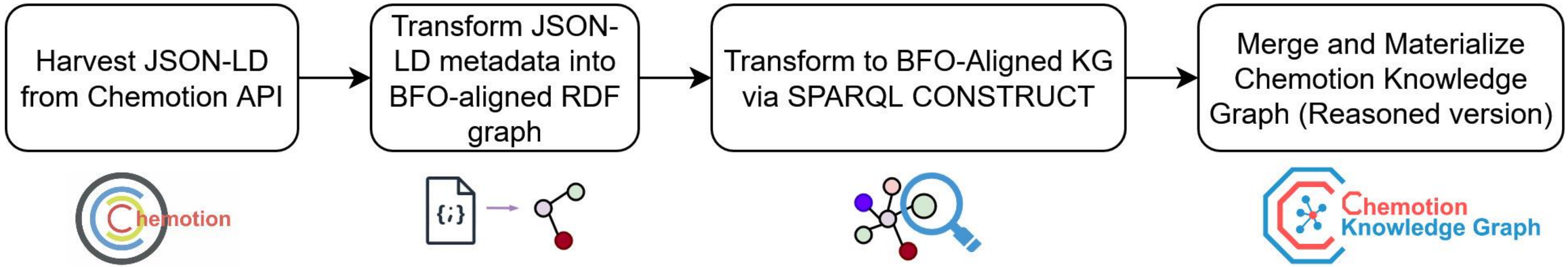
Introduction: What is BFO?

- Basic Formal Ontology (BFO) [1] ⟶ one of the leading upper-level ontology
- Widely adopted, philosophically grounded in realism
- Ensures interoperability & reusability across domains
- Clear and consistent class hierarchy
- Stable across versions ⟶ trust and adoption
- Foundation for many domain ontologies

[1] BFO guidebook: https://direct.mit.edu/books/monograph/4044/Building-Ontologies-with-Basic-Formal-Ontology

Knowledge Graph Construction Approach

- Source: Chemotion Repository (schema.org structure): lacks upper-level semantics, Incompatible with external ontologies and reasoning
- Target: BFO-compliant, ontology-aligned RDF knowledge graph
- Objective: semantic enrichment → reasoning, integrate with external KGs, enable cross-domain queries
- Ontology Design Patterns (ODPs) applied during schema-to-ontology mapping stage



Ontology Integration & Selection Rationale

- Upper-level alignment:
  - BFO – Basic Formal Ontology (provides upper-level semantics for interoperability)
  - Widely adopted in life sciences & research data infrastructures
  - Enables logical entailment and cross-domain reasoning
- Middle-level alignment:
  - NFDICore Ontology (provides modeling for representing metadata related to resources, including individuals, organizations, projects, data portals, and more) - Already BFO-aligned
  - Acts as a bridge layer between abstract BFO classes and domain-specific ontologies.
  - Developed by FIZ Karlsruhe for representing resources in the National Research Data Infrastructure (NFDI) initiatives & facilitates interoperability within NFDI ecosystems (e.g., NFDI4Culture, NFDIMatWerk)
- Domain-level alignment:
  - ChEBI – Chemical Entities of Biological Interest (provides controlled vocabulary for chemical substances and molecular entities) - Already BFO-aligned

Ontology Design Patterns [1]

- Reusable, modular solutions for recurring ontology modeling problems.
- Use of ODPs:
  - Applied during schema-to-ontology mapping stage
  - Ensure uniform modeling logic across Chemotion entities
  - Provide reusable templates for recurring concepts
  - Serve as building blocks for Chemotion-KG modular structure

[1] Gangemi, A., & Presutti, V. (2009). Ontology design patterns. In Handbook on ontologies (pp. 221-243). Berlin, Heidelberg: Springer Berlin Heidelberg.

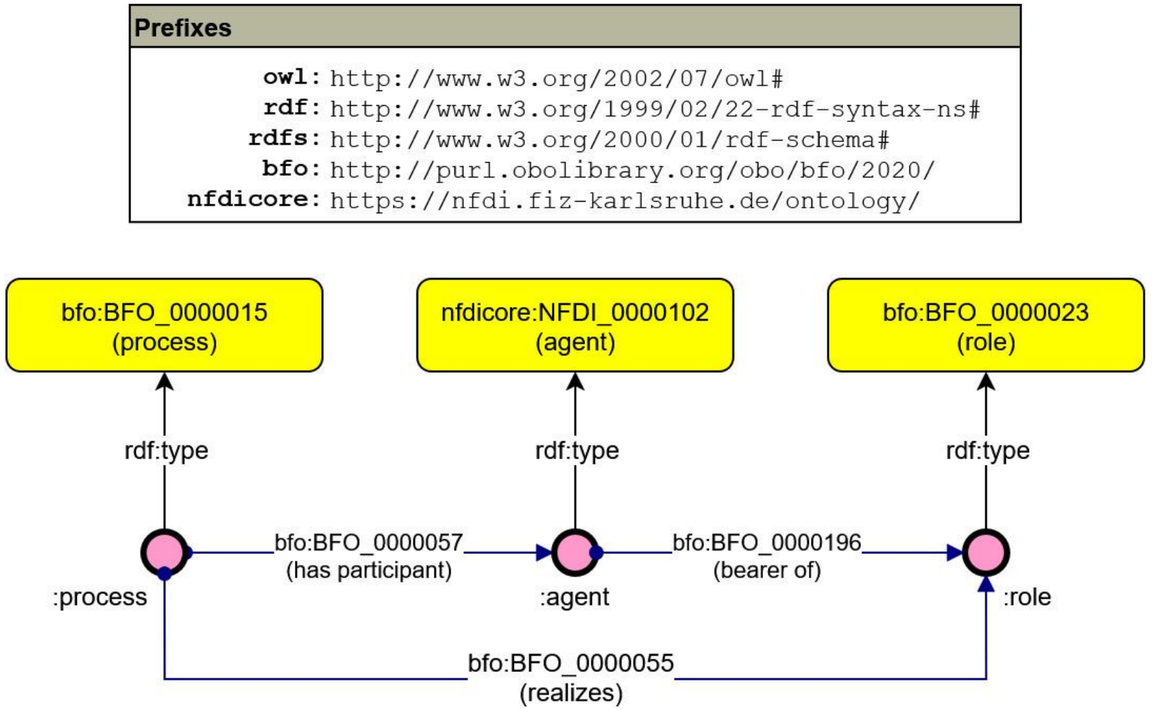
SPARQL CONSTRUCT

- Example: schema:Dataset ⟶ nfdicore:NFDI_0000009 alignment
- Preserve original values + add explicit BFO semantics

```
CONSTRUCT {
  ?dataset a nfdicore:NFDI_0000009 ;
        nfdicore:NFDI_0001027 ?creator ;
        obo:IAO_0000235 ?descriptionNode ;
        ...
}
WHERE {
  ?dataset a schema:Dataset ;
        schema:creator ?creator ;
        ...
}
```

Ontology Design Patterns (ODPs): Concept & Role

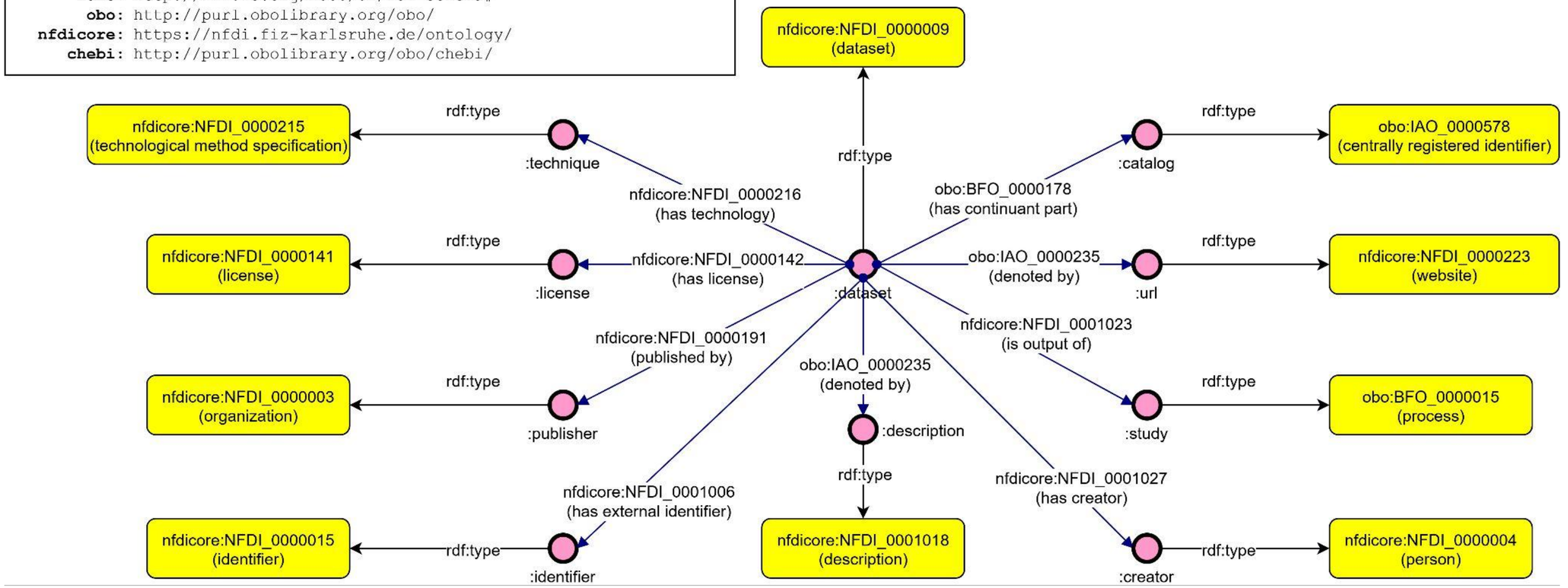
- Core ODP linking bfo:Process ↔ nfdicore:Agent ↔ bfo:Role
- Relations: has participant / bearer of / realizes
- Enables provenance and role attribution for research activities
- Used for creators and organizational affiliations



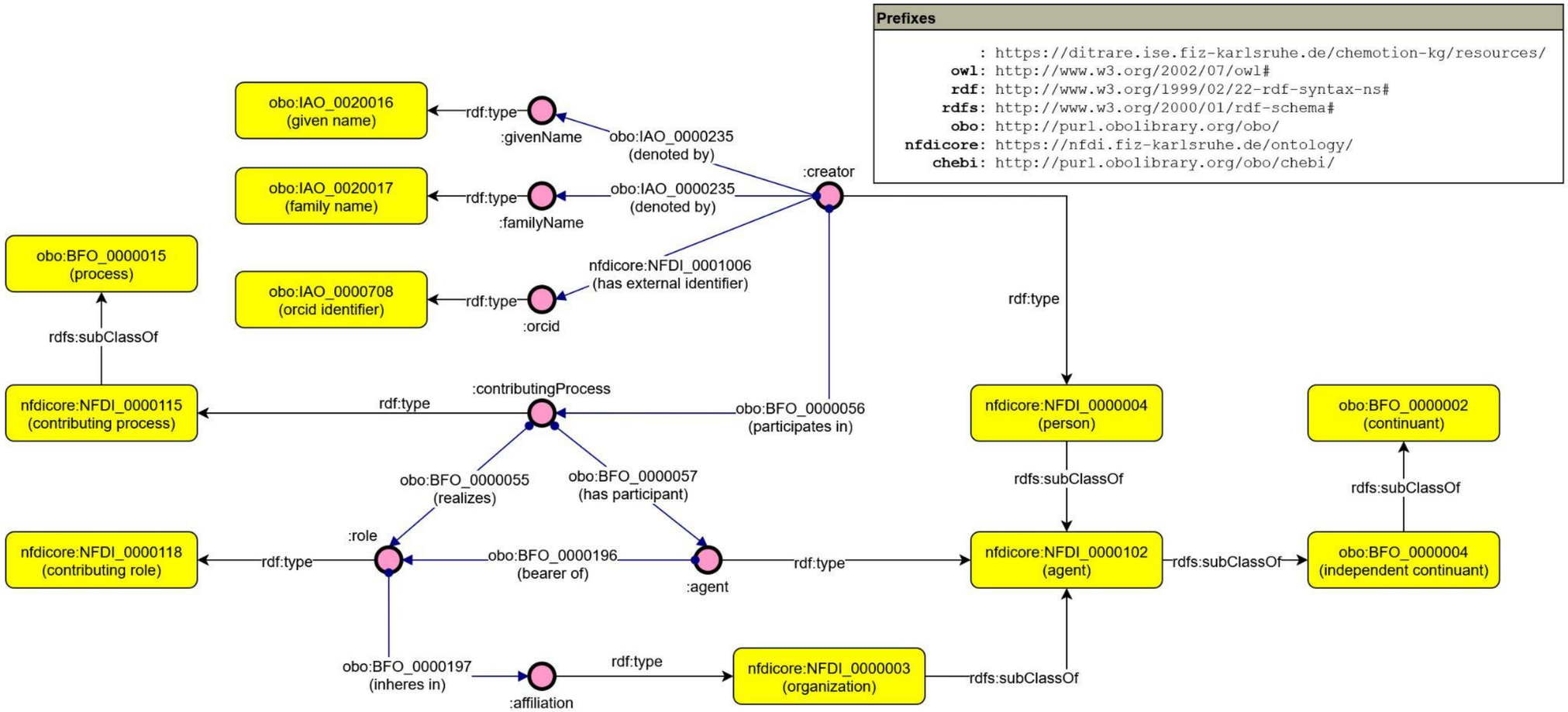
Dataset Pattern

Prefixes

```
        : https://ditrare.ise.fiz-karlsruhe.de/chemotion-kg/resources/
  owl: http://www.w3.org/2002/07/owl#
  rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
 rdfs: http://www.w3.org/2000/01/rdf-schema#
  obo: http://purl.obolibrary.org/obo/
nfdicore: https://nfdi.fiz-karlsruhe.de/ontology/
chebi: http://purl.obolibrary.org/obo/chebi/
```



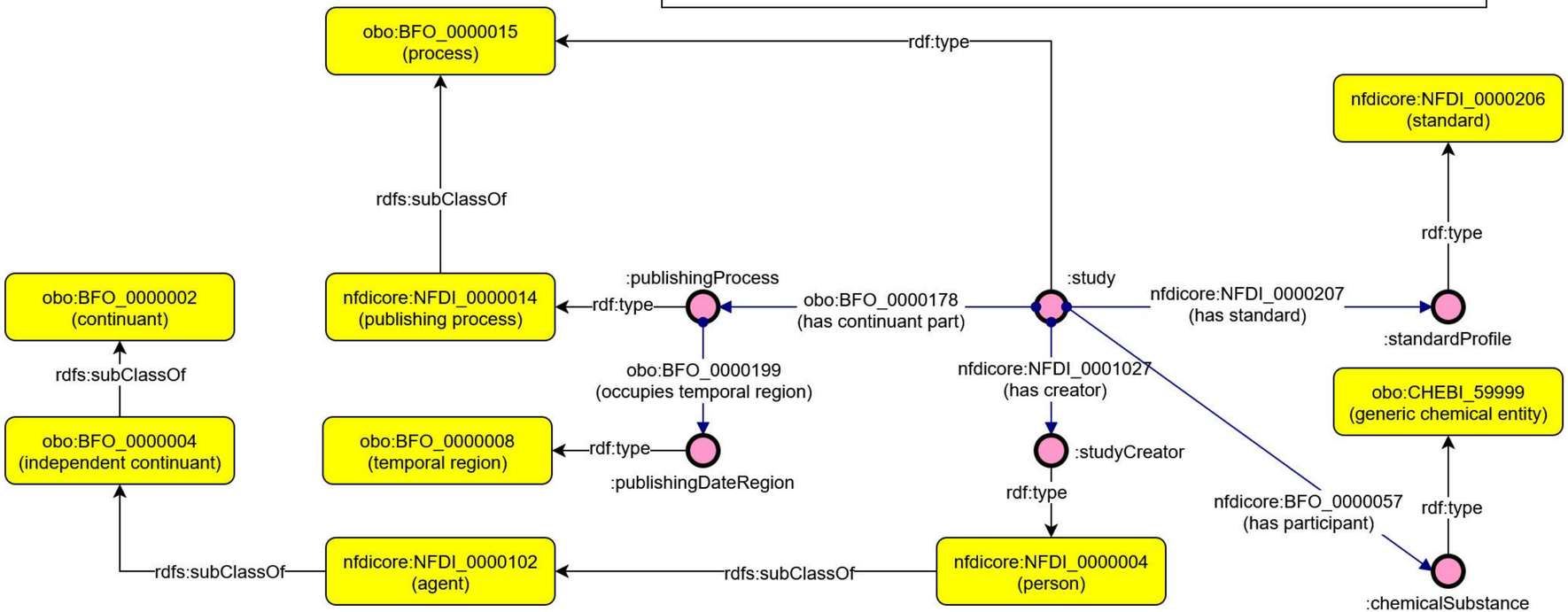
Creator Pattern



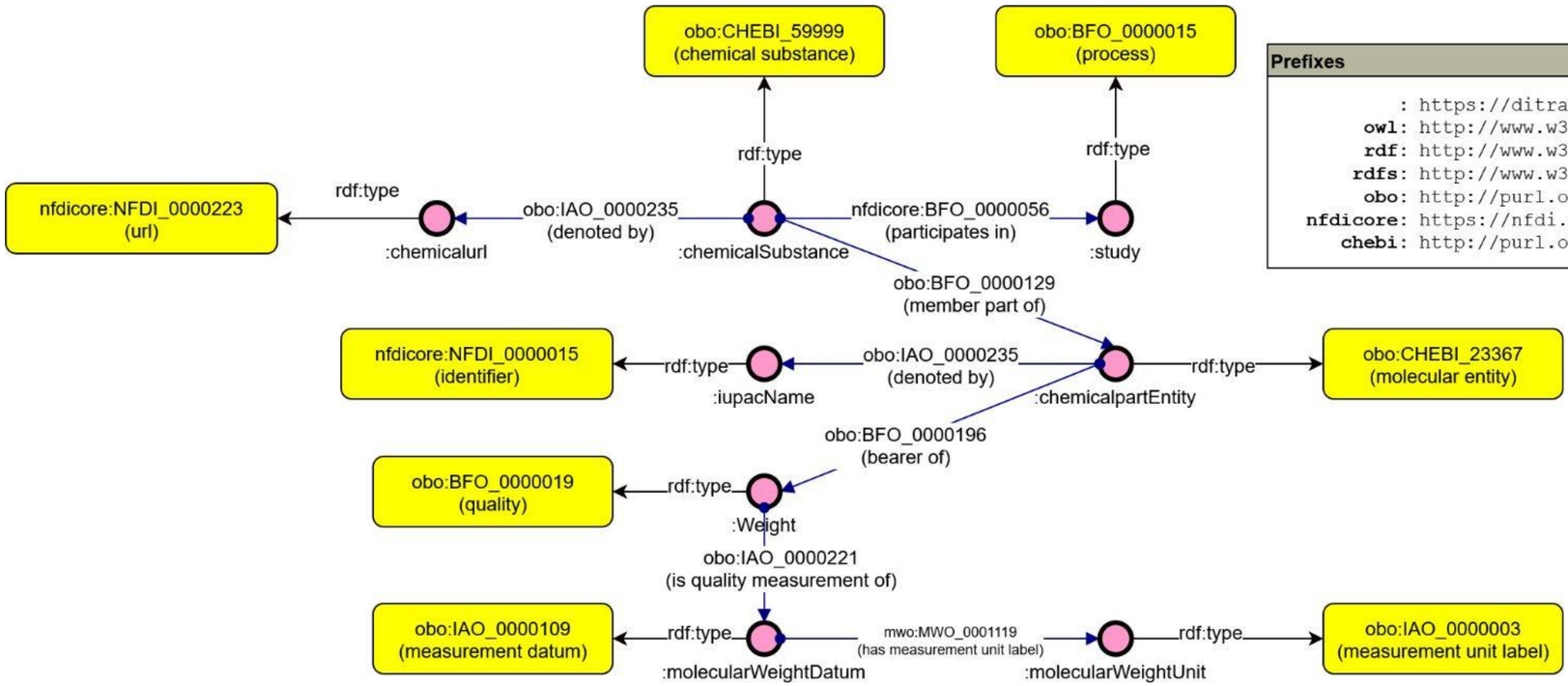
Study Pattern

Prefixes

```
      : https://ditrare.ise.fiz-karlsruhe.de/chemotion-kg/resources/
    owl: http://www.w3.org/2002/07/owl#
    rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
   rdfs: http://www.w3.org/2000/01/rdf-schema#
    obo: http://purl.obolibrary.org/obo/
nfdicore: https://nfdi.fiz-karlsruhe.de/ontology/
  chebi: http://purl.obolibrary.org/obo/chebi/
```



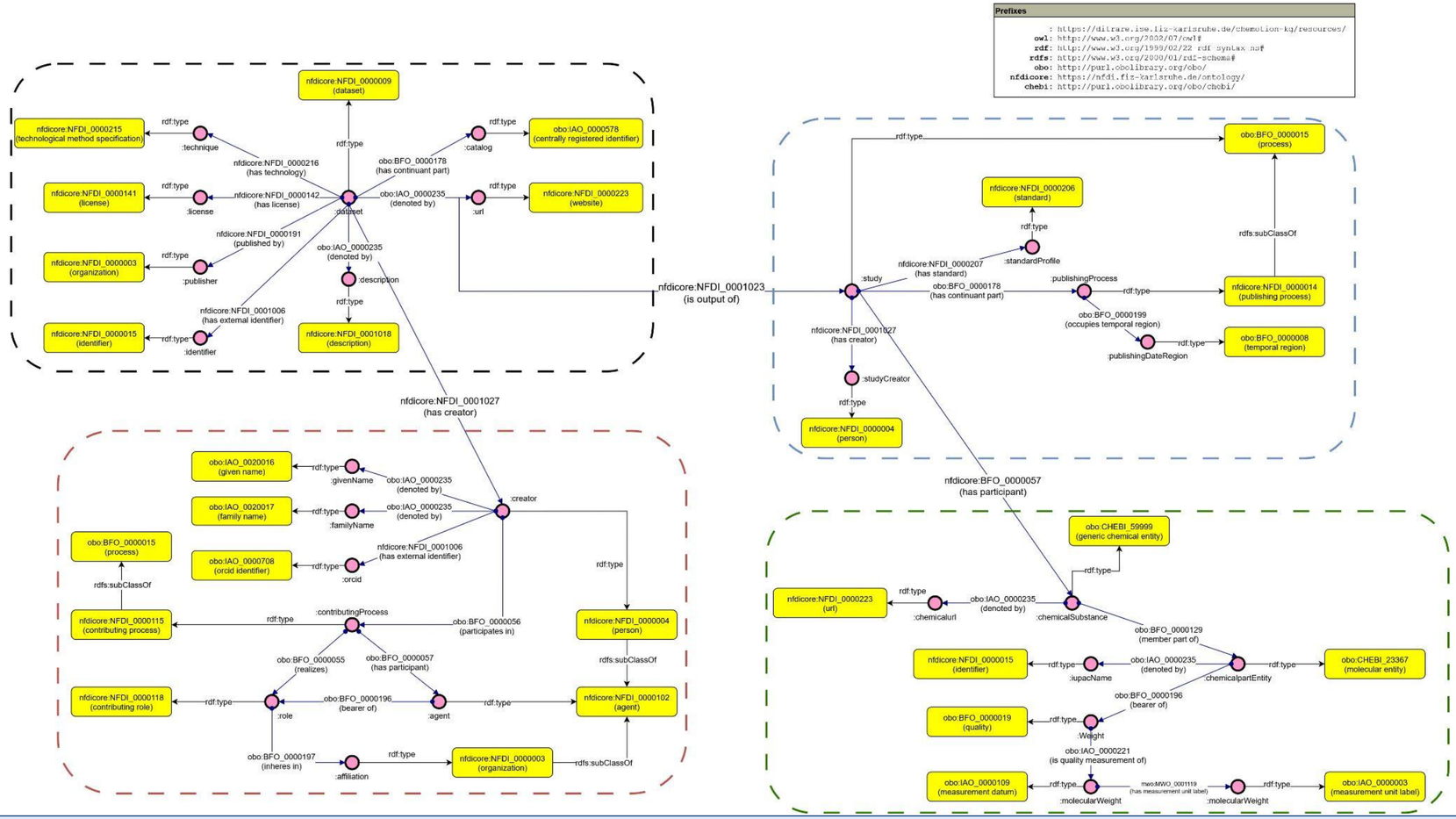
Chemical Substance Pattern



Prefixes

```
        : https://ditrare.ise.fiz-karlsruhe.de/chemotion-kg/resources/
      owl: http://www.w3.org/2002/07/owl#
      rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
     rdfs: http://www.w3.org/2000/01/rdf-schema#
      obo: http://purl.obolibrary.org/obo/
nfdicore: https://nfdi.fiz-karlsruhe.de/ontology/
    chebi: http://purl.obolibrary.org/obo/chebi/
```

Integrated Graph View



Knowledge Graph Statistics

- Total triples: 1,462,187
- Total instances: 87,782
  - Datasets 20,701
  - Studies 20,563
  - Chemicals 3,746
  - Creators 250
  - Chemical substances (obo:CHEBI_59999): 4,923 instances

Chemotion Knowledge Graph

- Public SPARQL endpoint: https://ditrare.ise.fiz-karlsruhe.de/chemotion-kg/sparql

Limitations

- Scope restriction
  - Current KG models metadata layer only -> Raw instrument outputs (e.g., NMR, MS, IR spectral files) not yet semantically integrated

- Canonical IRIs generated by concatenating Chemotion resource IDs (e.g. “https://ditrare.ise.fiz-karlsruhe.de/chemotion-kg/resources/2025-07/Dataset_1e332f08-cd6d-41b9-a6b8”) are Long.

Future Work & Summary

- Extend KG ⟶ include instrument and measurement data
- Cross-link to PubChem RDF, ChemSpider, NFDI4Chem
- Develop SHACL validations
- Evaluate user queries and domain relevance
- Integrate LLMs semantic alignment and data enrichment
- **Chemotion-KG = ontology-driven, BFO-compliant chemistry knowledge graph**
- **SPARQL CONSTRUCT + ODPs ⟶ modular semantic integration**

Leibniz ScienceCampus
Digital Transformation of Research

DiTraRe

Leibniz Science Campus *Digital Transformation of Research*

- Stay **connected** with DiTraRe
  - Website: www.ditrare.de/en
  - Email: ditrare@fiz-karlsruhe.de
  - LinkedIn: www.linkedin.com/company/ditrare
  - Mastodon: social.kit.edu/@DiTraRe
  - YouTube: www.youtube.com/@DiTraRe
  - Zenodo: zenodo.org/communities/ditrare
- **Symposium** on Digitalisation of Research: 2-3 December 2025, Karlsruhe
- Monthly **Interdisciplinary Colloquium** on Digitalisation of Research

www.ditrare.de/en

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Karlsruhe Institute of Technology

funded by Leibniz Competition

Thank you very much for your attention!

Leibniz ScienceCampus
Digital Transformation of Research

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

KIT
Karlsruher Institut für Technologie

ebrahim.norouzi@fiz-Karlsruhe.de

https://sigmoid.social/@enorouzi

https://ebrahimnorouzi.github.io/

https://ditrare.ise.fiz-karlsruhe.de/chemotion-kg/