Demonstrator 19: Materials Databases Integration using the Materials Design Ontology

Huanyu Li, Rickard Armiento, Olaf Hartig, Mina Abd Nikooie Pour, Ying Li, Patrick Lambrix

Linköping University
Sweden

LINKÖPING UNIVERSITY

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

Agenda

- Background and Motivation
- Materials Design Ontology (MDO)
- Ontology-driven Data Access and Data Integration
- Summary and Ongoing Work

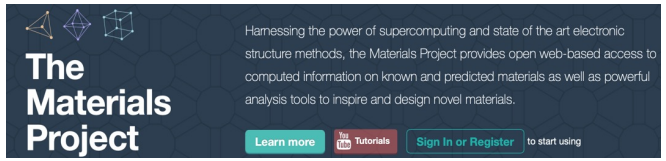
ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

Background and Motivation

LINKÖPING UNIVERSITY

Materials Design Background

- Many software programs can achieve materials calculations
- A lot of databases provide materials calculation data via Web APIs (e.g., REST APIs)
  - Materials databases are heterogenous in nature
  - APIs follow different data schemas
- A data-driven workflow of materials design will search these databases with desired combinations of properties

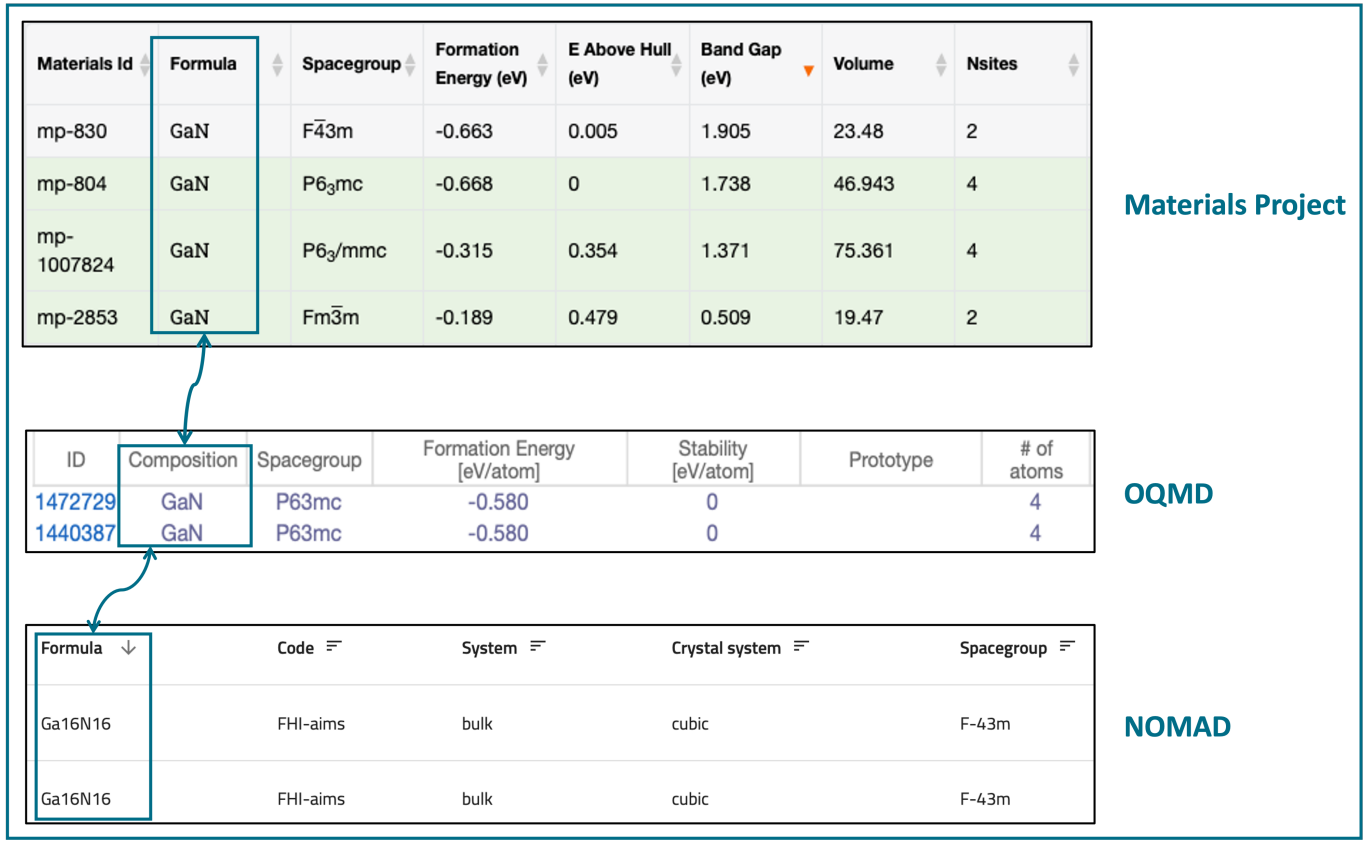


Querying Materials Databases

- Searching 'GaN' in Materials Project, OQMD, NOMAD
  - Different number of fields returned, different terminology for the same concept
  - Some semantics could be added
- OPTIMADE (Open Databases Integration for Materials Design)
  - To design a common API
  - To make materials databases interoperable
- Ontologies can help in,
  - accessing/integrating data with semantics-aware
  - standardizing terminologies
  - making data FAIR

  …

OPTIMADE
Open Databases Integration for Materials Design



Materials Id	Formula	Spacegroup	Formation Energy (eV)	E Above Hull (eV)	Band Gap (eV)	Volume	Nsites
mp-830	GaN	$F\bar{4}3m$	-0.663	0.005	1.905	23.48	2
mp-804	GaN	$P6_3mc$	-0.668	0	1.738	46.943	4
mp-1007824	GaN	$P6_3/mmc$	-0.315	0.354	1.371	75.361	4
mp-2853	GaN	$Fm\bar{3}m$	-0.189	0.479	0.509	19.47	2

Materials Project

ID	Composition	Spacegroup	Formation Energy [eV/atom]	Stability [eV/atom]	Prototype	# of atoms
1472729	GaN	P63mc	-0.580	0		4
1440387	GaN	P63mc	-0.580	0		4

OQMD

Formula	Code	System	Crystal system	Spacegroup
Ga16N16	FHI-aims	bulk	cubic	F-43m
Ga16N16	FHI-aims	bulk	cubic	F-43m

NOMAD

Motivation

- A semantics-aware and integrated way for querying data among heterogeneous data sources
  - Ontology-driven approaches are needed
  - Ontologies for representing materials design domain knowledge are needed

- There is a lack of methods accessing and integrating data over multiple heterogeneous data sources where data is shared via different ways
  - e.g., tabular data/SQL queries, JSON-formatted data/API requests
  - Classical ontology-based data access and integration (OBDA/OBDI) methods focus on relational data

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

Materials Design Ontology (MDO)

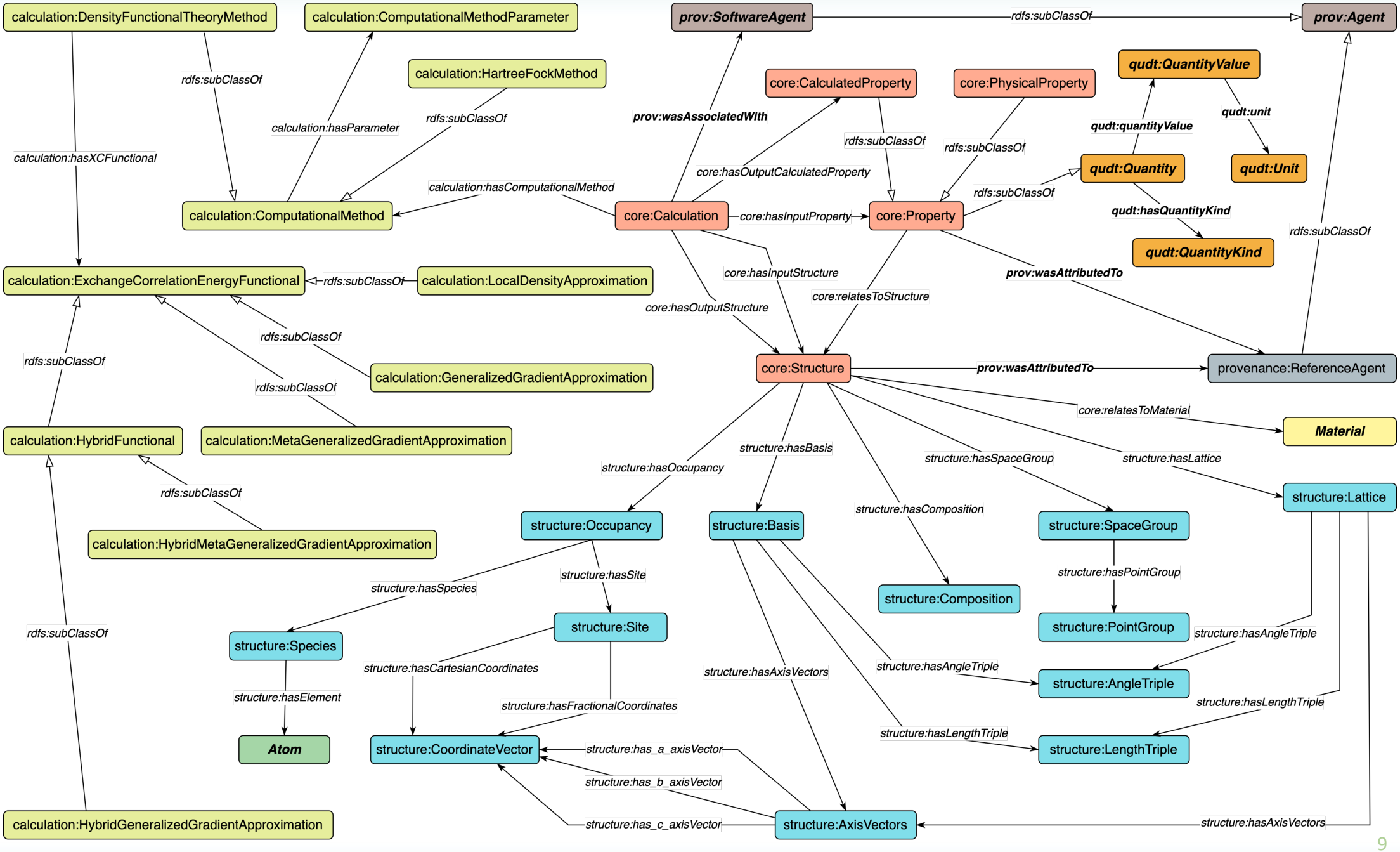
LINKÖPING UNIVERSITY

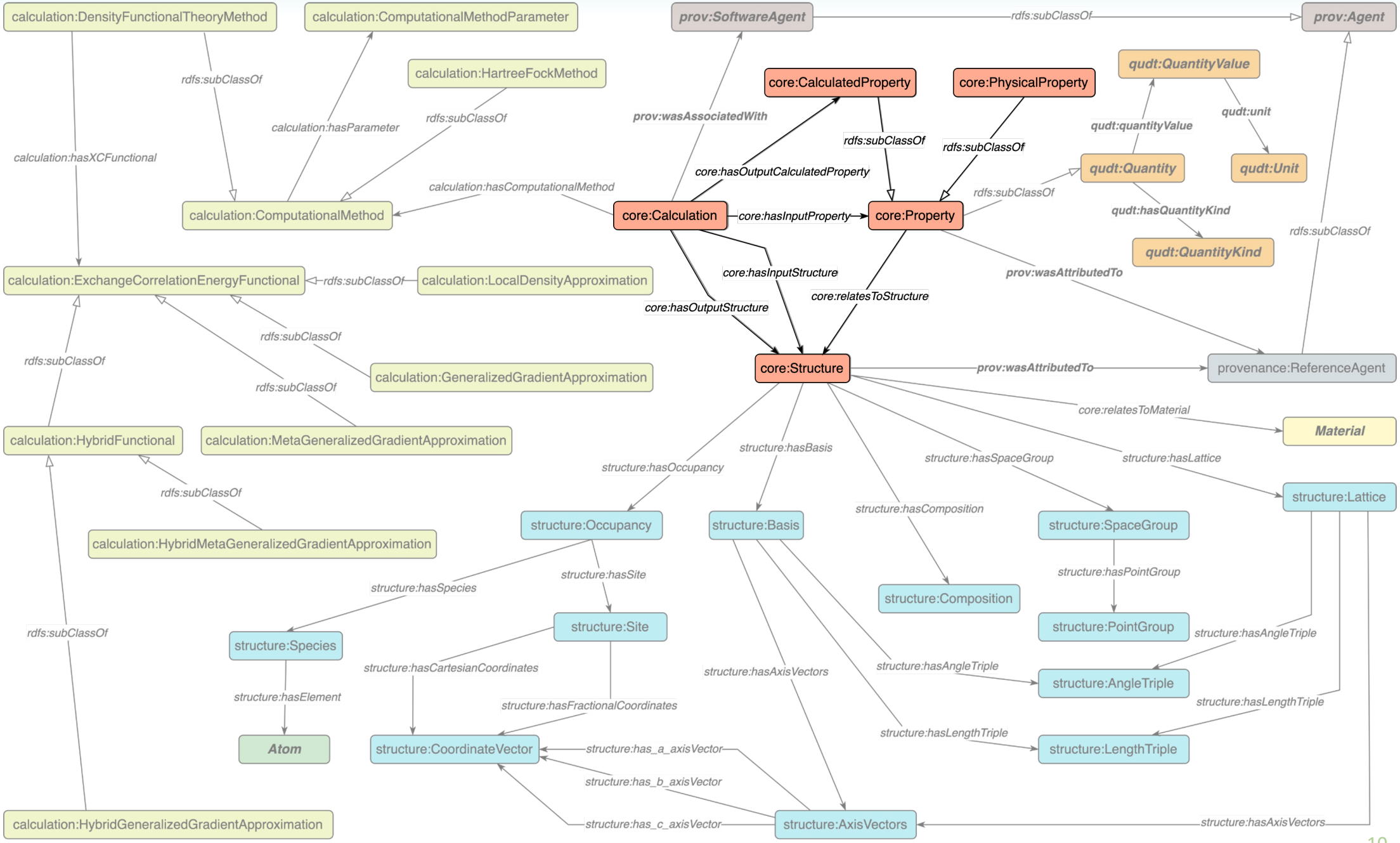
H. Li, R. Armiento and P. Lambrix, An Ontology for the Materials Design Domain, in: *The Semantic Web – ISWC 2020 – 19th International Semantic Web Conference, November 2-6, 2020.*

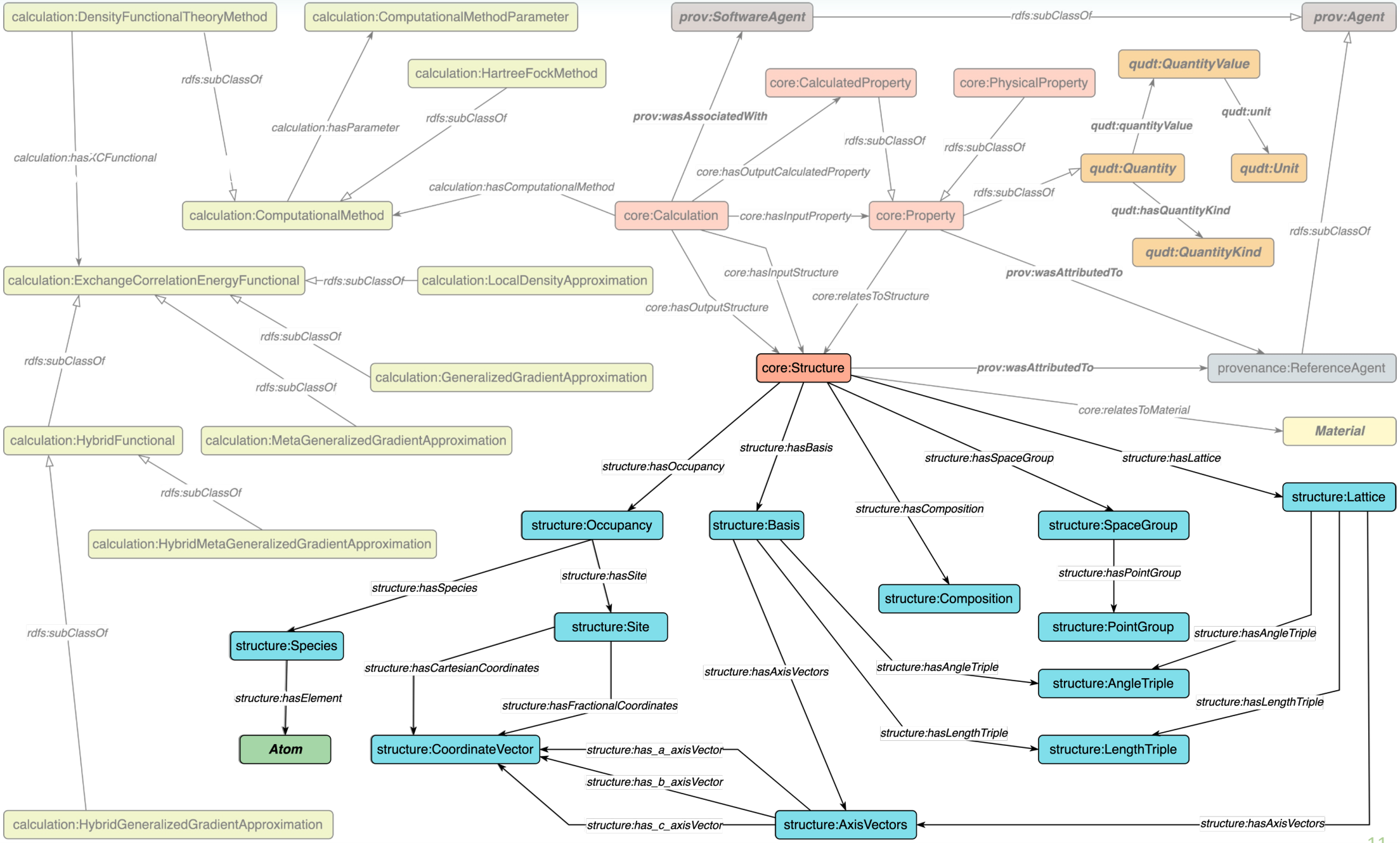
MDO

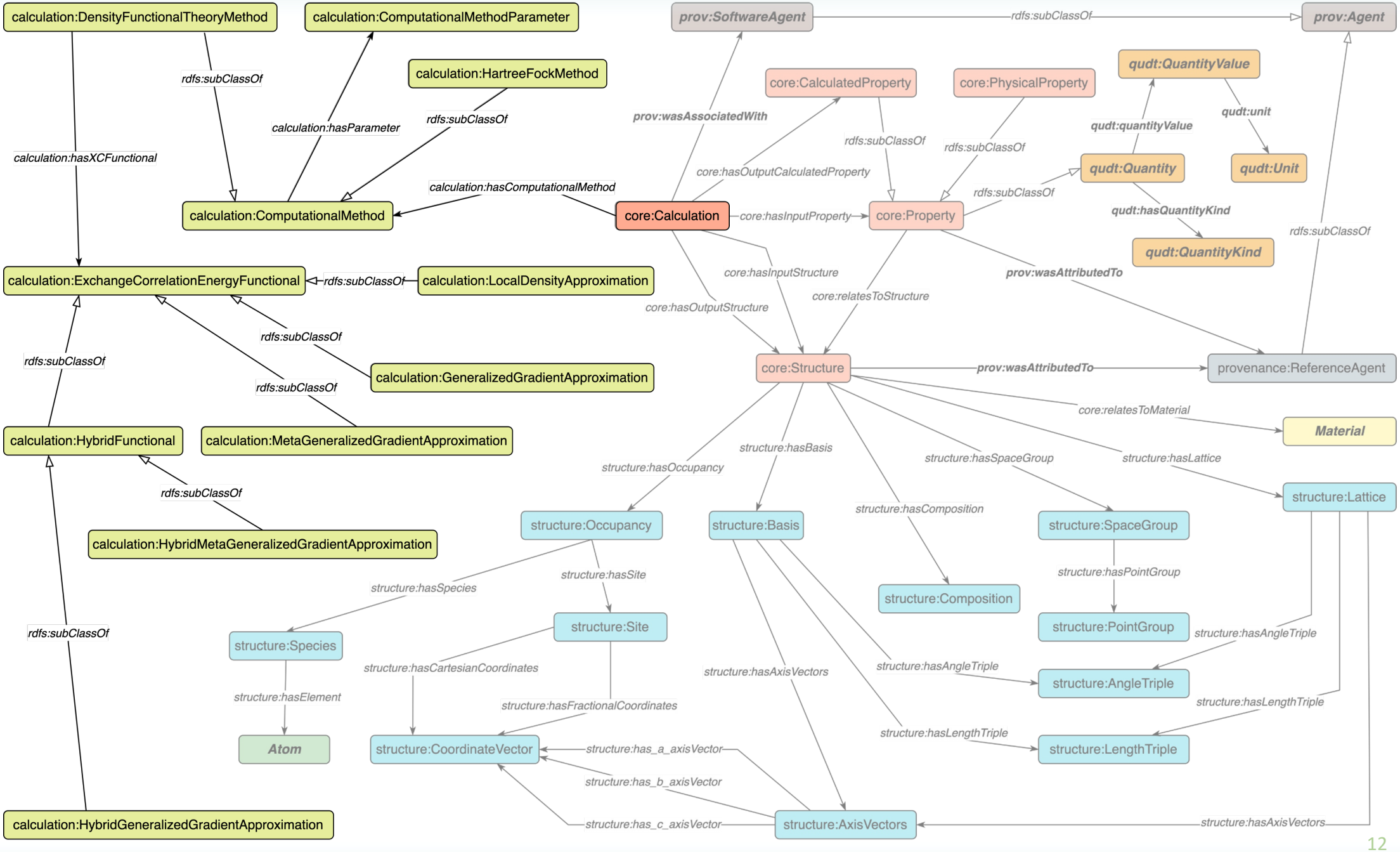
- NeOn methodology in ontology engineering
  - Requirements analysis (Use Cases, Competency Questions, Additional Restrictions)
  - Reusing concepts from existing ontologies
    - PROV-O (PROVenance Ontology), QUDT (Quantities, Units, Dimensions, and Type Ontology)
- Modular Design
  - Core Module, Structure Module, Calculation Module, Provenance Module
- Discussions with a domain expert
- https://github.com/LiUSemWeb/Materials-Design-Ontology

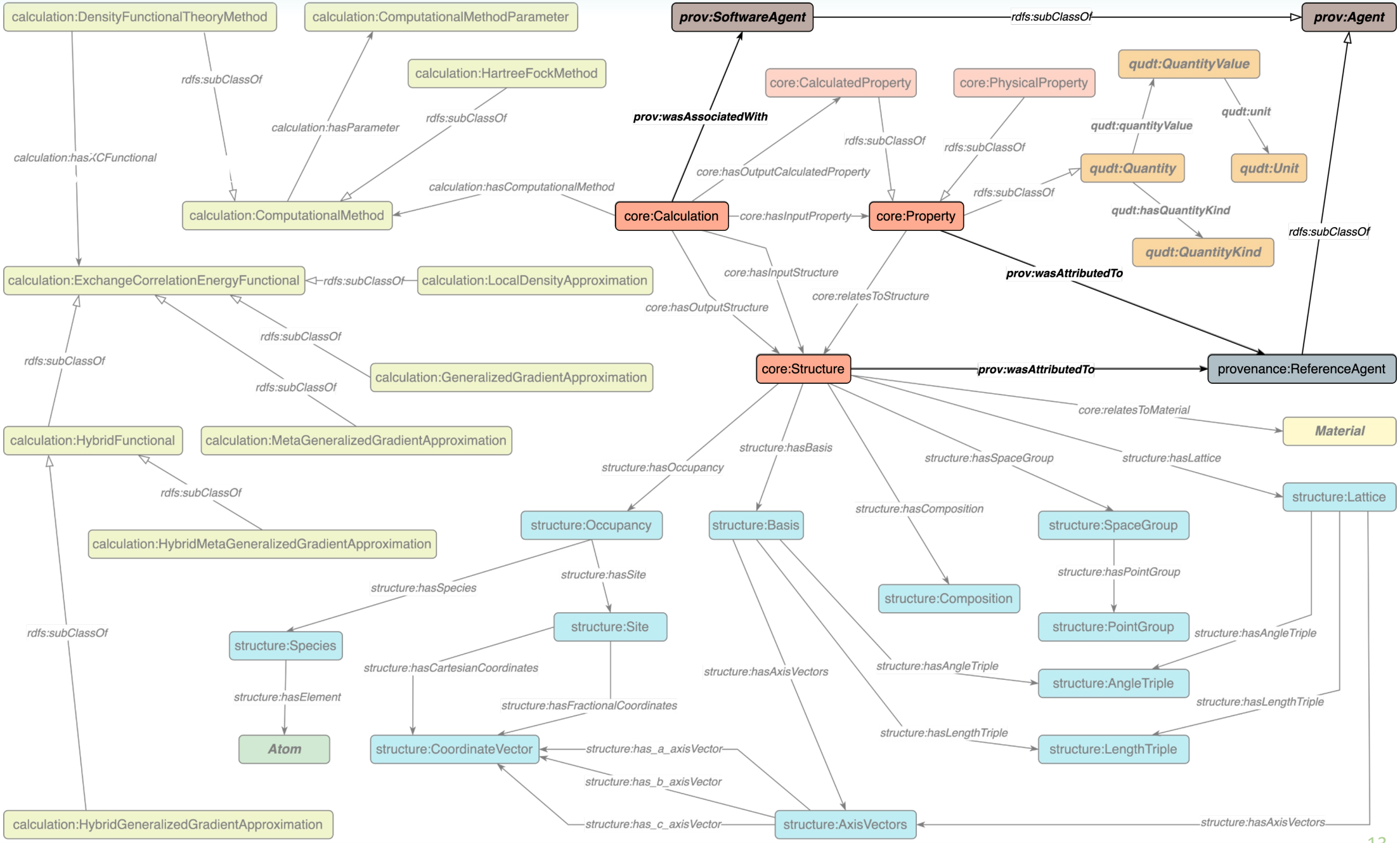
H. Li, R. Armiento and P. Lambrix, An Ontology for the Materials Design Domain, in: *The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, November 2-6, 2020.*

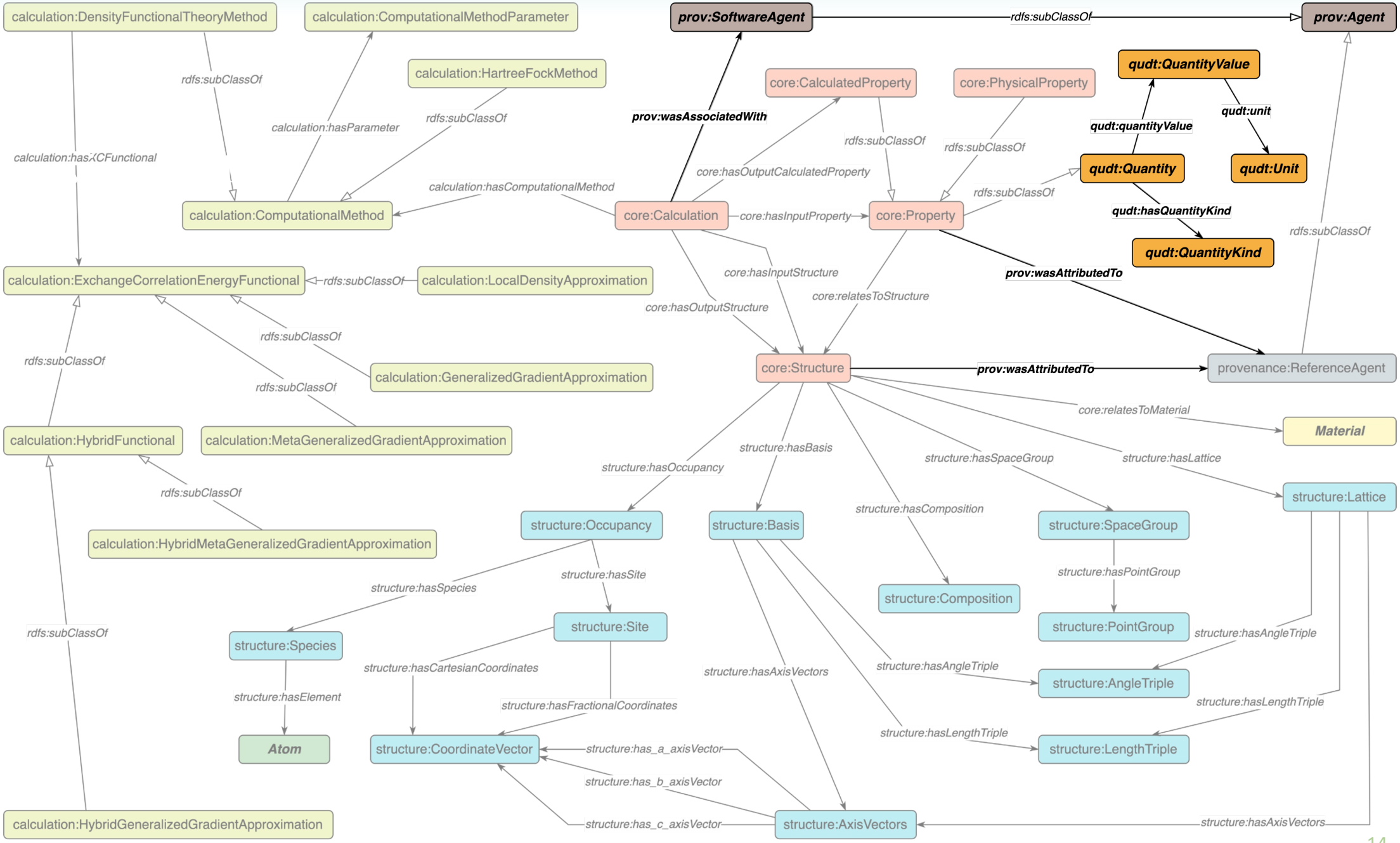

calculation:DensityFunctionalTheoryMethod
calculation:ComputationalMethodParameter
prov:SoftwareAgent
rdfs:subClassOf
prov:Agent
rdfs:subClassOf
calculation:HartreeFockMethod
calculation:hasParameter
rdfs:subClassOf
prov:wasAssociatedWith
core:CalculatedProperty
core:PhysicalProperty
qudt:QuantityValue
qudt:unit
qudt:quantityValue
calculation:hasXCFunctional
rdfs:subClassOf
rdfs:subClassOf
qudt:Quantity
qudt:Unit
core:hasOutputCalculatedProperty
calculation:hasComputationalMethod
calculation:ComputationalMethod
core:Calculation
core:hasInputProperty
core:Property
rdfs:subClassOf
qudt:hasQuantityKind
qudt:QuantityKind
rdfs:subClassOf
calculation:ExchangeCorrelationEnergyFunctional
rdfs:subClassOf
calculation:LocalDensityApproximation
core:hasInputStructure
prov:wasAttributedTo
core:hasOutputStructure
core:relatesToStructure
rdfs:subClassOf
rdfs:subClassOf
calculation:GeneralizedGradientApproximation
core:Structure
prov:wasAttributedTo
provenance:ReferenceAgent
rdfs:subClassOf
core:relatesToMaterial
Material
calculation:HybridFunctional
calculation:MetaGeneralizedGradientApproximation
structure:hasBasis
structure:hasOccupancy
structure:hasSpaceGroup
structure:hasLattice
structure:Lattice
rdfs:subClassOf
structure:hasComposition
calculation:HybridMetaGeneralizedGradientApproximation
structure:Occupancy
structure:Basis
structure:SpaceGroup
structure:hasSite
structure:hasPointGroup
structure:hasSpecies
structure:Composition
structure:Site
structure:PointGroup
structure:hasAngleTriple
rdfs:subClassOf
structure:Species
structure:hasCartesianCoordinates
structure:hasAxisVectors
structure:hasAngleTriple
structure:AngleTriple
structure:hasElement
structure:hasFractionalCoordinates
structure:hasLengthTriple
Atom
structure:CoordinateVector
structure:has_a_axisVector
structure:hasLengthTriple
structure:LengthTriple
structure:has_b_axisVector
structure:has_c_axisVector
structure:AxisVectors
structure:hasAxisVectors
calculation:HybridGeneralizedGradientApproximation


calculation:DensityFunctionalTheoryMethod
calculation:ComputationalMethodParameter
prov:SoftwareAgent
rdfs:subClassOf
prov:Agent
rdfs:subClassOf
calculation:HartreeFockMethod
core:CalculatedProperty
core:PhysicalProperty
qudt:QuantityValue
qudt:unit
calculation:hasParameter
rdfs:subClassOf
prov:wasAssociatedWith
rdfs:subClassOf
rdfs:subClassOf
qudt:quantityValue
calculation:hasXCFunctional
core:hasOutputCalculatedProperty
qudt:Quantity
qudt:Unit
calculation:hasComputationalMethod
calculation:ComputationalMethod
core:Calculation
core:hasInputProperty
core:Property
rdfs:subClassOf
qudt:hasQuantityKind
qudt:QuantityKind
rdfs:subClassOf
calculation:ExchangeCorrelationEnergyFunctional
rdfs:subClassOf
calculation:LocalDensityApproximation
core:hasInputStructure
prov:wasAttributedTo
core:relatesToStructure
core:hasOutputStructure
rdfs:subClassOf
rdfs:subClassOf
core:Structure
prov:wasAttributedTo
provenance:ReferenceAgent
calculation:GeneralizedGradientApproximation
rdfs:subClassOf
core:relatesToMaterial
Material
calculation:HybridFunctional
calculation:MetaGeneralizedGradientApproximation
structure:hasBasis
structure:hasSpaceGroup
structure:hasLattice
structure:hasOccupancy
structure:Lattice
rdfs:subClassOf
structure:hasComposition
calculation:HybridMetaGeneralizedGradientApproximation
structure:Occupancy
structure:Basis
structure:SpaceGroup
structure:hasPointGroup
structure:hasSite
structure:hasSpecies
structure:Composition
structure:Site
structure:PointGroup
structure:hasAngleTriple
rdfs:subClassOf
structure:Species
structure:hasCartesianCoordinates
structure:hasAxisVectors
structure:hasAngleTriple
structure:AngleTriple
structure:hasLengthTriple
structure:hasElement
structure:hasFractionalCoordinates
structure:hasLengthTriple
Atom
structure:CoordinateVector
structure:has_a_axisVector
structure:LengthTriple
structure:has_b_axisVector
structure:has_c_axisVector
structure:AxisVectors
structure:hasAxisVectors
calculation:HybridGeneralizedGradientApproximation


calculation:DensityFunctionalTheoryMethod
calculation:ComputationalMethodParameter
prov:SoftwareAgent
rdfs:subClassOf
prov:Agent
rdfs:subClassOf
calculation:HartreeFockMethod
core:CalculatedProperty
core:PhysicalProperty
qudt:QuantityValue
qudt:unit
calculation:hasParameter
rdfs:subClassOf
prov:wasAssociatedWith
rdfs:subClassOf
rdfs:subClassOf
qudt:quantityValue
calculation:hasXCFunctional
core:hasOutputCalculatedProperty
qudt:Quantity
qudt:Unit
calculation:hasComputationalMethod
rdfs:subClassOf
qudt:hasQuantityKind
calculation:ComputationalMethod
core:Calculation
core:hasInputProperty
core:Property
qudt:QuantityKind
rdfs:subClassOf
calculation:ExchangeCorrelationEnergyFunctional
rdfs:subClassOf
calculation:LocalDensityApproximation
core:hasInputStructure
prov:wasAttributedTo
core:relatesToStructure
core:hasOutputStructure
rdfs:subClassOf
rdfs:subClassOf
core:Structure
prov:wasAttributedTo
provenance:ReferenceAgent
calculation:GeneralizedGradientApproximation
rdfs:subClassOf
core:relatesToMaterial
Material
calculation:HybridFunctional
calculation:MetaGeneralizedGradientApproximation
structure:hasOccupancy
structure:hasBasis
structure:hasSpaceGroup
structure:hasLattice
rdfs:subClassOf
structure:Lattice
structure:hasComposition
calculation:HybridMetaGeneralizedGradientApproximation
structure:Occupancy
structure:Basis
structure:SpaceGroup
structure:hasPointGroup
structure:hasSpecies
structure:hasSite
structure:Composition
rdfs:subClassOf
structure:Species
structure:Site
structure:PointGroup
structure:hasAngleTriple
structure:hasCartesianCoordinates
structure:hasAxisVectors
structure:hasAngleTriple
structure:AngleTriple
structure:hasElement
structure:hasFractionalCoordinates
structure:hasLengthTriple
Atom
structure:CoordinateVector
structure:hasLengthTriple
structure:LengthTriple
structure:has_a_axisVector
structure:has_b_axisVector
structure:has_c_axisVector
structure:AxisVectors
structure:hasAxisVectors
calculation:HybridGeneralizedGradientApproximation


calculation:DensityFunctionalTheoryMethod
calculation:ComputationalMethodParameter
rdfs:subClassOf
calculation:HartreeFockMethod
calculation:hasParameter
rdfs:subClassOf
calculation:hasXCFunctional
calculation:ComputationalMethod
calculation:hasComputationalMethod
core:Calculation
calculation:ExchangeCorrelationEnergyFunctional
rdfs:subClassOf
calculation:LocalDensityApproximation
rdfs:subClassOf
calculation:GeneralizedGradientApproximation
rdfs:subClassOf
rdfs:subClassOf
calculation:MetaGeneralizedGradientApproximation
calculation:HybridFunctional
rdfs:subClassOf
calculation:HybridMetaGeneralizedGradientApproximation
rdfs:subClassOf
calculation:HybridGeneralizedGradientApproximation
prov:SoftwareAgent
rdfs:subClassOf
prov:Agent
prov:wasAssociatedWith
core:CalculatedProperty
core:PhysicalProperty
qudt:QuantityValue
qudt:quantityValue
qudt:unit
rdfs:subClassOf
rdfs:subClassOf
core:hasOutputCalculatedProperty
qudt:Quantity
qudt:Unit
rdfs:subClassOf
core:hasInputProperty
core:Property
qudt:hasQuantityKind
qudt:QuantityKind
rdfs:subClassOf
core:hasInputStructure
prov:wasAttributedTo
core:hasOutputStructure
core:relatesToStructure
core:Structure
prov:wasAttributedTo
provenance:ReferenceAgent
core:relatesToMaterial
Material
structure:hasOccupancy
structure:hasBasis
structure:hasSpaceGroup
structure:hasLattice
structure:Lattice
structure:hasComposition
structure:Occupancy
structure:Basis
structure:SpaceGroup
structure:hasPointGroup
structure:Composition
structure:PointGroup
structure:hasSpecies
structure:hasSite
structure:Species
structure:Site
structure:hasAngleTriple
structure:hasCartesianCoordinates
structure:hasAxisVectors
structure:hasAngleTriple
structure:AngleTriple
structure:hasElement
structure:hasFractionalCoordinates
structure:hasLengthTriple
structure:hasLengthTriple
Atom
structure:CoordinateVector
structure:LengthTriple
structure:has_a_axisVector
structure:has_b_axisVector
structure:has_c_axisVector
structure:AxisVectors
structure:hasAxisVectors


calculation:DensityFunctionalTheoryMethod
calculation:ComputationalMethodParameter
prov:SoftwareAgent
rdfs:subClassOf
prov:Agent
rdfs:subClassOf
calculation:HartreeFockMethod
calculation:hasParameter
rdfs:subClassOf
prov:wasAssociatedWith
core:CalculatedProperty
core:PhysicalProperty
qudt:QuantityValue
qudt:quantityValue
qudt:unit
calculation:hasXCFunctional
rdfs:subClassOf
rdfs:subClassOf
core:hasOutputCalculatedProperty
qudt:Quantity
qudt:Unit
calculation:hasComputationalMethod
rdfs:subClassOf
calculation:ComputationalMethod
core:Calculation
core:hasInputProperty
core:Property
qudt:hasQuantityKind
qudt:QuantityKind
rdfs:subClassOf
prov:wasAttributedTo
calculation:ExchangeCorrelationEnergyFunctional
rdfs:subClassOf
calculation:LocalDensityApproximation
core:hasInputStructure
core:hasOutputStructure
core:relatesToStructure
rdfs:subClassOf
rdfs:subClassOf
calculation:GeneralizedGradientApproximation
core:Structure
prov:wasAttributedTo
provenance:ReferenceAgent
rdfs:subClassOf
core:relatesToMaterial
Material
calculation:HybridFunctional
calculation:MetaGeneralizedGradientApproximation
structure:hasBasis
structure:hasOccupancy
structure:hasSpaceGroup
structure:hasLattice
structure:Lattice
rdfs:subClassOf
structure:hasComposition
calculation:HybridMetaGeneralizedGradientApproximation
structure:Occupancy
structure:Basis
structure:SpaceGroup
structure:hasSite
structure:hasPointGroup
structure:hasSpecies
structure:Composition
structure:Site
structure:PointGroup
structure:hasAngleTriple
rdfs:subClassOf
structure:Species
structure:hasCartesianCoordinates
structure:hasAxisVectors
structure:hasAngleTriple
structure:AngleTriple
structure:hasElement
structure:hasFractionalCoordinates
structure:hasLengthTriple
structure:hasLengthTriple
Atom
structure:CoordinateVector
structure:has_a_axisVector
structure:LengthTriple
structure:has_b_axisVector
structure:has_c_axisVector
structure:AxisVectors
structure:hasAxisVectors
calculation:HybridGeneralizedGradientApproximation


calculation:DensityFunctionalTheoryMethod
calculation:ComputationalMethodParameter
prov:SoftwareAgent
rdfs:subClassOf
prov:Agent
rdfs:subClassOf
calculation:HartreeFockMethod
qudt:QuantityValue
prov:wasAssociatedWith
core:CalculatedProperty
core:PhysicalProperty
qudt:unit
calculation:hasParameter
rdfs:subClassOf
qudt:quantityValue
calculation:hasXCFunctional
rdfs:subClassOf
rdfs:subClassOf
qudt:Quantity
qudt:Unit
core:hasOutputCalculatedProperty
rdfs:subClassOf
calculation:hasComputationalMethod
calculation:ComputationalMethod
core:Calculation
core:hasInputProperty
core:Property
rdfs:subClassOf
qudt:hasQuantityKind
qudt:QuantityKind
rdfs:subClassOf
prov:wasAttributedTo
calculation:ExchangeCorrelationEnergyFunctional
rdfs:subClassOf
calculation:LocalDensityApproximation
core:hasInputStructure
core:relatesToStructure
core:hasOutputStructure
rdfs:subClassOf
rdfs:subClassOf
calculation:GeneralizedGradientApproximation
core:Structure
prov:wasAttributedTo
provenance:ReferenceAgent
rdfs:subClassOf
core:relatesToMaterial
Material
calculation:HybridFunctional
calculation:MetaGeneralizedGradientApproximation
structure:hasBasis
structure:hasOccupancy
structure:hasSpaceGroup
structure:hasLattice
structure:Lattice
rdfs:subClassOf
structure:hasComposition
calculation:HybridMetaGeneralizedGradientApproximation
structure:Occupancy
structure:Basis
structure:SpaceGroup
structure:hasSite
structure:hasPointGroup
structure:hasSpecies
structure:Composition
structure:Site
structure:PointGroup
structure:hasAngleTriple
rdfs:subClassOf
structure:Species
structure:hasCartesianCoordinates
structure:hasAxisVectors
structure:hasAngleTriple
structure:AngleTriple
structure:hasElement
structure:hasFractionalCoordinates
structure:hasLengthTriple
structure:hasLengthTriple
Atom
structure:CoordinateVector
structure:has_a_axisVector
structure:LengthTriple
structure:has_b_axisVector
structure:has_c_axisVector
structure:AxisVectors
structure:hasAxisVectors
calculation:HybridGeneralizedGradientApproximation

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

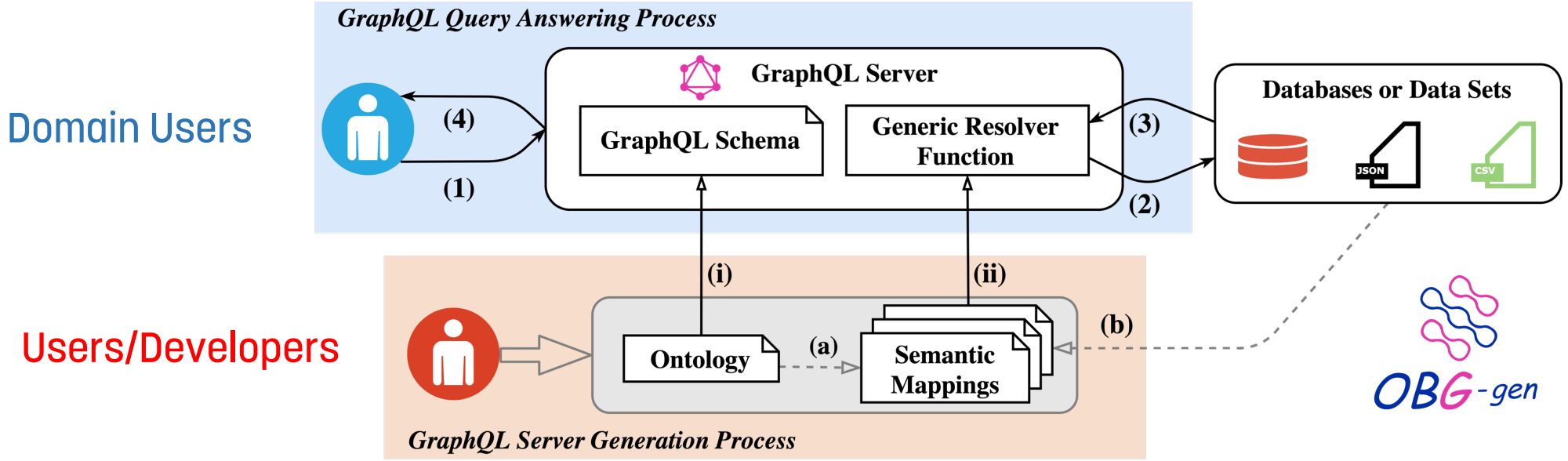
Ontology-driven Data Access and Data Integration

LINKÖPING UNIVERSITY

H. Li, Ontology-Driven Data Access and Data Integration with an Application in the Materials Design Domain, PhD thesis, 2022, Linköping University, Sweden.

GraphQL-based framework for data access and integration

- GraphQL Server Generation Process, arrows (i) and (ii)
  - Ontology-based GraphQL server generation (OBG-gen)
- GraphQL Query Answering Process, arrows (1)–(4)
- https://github.com/LiUSemWeb/OBG-gen

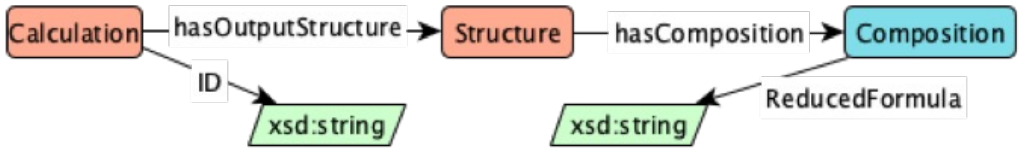


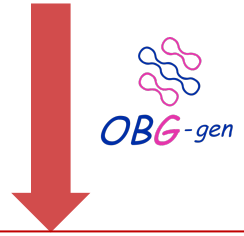
H. Li, Ontology-Driven Data Access and Data Integration with an Application in the Materials Design Domain, PhD thesis, 2022, Linköping University, Sweden.

GraphQL

- What is GraphQL?
  - GraphQL is a conceptual framework for building Web APIs
  - GraphQL can work with existing APIs of a system
  - A GraphQL server contains GraphQL schema and GraphQL resolver
  - Clients use the GraphQL query language to make requests to a GraphQL server

GraphQL






```
type Query {
    CalculationList(filter: CalculationFilter): [Calculation]
}
type Calculation {
    hasOutputStructure: [Structure]
    hasOutputCalculatedProperty: [CalculatedProperty]
    ID: String
}
type Structure {
    StructureID: String
    hasComposition: Composition
}
type Composition {
    ReducedFormula: String
}
```

- Get all the calculations where ID in a given list, and reduced chemical formula contains the chlorine (Cl) or oxygen (O) elements

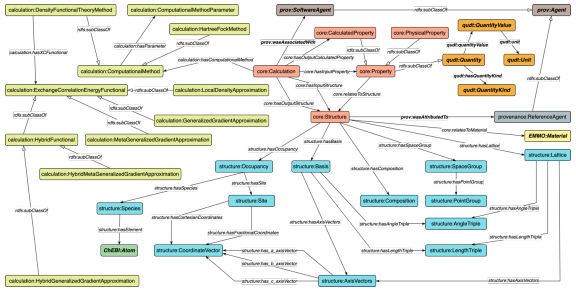
```
query My_First_Query{
  CalculationList(
    filter: {
      _and: [
        { ID: { _in: ["6332", "8088", "21331", "mp-561628", "mp-614919"] } },
        { _or: [
           {hasOutputStructure:{hasComposition:{ReducedFormula:{_like: "%Cl%"}}}},
           {hasOutputStructure:{hasComposition:{ReducedFormula:{_like: "%O%"}}}}]
        }]
    }
  )
  {
    ID
    hasOutputStructure{
      hasComposition{
        ReducedFormula
      }
    }
  }
}
```

Summary and Ongoing work

Summary and Ongoing Work

✓ Materials Design Ontology

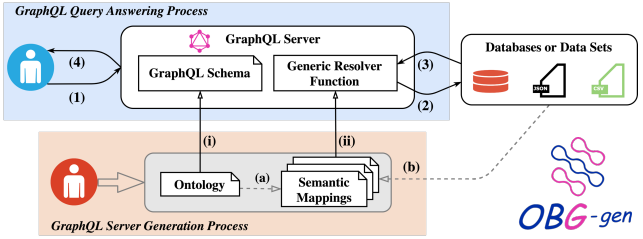
- MDO is capable to represent basic domain knowledge
- MDO can be used for mapping different materials databases' schemas
- MDO can be used for semantically enabling materials database search



MDO

✓ A GraphQL-based framework for data access and integration

- with an application in the materials design domain



➢ Investigating the compatibility of MDO and top-level ontologies (e.g., EMMO)

- aligning MDO-core with EMMO

➢ Extending MDO with new concepts and relationships

- A new MDO-property module