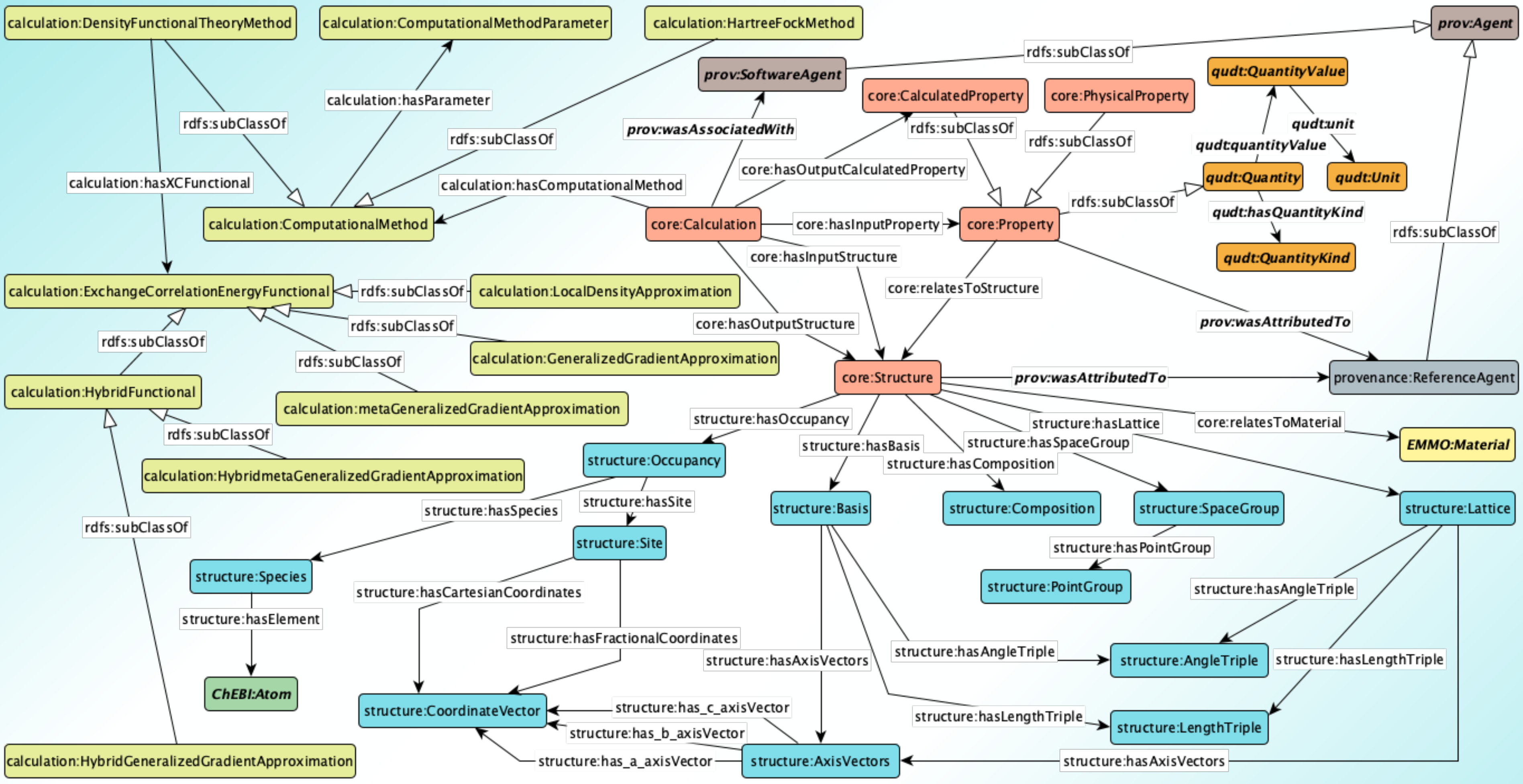
An Ontology for the Materials Design Domain

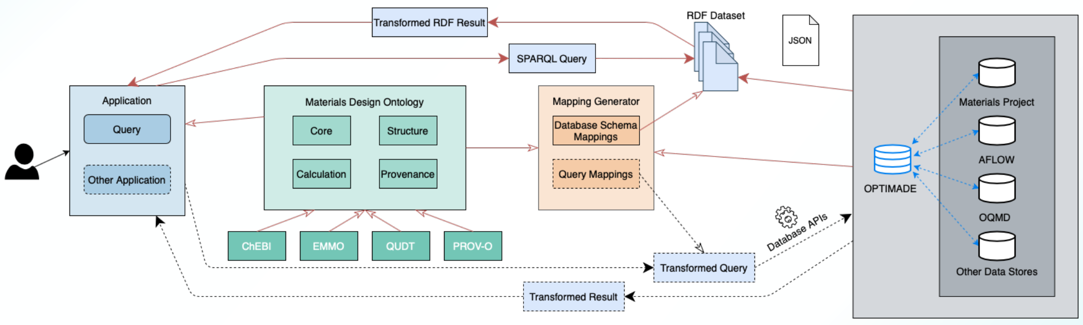
Huanyu Li, Rickard Armiento, Patrick Lambrix
The Swedish e-Science Research Centre & Linköping University, Linköping, Sweden

Motivation

In the materials science field, data-driven techniques have the potential to accelerate the discovery and design of new materials. Therefore, a large number of research groups and communities have developed data-driven workflows, including data repositories [1]. In the materials design domain, much of the data from materials calculations are stored in different heterogeneous databases. Such databases usually have different data models. It is challenging for users to find data and integrate data from multiple sources. To address such challenges and make data FAIR, ontologies and ontology based techniques can play a significant role. Therefore, we present the Materials Design Ontology (MDO) which defines concepts and relationships to cover knowledge in the field of materials design.



MDO Usage



✓ What are the materials of which the value of band gap is higher than 5eV?
- The whole data set contains 85 stable materials from [3].

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX core: <https://w3id.org/mdo/core/>
PREFIX structure: <https://w3id.org/mdo/structure/>
PREFIX qudt: <http://qudt.org/schema/qudt/>

SELECT ?formula ?value WHERE {
  ?calculation rdf:type core:Calculation;
               core:hasOutputCalculatedProperty ?property;
               core:hasOutputStructure ?output_structure.
  ?property qudt:quantityValue ?quantity_value;
            core:hasPropertyName ?name.
  ?quantity_value rdf:type qudt:QuantityValue;
                  qudt:numericValue ?value.
  ?output_structure structure:hasComposition ?composition.
  ?composition structure:hasDescriptiveFormula ?formula.
  FILTER (?value>5 && ?name="band_gap")
}
```

← SPARQL query

SPARQL query result →

formula	value
$Cs_2Rb_1In_1F_6$	5.3759
$Cs_2Rb_1Ga_1F_6$	5.9392
$Cs_2K_1In_1F_6$	5.4629
$Rb_2Na_1In_1F_6$	5.2687
$Cs_2Rb_1Ga_1F_6$	5.5428
$Rb_2Na_1Ga_1F_6$	5.9026
$Cs_2K_1Ga_1F_6$	6.0426

MDO

✓ 4 main Components
- Core: the top-level concepts and relationships
- Structure: the structural information of materials
- Calculation: the classification of different computational methods
- Provenance: the provenance information of materials data and calculation

✓ Inspiration from OPTIMADE [2]

✓ Reuse of ontological resources
- EMMO (European Materials & Modelling Ontology)
- ChEBI (Chemical Entities of Biological Interest)
- QUDT (Quantities, Units, Dimensions and Data Types)
- PROV-O (The PROV ontology)

✓ Analysis of non-ontological resources
- CIF, International Tables for Crystallography
- APIs of well known materials databases

✓ Discussions with domain expert

✓ Available at https://w3id.org/mdo

Future Work

✓ Refining the current ontology
- e.g., with workflows containing multiple calculations

✓ Extending the ontology with additional modules

[1] Lambrix, P., Armiento, R., Delin, A., Li, H., 2018. Big Semantic Data Processing in the Materials Design Domain. Encyclopedia of Big data Technologies, DOI:10.1007/978-3-319-63962-8_293-1
[2] Open Databases Integration for Materials Design: https://www.optimade.org
[3] Faber, F.A., Lindmaa, A., Von Lilienfeld, O.A., Armiento, R.: Machine learning energies of 2 million elpasolite (a b c 2 d 6) crystals. Physical review letters 117(13),135502 (2016)

Swedish e-Science Research Centre
LINKÖPING UNIVERSITY