Introduction to Ontologies

Maja-Olivia Lenz-Himmer

Coffee Talk

November 11, 2019

What is an ontology?

There is no common definition of an ontology in literature.

- "explicit specification of a conceptualization" (1993, Gruber) $\rightarrow$ Huh???
- "An ontology encompasses a representation, formal naming and definition of the categories, properties and relations between the concepts, data and entities that substantiate one, many or all domains of discourse." (Wikipedia)
- "A set of concepts and categories in a subject area or domain that shows their properties and the relations between them." (Oxford dictionary)

What is an ontology?

More helpful: An ontology is a formal description of knowledge as a set of concepts within a domain and the relationships that hold between them.

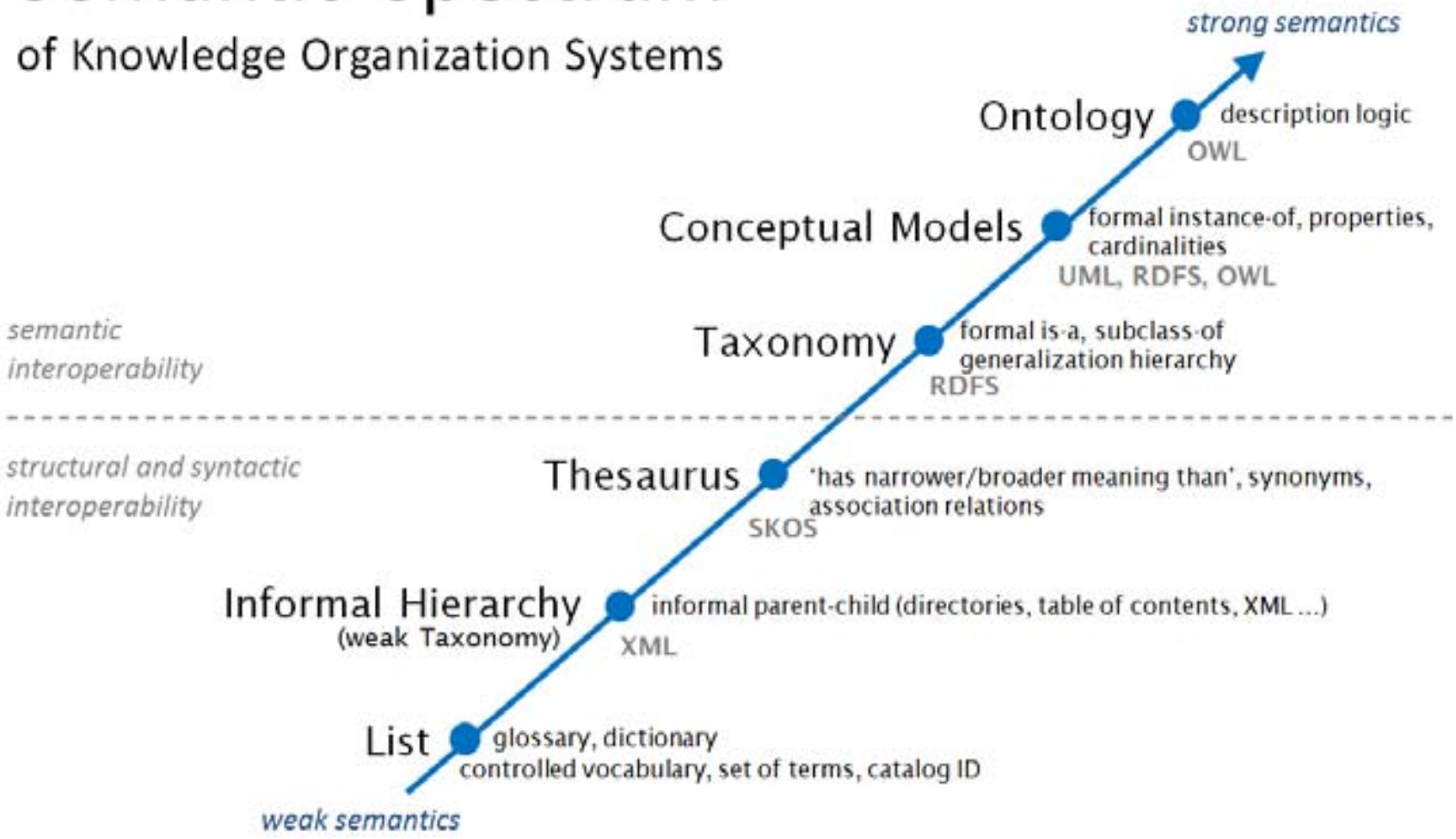
To enable such a description, we need to formally specify components such as

- individuals (instances of objects)
- classes
- attributes
- relations
- restrictions
- rules
- axioms

As a result, ontologies do not only introduce a sharable and reusable knowledge representation but can also add new knowledge about the domain.
(ontotext.com)

Semantic Spectrum

of Knowledge Organization Systems



Why develop an ontology?

- Share common understanding of the structure of information among people or software agents
  → Example: Different Websites containing medical information or providing medical services
  - Automated extraction and aggregation of information from these different sites
  - Use to answer user queries or as input to other applications

- Enable reuse of domain knowledge
  → Example: Representation of time in different domains
  - Time intervals, points in time, relative measures of time
  - Usage of existing detailed "time ontology" instead of re-inventing the wheel

Why develop an ontology?

- Make domain assumptions explicit
  - Change assumptions easily if knowledge changes
  - Useful for new users who want to learn about terms in a domain
- Separate domain knowledge from the operational knowledge
  - Specification of building a product from components independent of the product and its components themselves
  - Imagine: Someone builds tools to calculate materials properties without knowing anything about these materials and properties
- Analyze domain knowledge
  - Useful for reusing existing ontologies and extending them

How to?

OWL – Web ontology language

- Based on RDF/XML
- Everything is a Resource and has a unique identifier (URI)
- Consists of Triples connecting things (subject predicate object)
- Different versions of OWL with different orders of logic (description logic, first-order logic, ...)

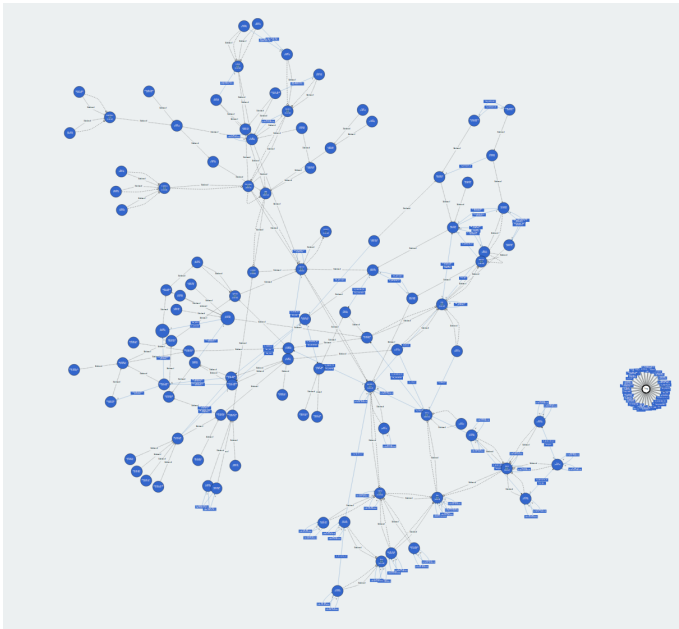
Reasoning

- Software tool that infers logical consequences from given facts/axioms in ontology

$\rightarrow$ New (transitive) relationships and facts enter the ontology automatically

EMMO – European Materials and Modeling Ontology

Remember Jesper Friis? Remember the talk on EMMO?



- Very general ontology about the universe starting from the concept of spacetime
- Allows to define in principle *everything*

Main relations

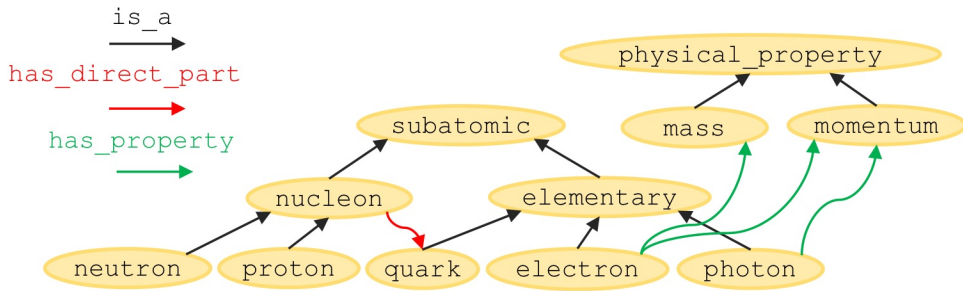
Only 4 primitive families

TAXONOMY
Classification
Ex: `is_a`

MEREOTOPOLOGY
Parthood and *Slicing*
Ex: `has_part`

SEMIOTIC
Representation
Ex: `has_property`

SET THEORY
Membership
Ex: `has_member`



Items that unfolds in space and time
Granularity (multi-scale modelling)

Signs that stands for something else
Represents real-world objects

Abstract collections of items

STREAM

Semantic representation, linking and curating of quality-ensured material data

The main goals of STREAM are

- Developing a community-driven common vocabulary
- Developing a materials ontology
- Developing a quality measure for material data
- Mapping from existing data structures onto vocabulary
- Inclusion of / Distribution in the communities

FRITZ-HABER-INSTITUT
MAX-PLANCK-GESELLSCHAFT

TIB LEIBNIZ-INFORMATIONSZENTRUM TECHNIK UND NATURWISSENSCHAFTEN UNIVERSITÄTSBIBLIOTHEK

Fraunhofer IWM

InfAI Institut für Angewandte Informatik

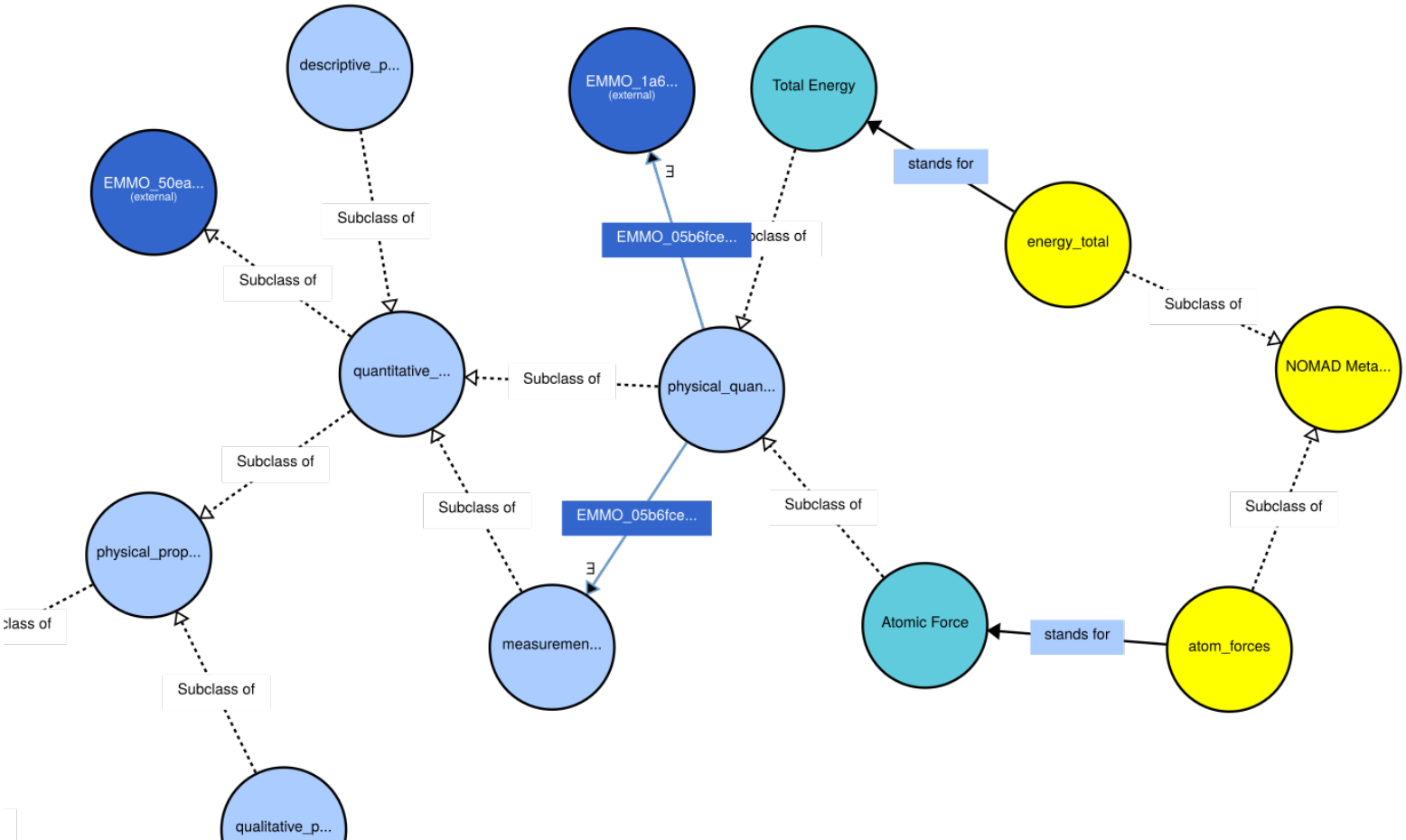
KIT Karlsruhe Institute of Technology

NOMAD Meta Info

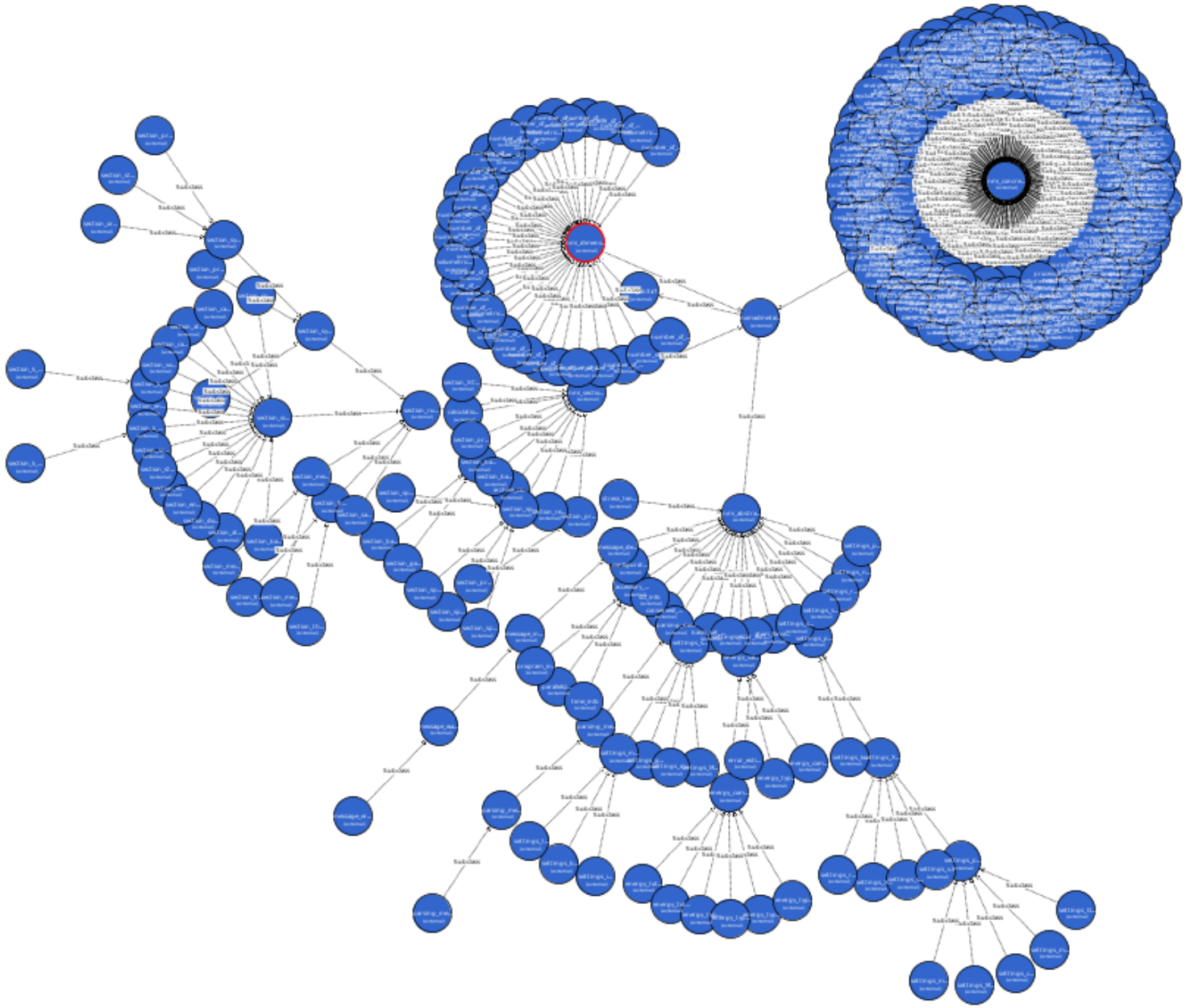
"Brand name" for Metadata contained in the NOMAD Archive

Descriptive and structured, code-independent information about materials-science data contained in the NOMAD Archive

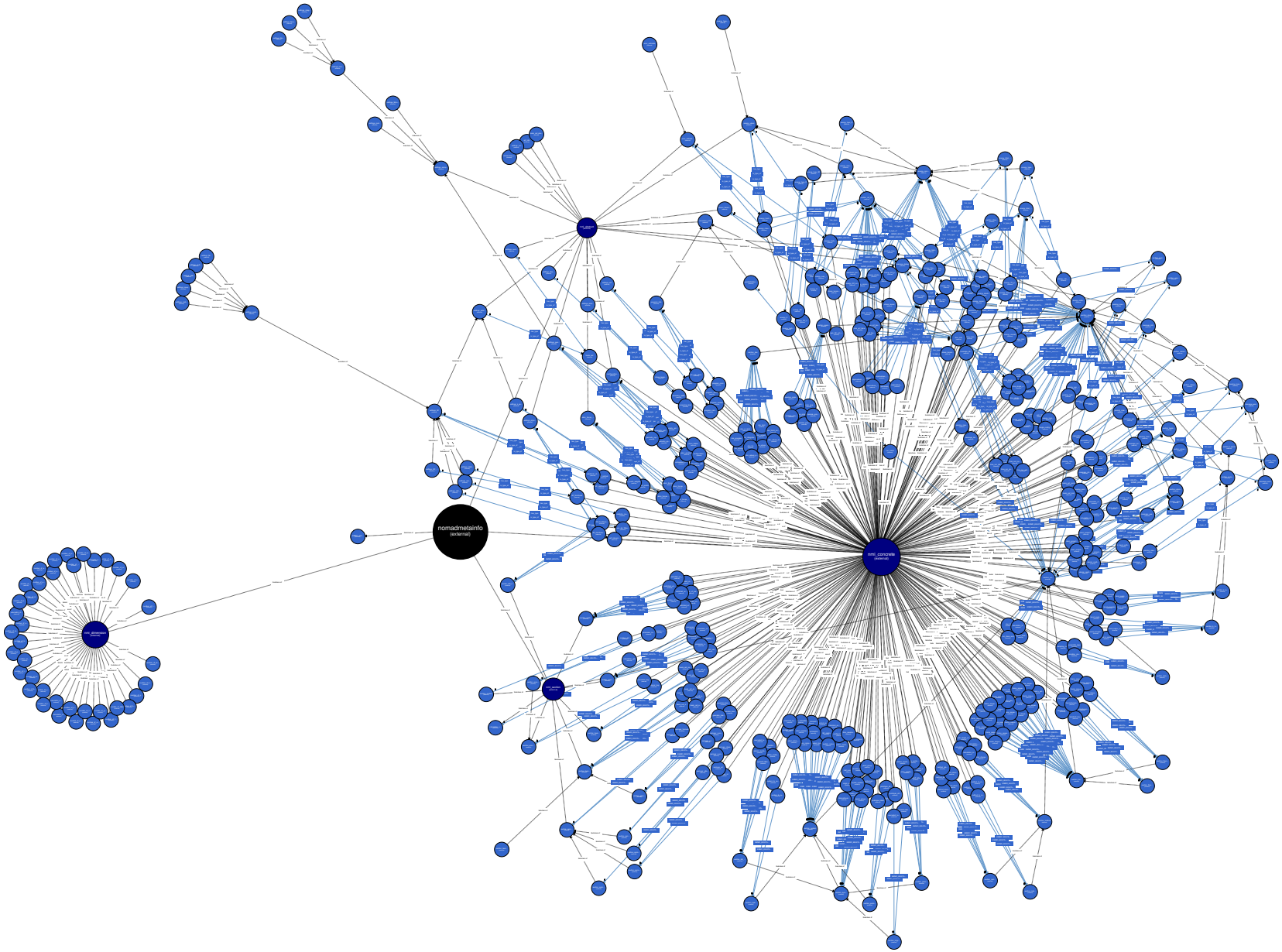
→ Include Meta Info in "Semiotic" branch of EMMO as "signs" for common terms



NOMAD Meta Info – only hierarchy



NOMAD Meta Info – including relations



Hands-on: Challenges and Examples

Crystal unit cell

- *Supercell* and *Primitive cell*: Subclass or part of *crystal unit cell*?
- Different representations of the same thing?
  (But different properties (e.g. energy)...
- Direct and proper parthood?

Workflow

- Most properties not direct output of single program call
- Simple example: Heat capacity (output of Phonopy)
  Input geometry $\rightarrow$ FHI-aims $\rightarrow$ relaxed geometry $\rightarrow$ Phonopy $\rightarrow$ displaced supercells $\rightarrow$ multiple FHI-aims calls $\rightarrow$ forces $\rightarrow$ Phonopy $\rightarrow$ Heat capacity