National Physical Laboratory

What are semantic representations and where to find them?

Jean-Laurent Hippolyte

Michael Chrubasik

Keith Lines

Marina Romanchikova

Image source: https://www.bbc.co.uk/programmes/articles/pzx5tJ7XHjZDLxH4SCllvp/which-fantastic-beast-are-you

Agenda

- What are semantic representations? What are ontologies?
- How are ontologies useful to research data management?
- What methodology did we use to find semantic representations for metrology?
- Highlights and lowlights of the selected representations
- General data management considerations
- Appendices

Semantic Representation

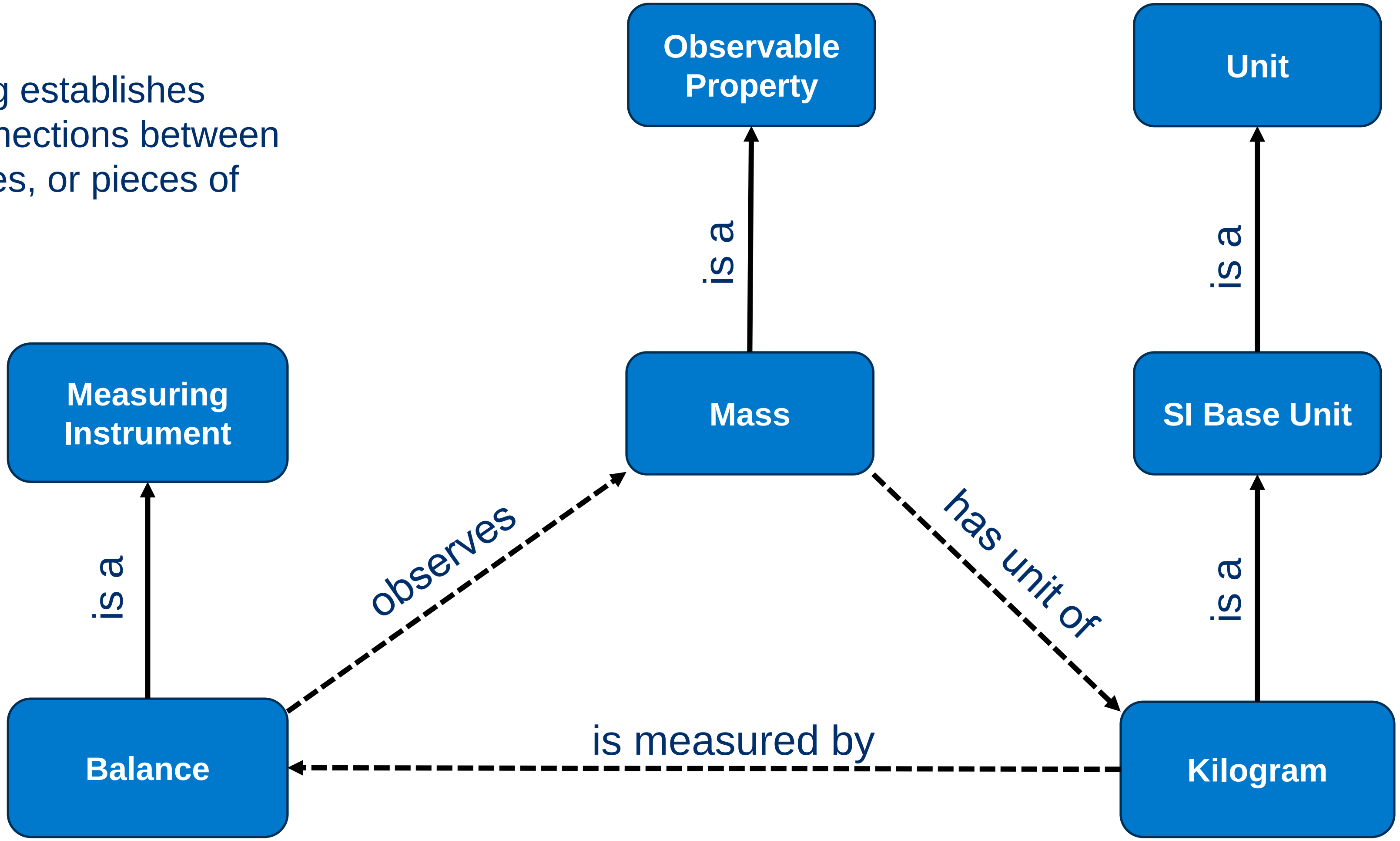
Semantic representation refers to the way meanings are encoded or represented in a language. It facilitates the understanding and communication of meanings of words, phrases, and sentences.

What would a semantic representation of the term "Kilogram" look like? Need to answer a couple of questions like:

- What is a kilogram?
  - A unit of measurement, abstract concept
  - Linked to observable property of mass
- Where does its meaning come from?
  - Originally represented through an artefact and compared against
  - It is defined by relation to system of measurement and dependent on international standards
  - Through the SI has a precise definition: Planck constant h, at 6.62607015×10−34 when expressed in the unit J·s, which is equal to kg·m2·s−1, where the metre and the second are defined in terms of c and ΔvCs
- What is its role? How do you measure it? Relations? …

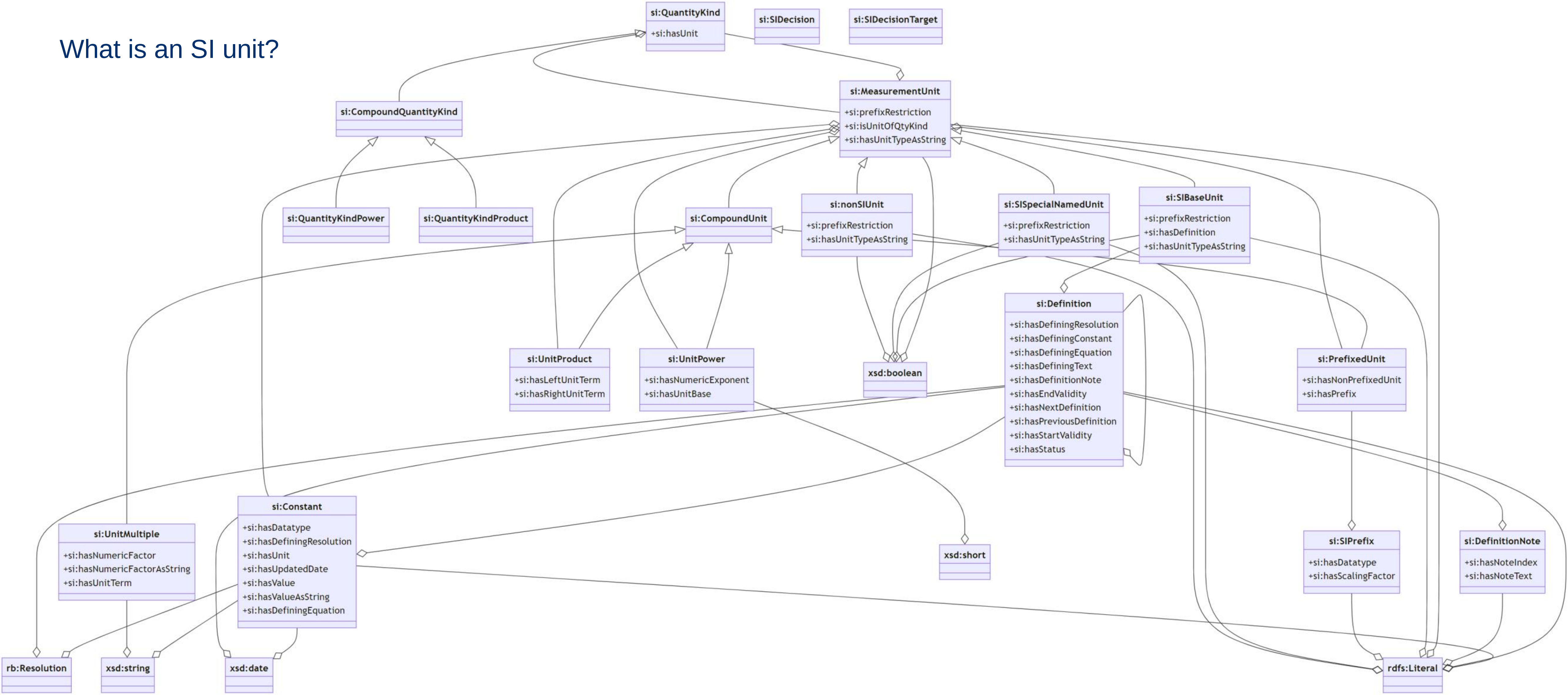
Semantic Representation

Semantic linking establishes meaningful connections between concepts, entities, or pieces of information.



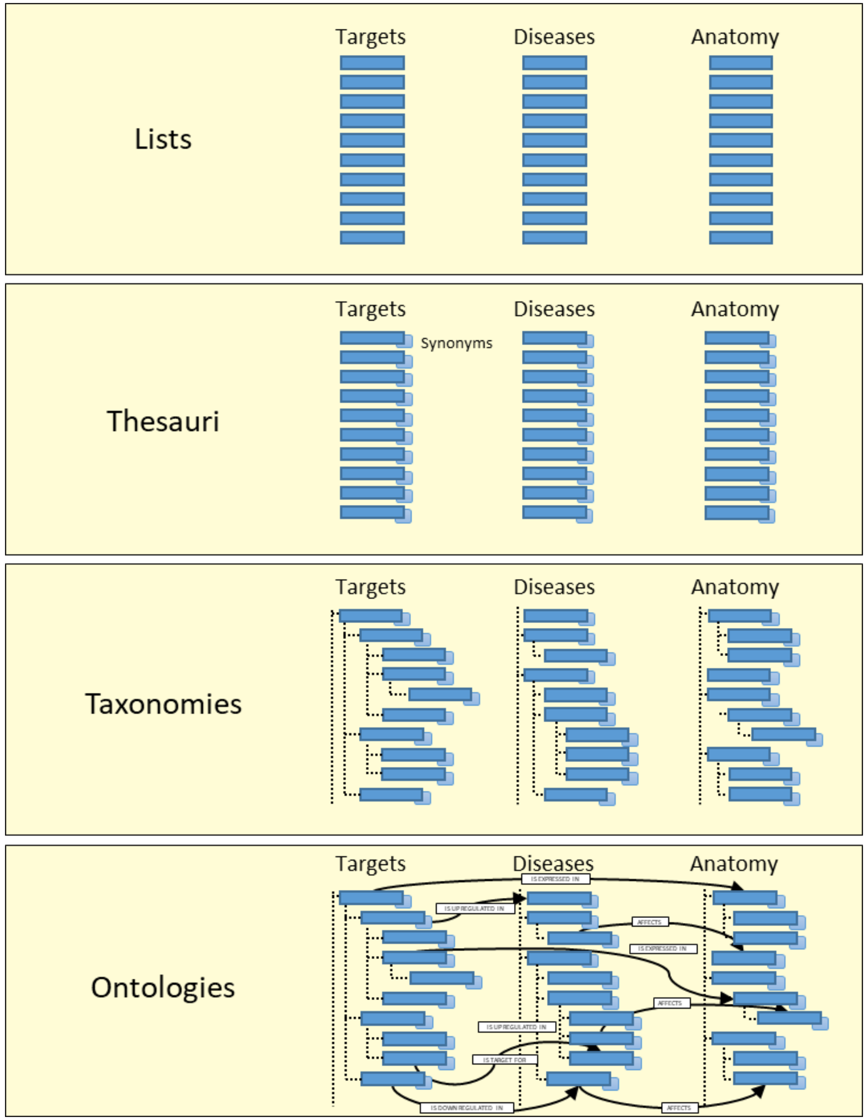
Semantic Representation

What is an SI unit?



Ontologies

- Ontologies* are shareable, reusable and computable representations of connected pieces of knowledge



- OWL semantic web language is a W3C standard for sharing ontologies on the web

Image: Swain, M. (2013). Knowledge Representation. In: Dubitzky, W., Wolkenhauer, O., Cho, KH., Yokota, H. (eds) Encyclopedia of Systems Biology. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-9863-7_595

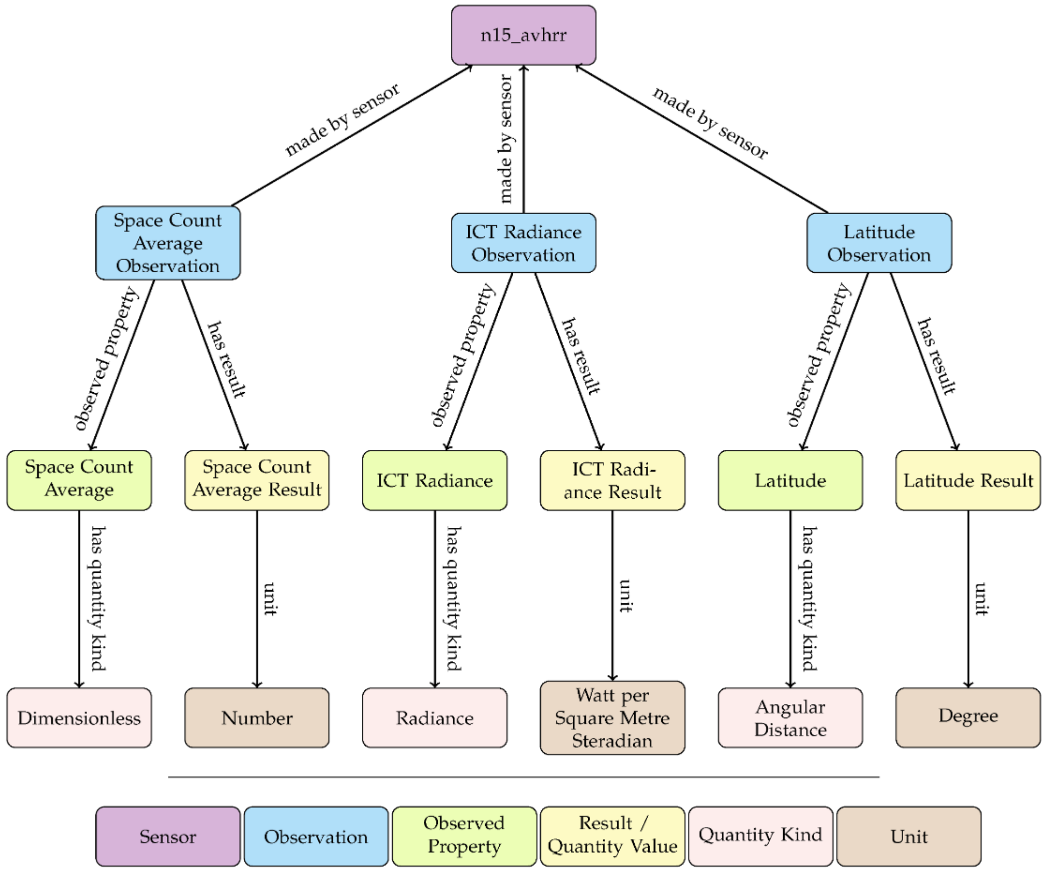
*computational ontologies

https://www.w3.org/OWL/

Semantic Representation through Ontologies

Challenges

- **Domain Complexity**: Capturing the full complexity and nuances of a domain in a formal representation can be difficult and time-consuming.
- **Knowledge Acquisition**: Acquiring and formalising domain knowledge from experts and existing sources can be a substantial effort.
- **Consensus Building**: Achieving consensus among stakeholders on the concepts, relationships, and definitions within a taxonomy/ontology can be challenging.
- **Maintenance and Evolution**: Keeping models up-to-date and aligned with evolving domain knowledge requires ongoing effort and collaboration.



Semantic Representation through Ontologies

Common Pitfalls

- **Overcomplicating the Ontology**: Striking the right balance between expressiveness and simplicity is crucial. An overly complex ontology can be difficult to understand, maintain, and use effectively.
- **Neglecting User Requirements**: Failing to align the model with the needs and requirements of its intended users can limit its adoption and usefulness.
- **Insufficient Documentation**: Providing clear and comprehensive documentation is essential for understanding and using models effectively.
- **Ignoring Best Practices and Standards**: Following established best practices and standards for model development can help ensure interoperability, reusability, and maintainability.

Foundations of the semantic web ontologies

NPL

Semantic web (OWL) ontologies bring formal aspects of:

- Logic
- Set theory

To meta(data) modelling

Foundations of semantic web ontologies : first-order logic

- First-order logic (a.k.a. first-order predicate calculus)
  - To express and resolve the truth value of complex statements that depends on multiple variables
  - *Example: If the observer is facing a screen of rain droplets and the sun is behind the observer at a sufficiently low angle, then the observer can see a rainbow.*

$$IsFacing(observer, droplets) \wedge IsBehind(sun, observer) \wedge IsAngleLowEnough(observer, sun) \Longrightarrow CanSee(observer, rainbow)$$

- ❑ *IsFacing*, *IsBehind*, *IsAngleLowEnough* and *CanSee* are predicates;
- ❑ *me*, *droplets*, *sun*, *rainbow* are variables.

Foundations of semantic web ontologies : first order logic

$A \implies B \equiv \neg A \vee B$

- Very expressive and allows manipulation of concepts/objects and relationships between concepts/objects
- Axioms, simplification rules and inference rules **do not depend on the nature of the concepts/objects that are manipulated**.

A = Today is Easter
B = Tomorrow is a Monday

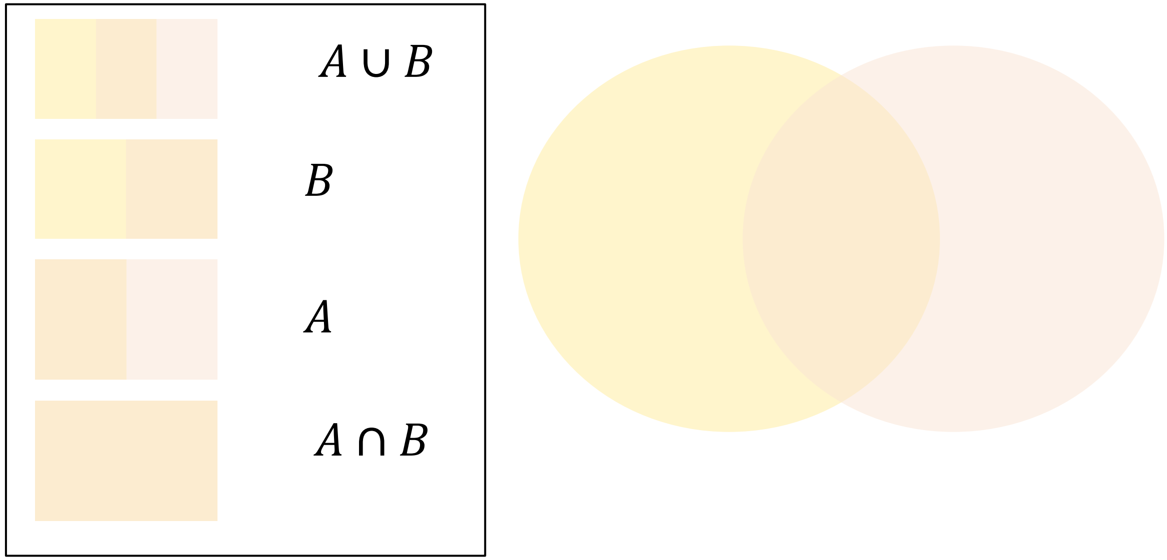
A = Paul is a human
B = Paul has a brain

A= $IsFacing(observer, droplets) \wedge \mathrm{IsBehind}(\mathrm{sun}, \mathrm{observer}) \wedge IsAngleLowEnough$

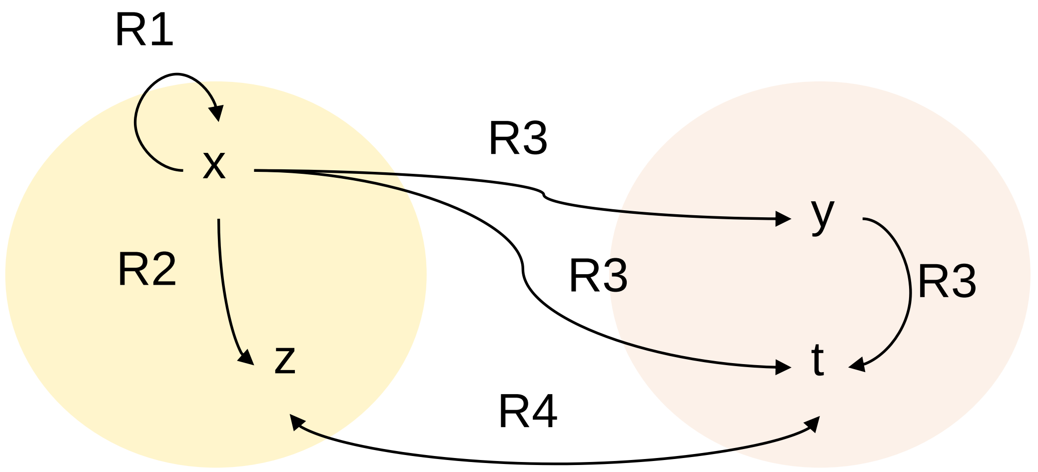
B= $CanSee(\mathrm{observer}, rainbow)$

Foundations of semantic web ontologies : set theory

- Set = a sub set of a predetermined collection of objects
  - Sets regarded as entities, set theory looks at the behaviour of these entities, **regardless of what their elements are**.
- Operations can be applied to sets



Foundations of semantic web ontologies : set theory



- Relations can be defined between elements of:
  - the same set or,
  - different sets.
- Unary and binary relations important
- Binary relations can have interesting properties for knowledge representation and discovery: reflexivity (e.g. *knows, R1*), symmetry (e.g. *is friends with, R4*), transitivity (e.g. *happens before, R3*) etc

Foundations of semantic web ontologies : links with OWL

- OWL based on Description Logics (DL) family of formal languages
- DL subset of first order logic with set theoretical aspects
- DL $\neq$ first order-logic because it is decidable (algorithms exist to compute truth value of DL expressions)

- OWL adds to DL
  - Entities in OWL identified by URIs (or IRIs)
  - OWL has datatypes
  - Integrated with other semantic web languages RDF and RDFS

FAIR principles

The FAIR (Findable, Accessible, Interoperable, Reusable) guiding principles established in the seminal paper

- To ensure credibility of research data
- To improve research data curation
- To achieve wider dissemination of scientific results
- To promote data-driven innovation

Using ontologies can improve compliance with the FAIR principles.

Wilkinson, M. D. et al. (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', Scientific Data, 3(1), p. 160018. doi: 10.1038/sdata.2016.18

FAIR principles

Findability, Accessibility

- Ontologies* supports **Findability** through:
  - URIs as identifiers for concepts and data
  - Resolvable URI-based ontology namespaces

Ontologies* support **Accessibility** through:

- Standard web languages and protocols
- OWL for expression
- SPARQL for querying
- HTTP protocol

<table>
<tr><th colspan="4">Findable</th></tr>
<tr><td>F1. (Meta)data are assigned a globally unique and persistent identifier</td><td>F2. Data are described with rich metadata (defined by R1)</td><td>F3. Metadata clearly and explicitly include the identifier of the data they describe</td><td>F4. (Meta)data are registered or indexed in a searchable resource</td></tr>
<tr><th colspan="4">Accessible</th></tr>
<tr><td colspan="2">A1. (Meta)data are retrievable by their identifier using a standardised communications protocol</td><td colspan="2" rowspan="2">A2. Metadata are accessible, even when the data are no longer available</td></tr>
<tr><td>A1.1 The protocol is open, free, and universally implementable</td><td>A1.2 The protocol allows for an authentication and authorisation procedure, where necessary</td></tr>
</table>

https://www.go-fair.org/fair-principles/

*semantic web ontologies

Drivers: FAIR principles

Interoperability, Reusability

Ontologies* support **Interoperability** through:

- Standardised exchange formats
- Modularity/Linked data/Networked ontologies
- Existing FAIR ontologies
- Variety of existing tools to create and process ontologies

Ontologies* support **Reusability** through:

- Reasoning: Allow for inference and based on the defined relationships and rules
- Good practices in semantic web

Interoperable		
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	I2. (Meta)data use vocabularies that follow FAIR principles	I3. (Meta)data include qualified references to other (meta)data
Reusable		
R1. (Meta)data are richly described with a plurality of accurate and relevant attributes		
R1.1. (Meta)data are released with a clear and accessible data usage license	R1.2. (Meta)data are associated with detailed provenance	R1.3. (Meta)data meet domain-relevant community standards

https://www.go-fair.org/fair-principles/

*semantic web ontologies

Resource selection and evaluation

NPL

Material collection

- Used expert knowledge and two reviews [1, 2] to identify resource candidates
- 119 resources suggested

Criteria definition

- 6 criteria categories: quantities & units; measurement process & equipment; uncertainty & traceability; quality infrastructure; data management.

Classification & selection

- Exclude materials that did not meet criteria
- Categorise the remainder (N=97)

1. Use Cases and Suitability Metrics for Unit Ontologies, https://doi.org/10.1007/978-3-319-54627-8_4
2. Comparison and evaluation of ontologies for units of measurement, https://doi.org/10.3233/SW-180310

List of criteria

Quantities and Units
Measurement Process
Measuring Equipment
Uncertainty and Traceability
Quality Infrastructure
Data Management

Disclaimer

- Concepts are not exclusive to one model category
- Some models overlap multiple criteria

1 of 5: Quantities and Units

Content and Explanation

- Units of measurement and scales, as well as their definitions, interrelationships, conversions, and standards;
- Modelling of how quantities are measured, including the relationships between the measurand, the measurement method/instrument, the measurement unit/scale, and the measurement result and values

1 of 5: Quantities and Units

31 out of 58 papers in literature review included this theme.

Highlight

- QUDT Quantities, Units, Dimensions and Data Types Ontologies
- **Papers:**
  - https://doi.org/10.1109/GIOTS.2018.8534561

- **Resource:** https://www.qudt.org/2.1/catalog/qudt-catalog.html
- **Description:** large set of ontologies based on dimensional analysis
- **Reasons:** Thorough web documentation, sensible ontology network structure, comprehensiveness
- **Case study:** https://doi.org/10.1109/IGARSS47720.2021.9553547

2 of 5: Measurement Process

NPL

23 out of 54 papers in literature review included this theme.

Content and Explanation

- Types of measurement processes, procedures, methods, instruments, tools, and calibration approaches;
- Metadata about measurements and calibration data, including software, instruments and anyone involved in the measurement.

2 of 5: Measurement Process

25 out of 58 papers in literature review included this theme.

Highlight

- An Ontology for the Materials Design Domain,

  Huanyu Li, Rickard Armiento, Patrick Lambrix, Linköping University, Sweden

- **Paper:** https://doi.org/10.48550/arXiv.2006.07712
- **Resource:** https://github.com/LiUSemWeb/Materials-Design-Ontology
- **Description:** Materials Design Ontology (MDO), concepts and relations to cover knowledge in materials design. Designed using domain knowledge in materials science (especially in solid-state physics)
- **Reasons:** Thoroughly documented from introductory information to underlying axioms. Easily accessible.
- **Case study:** https://huanyu-li.github.io/posters/mdo-poster-paper276.pdf

2 of 5: Measurement Process

23 out of 58 papers in literature review included this theme.

Highlight

- SmartCOM DCC XSD schema, PTB
- **Paper:** https://zenodo.org/doi/10.5281/zenodo.3696566
- **Resource:** https://dccwiki.ptb.de/en/home
- **Description:** an XML schema for the exchange of calibration certificate
- **Reasons:** online documentation, easy access, growing support from metrology community, demonstrators and good practice guides available
- **Case study:**
  - https://doi.org/10.3390/metrology2010003

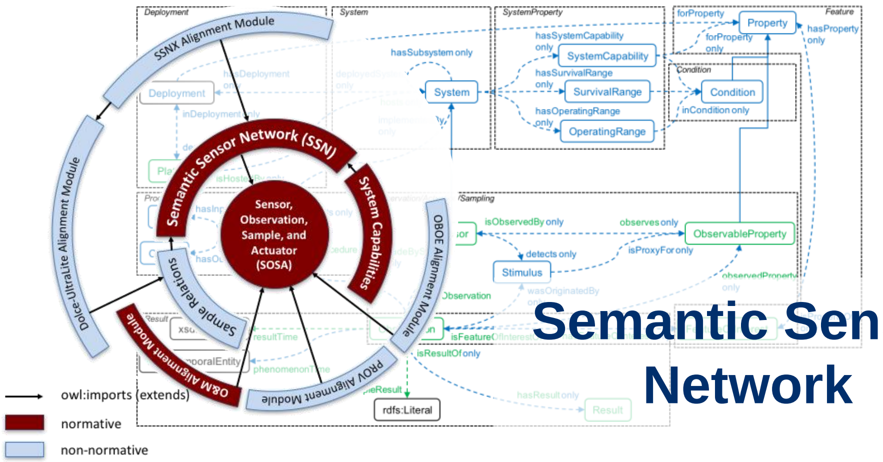
3 of 5: Devices for measurement

NPL

Content and Explanation

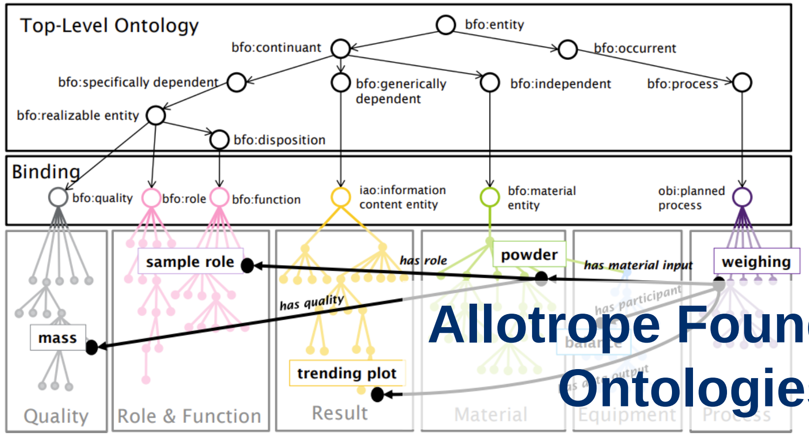
- Measuring instrument, software, measurement standard, reference material or auxiliary apparatus or combination thereof necessary to realize a measurement process.
- Category Classification Ratio 16/58

3 of 5: Devices for measurement



Semantic Sensor Network

- Provides framework for describing sensors, their observations, and the procedures involved
- SSN is well-documented modular, and offers various examples to guide new users in its application
- Highly relevant to metrology
- Trialled and tested through large amount of use cases, constant expansion with additional models
- Can easily be integrated with other foundational ontologies



Allotrope Foundation Ontologies

- Provides a standard vocabulary and semantic model for representation of laboratory analytical processes and instruments involved
- Ontology application and documentation more complex and documentation not for beginners
- Relevant to metrology but specific application area: Currently largely applied to chemical analysis
- Large team working on ontologies, data models and data formats to be used and integrated
- Based on BFO facilitates easier integration with large amount of already available ontologies

4 of 5: Uncertainty evaluation and Traceability

NPL

Content and Explanation

- Characteristics of measurements and measuring systems such as accuracy, precision, uncertainty, traceability;
- Concepts related to metrological confirmation, validation, traceability chains.
- Codified mathematical notations (for example MathML).

4 of 5: Uncertainty evaluation and Traceability

15 out of 58 papers in literature review included this theme.

Highlight

- vim: Ontology for Units of Measurement
- **Paper:** https://doi.org/10.3390/electronics12183783
- **Resource:** https://zenodo.org/doi/10.5281/zenodo.8312872
- **Description:** an OWL ontology that captures concepts from the International Vocabulary of Metrology—Basic and general concepts (VIM) and its Chinese counterpart
- **Reasons:** easy (but not entirely FAIR) access to OWL files, developed using a systematic methodology, covers VIM terms
- **Case study:**
  - ?

5 of 5: Quality Infrastructure

Content and Explanation

- Organisations, people, labs, facilities related to metrological standardisation, accreditation, testing and calibration.

5 of 5: Quality Infrastructure

7 out of 58 papers in literature review included this theme.

Highlight

- The Organization Ontology
- **Paper:**
- **Resource:** https://www.w3.org/TR/vocab-org/
- **Description:** an OWL ontology for organizational structures
- **Reasons:** FAIR access to OWL files, W3C standard
- **Note:** we didn’t find a metrology-related case study, ORG on its own not enough to cover whole quality infrastructure

Addon: Data Management

- General tools for data curation, FAIR principle realisation and semantic annotation.

General

- DCTERMS: Up-to-date, authoritative specification of all metadata terms maintained by the Dublin Core™ Metadata Initiative
- DCAT: Data Catalogue is an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web.
- PROV-O: Provides a set of classes, properties, and restrictions that can be used to represent and interchange provenance information generated in different systems and under different contexts. (can be BFO-aligned)
- REPRODUCEME: The REPRODUCE-ME is an OWL2 ontology to describe a complete path of a scientific experiment.
- Quickgraph: Advanced annotation application built to produce high-quality training data for knowledge graph extraction from text-based content
- OOPS: (OntOlogy Pitfall Scanner!) is a web-based tool designed to detect potential pitfalls that could lead to modelling errors.
- FOOPS: FOOPS! (Findable, Accessible, Interoperable, and Reusable Ontology Pitfall Scanner) is a web-based tool designed to assess the compliance of vocabularies or ontologies against the FAIR principles.
- Many more..

BFO-based

- OBO: Open Biological and Biomedical Ontologies (OBO) provide a suite of high-quality, interoperable, free and open-source tools for sharing scientific knowledge and making new discoveries.
- Information Artifact Ontology (BFO and OBO based): The Information Artifact Ontology (IAO) is an ontology of information entities, originally driven by work by the OBI digital entity and realizable information entity branch.
- Many more..

Sources for ontologies

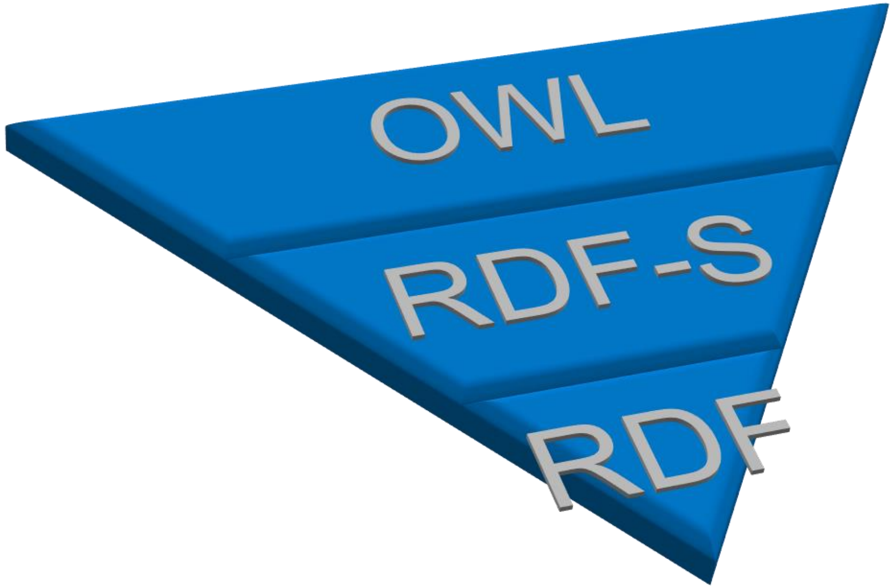
Linked Open Vocabularies (linkeddata.es)
A collection of reusable, linked data-friendly vocabularies covering a wide range of subjects like IoT, industry, environment, and more.

Welcome to the NCBO BioPortal | NCBO BioPortal (bioontology.org)
The most comprehensive repository of biomedical ontologies based on the OBO standard

Ontology Lookup Service (OLS) (ebi.ac.uk)
The Ontology Lookup Service (OLS) is a repository for biomedical ontologies that aims to provide a single point of access to the latest ontology versions.

OWL and the semantic web stack

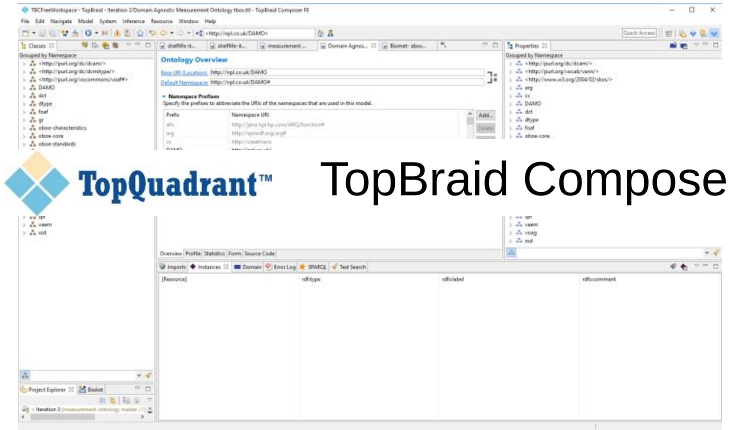
- RDF
  - Open triples (= Subject Predicate Object)
- RDFS
  - Closed vocabulary for RDF triples
- OWL 2
  - Constructs based on logic and set theories
- OWL formal language with several syntaxes:
  - RDF/XML, **TURTLE**, JSON-LD, **Functional**



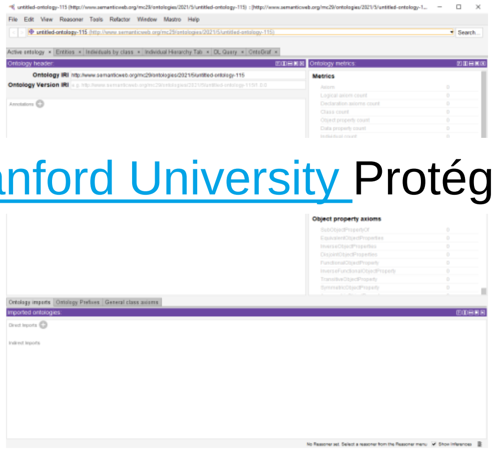
Some tools for ontologies

NPL

Edit ontologies via GUIs



TopQuadrant™ TopBraid Composer



Stanford University Protégé

Store/maintain/serve ontologies

GraphDB by ontotext

Apache Jena Fuseki

STARDOG

Process ontologies programmatically

dotNetRDF

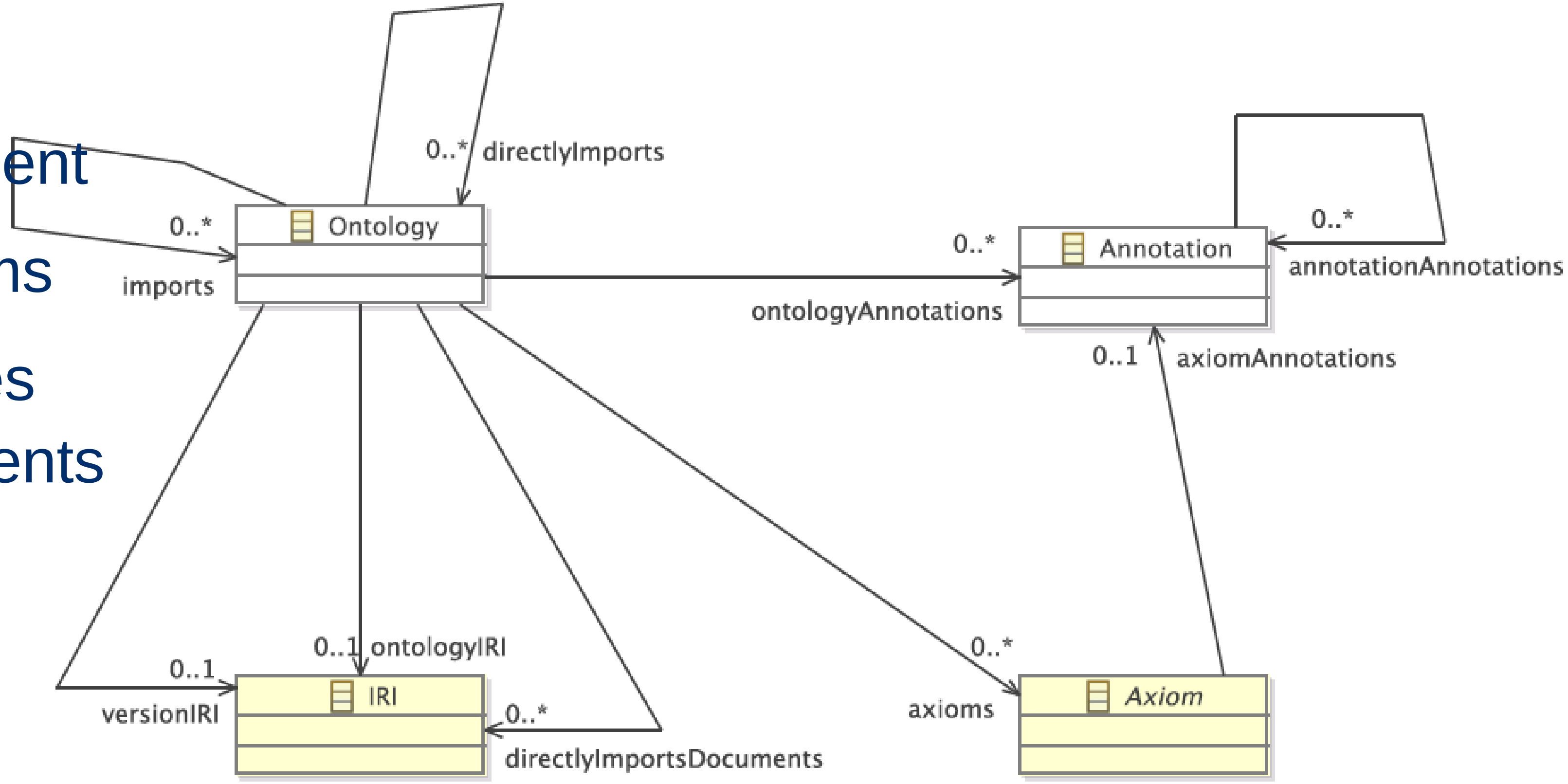
rdf4j

The OWL API

Owlready2

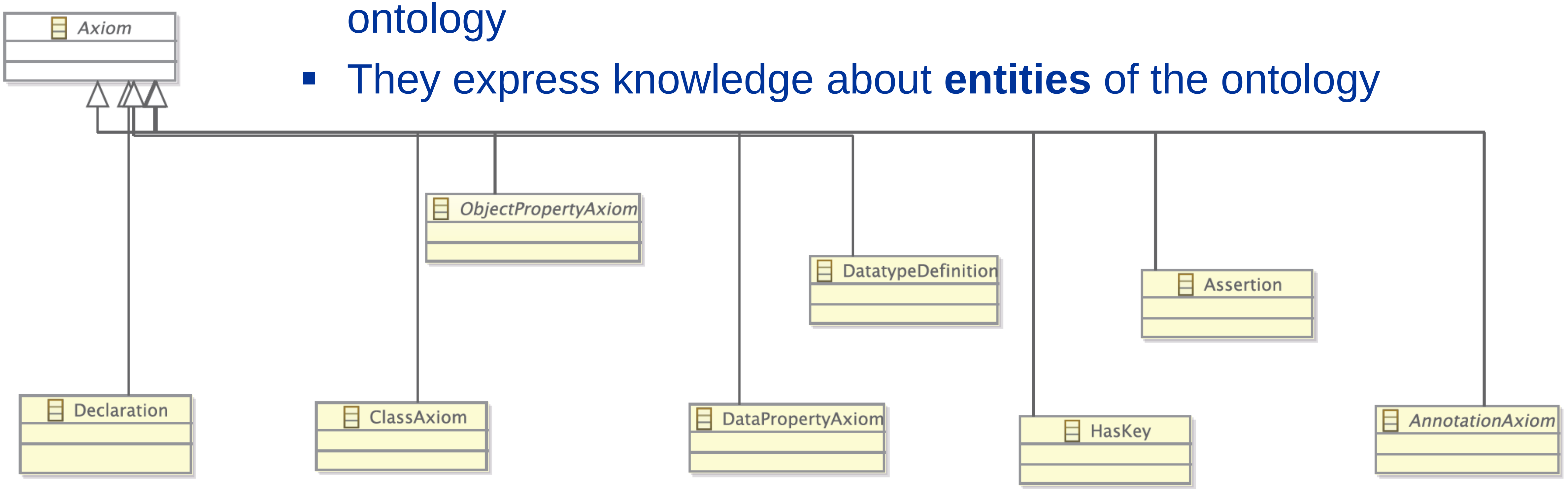
Features of OWL: ontologies

- Identified by a URI (or IRI), as any other ontology entity
- Lives in an ontology document
- Composed of a set of axioms
- Can refer to other ontologies from other ontology documents
- Optionally one or more ontology annotations



Features of OWL: axioms

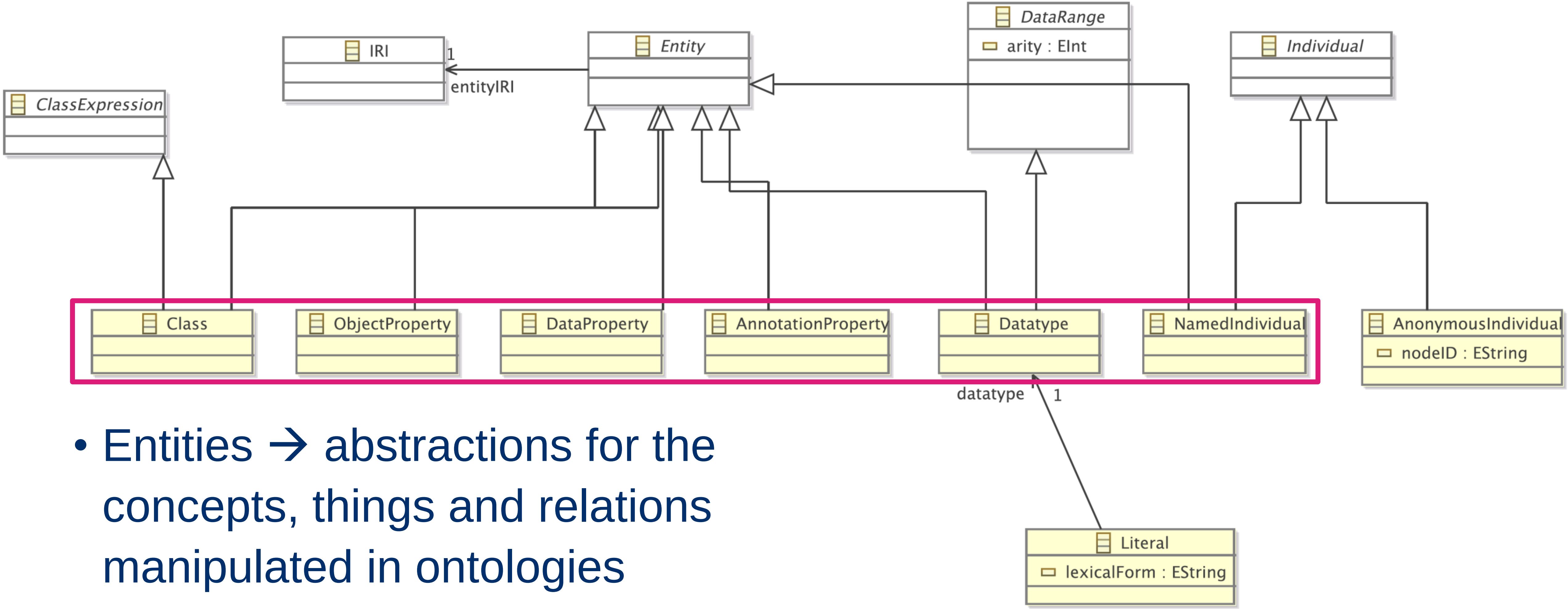
- Axioms = statements that are asserted to be true in the ontology
- They express knowledge about **entities** of the ontology



:someone rdf:type Person is a class assertion. (= someone is a Person)

:canSee rdfs:domain :Person is an object property axiom.
(the canSee property's domain is the Person class)

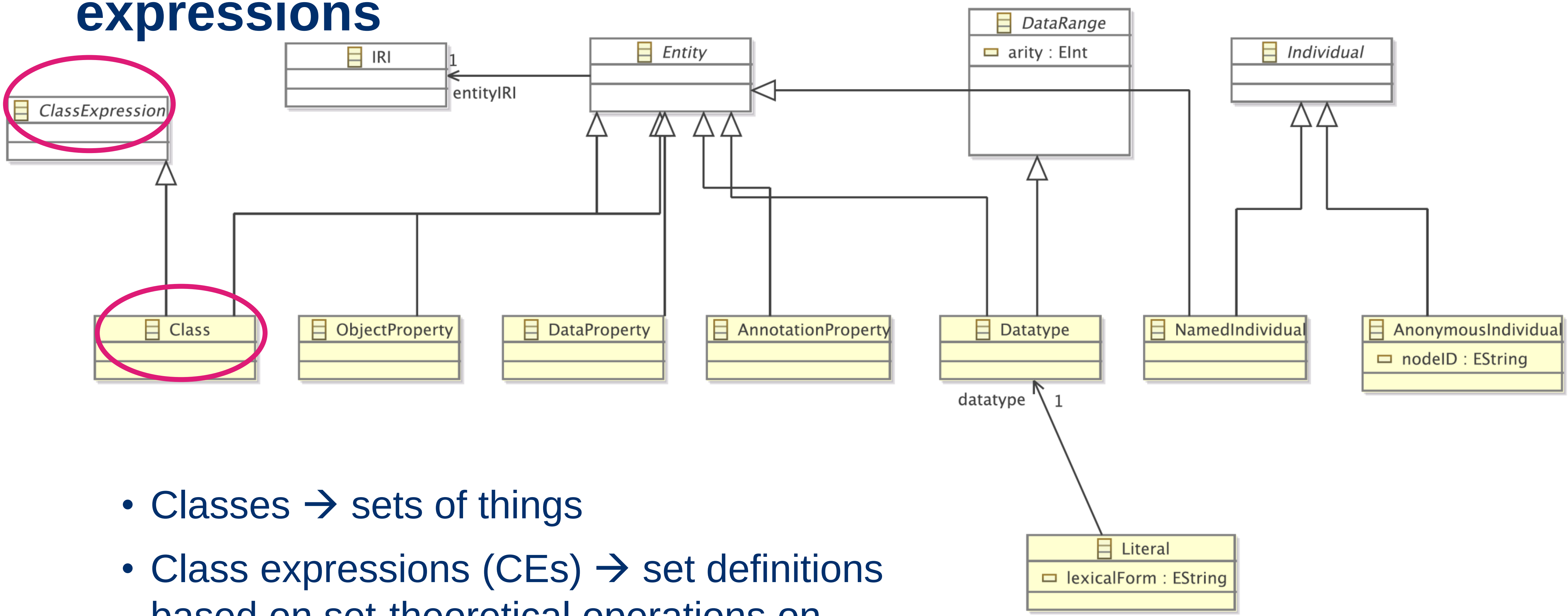
Features of OWL: entities



- Entities → abstractions for the concepts, things and relations manipulated in ontologies
- Entities have a URI (or IRI)

:someone is a named individual, its has a URI http://npl.co.uk/ontology101#someone

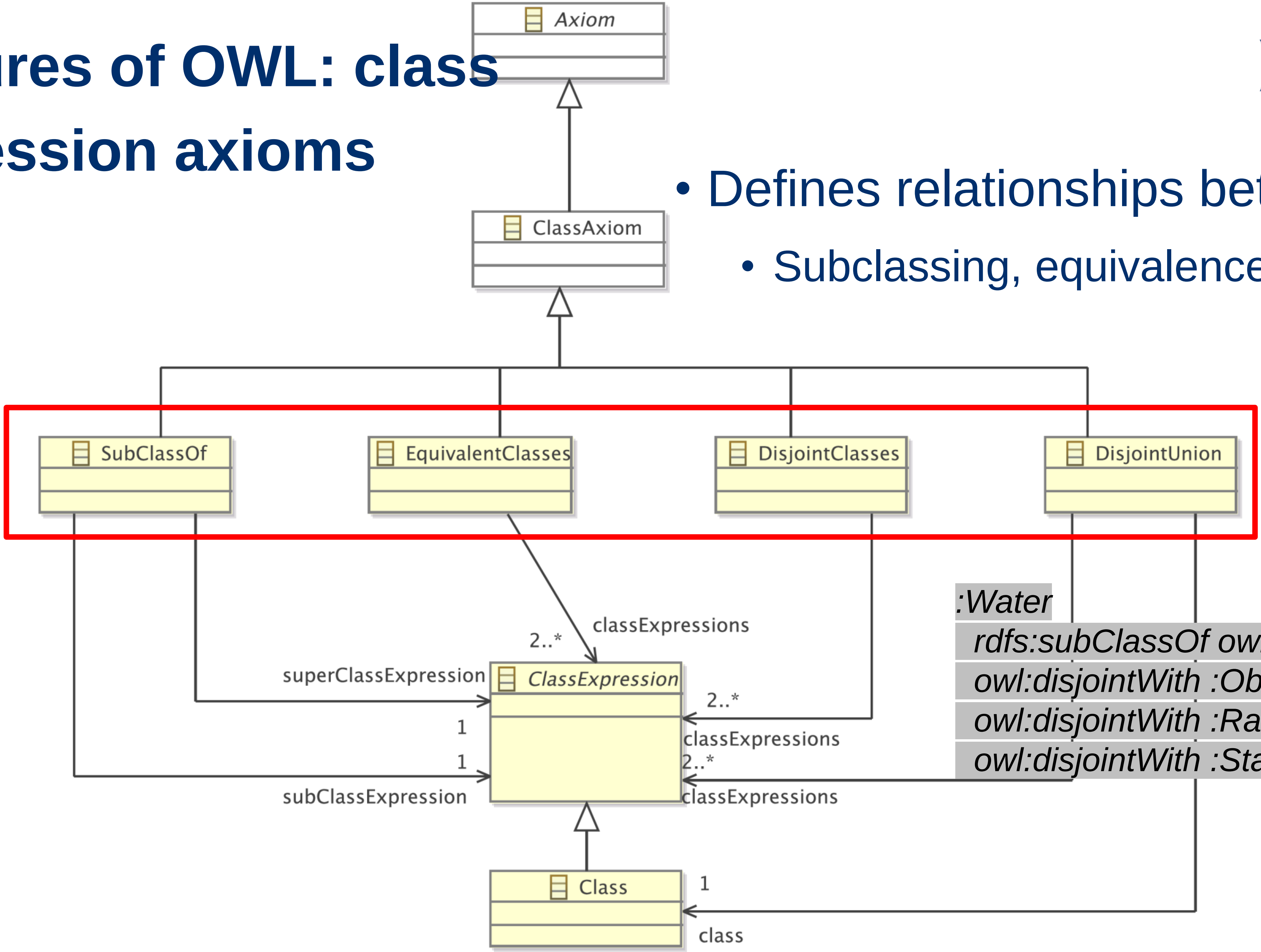
Features of OWL: classes and class expressions



- Classes → sets of things
- Class expressions (CEs) → set definitions based on set-theoretical operations on classes or CEs: union, complement, intersection

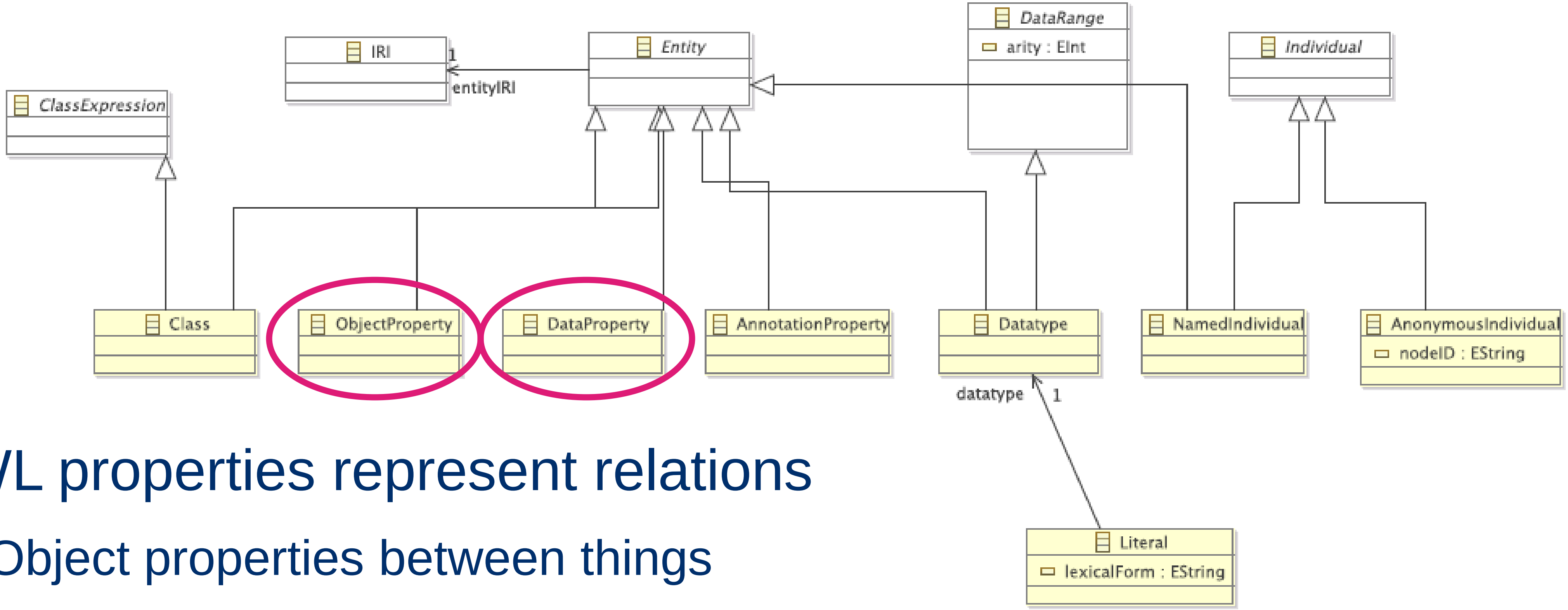
Features of OWL: class expression axioms

- Defines relationships between class
  - Subclassing, equivalence, disjointness




```
:Water
  rdfs:subClassOf owl:Thing ;
  owl:disjointWith :Observer ;
  owl:disjointWith :Rainbow ;
  owl:disjointWith :Star
```

Features of OWL: properties



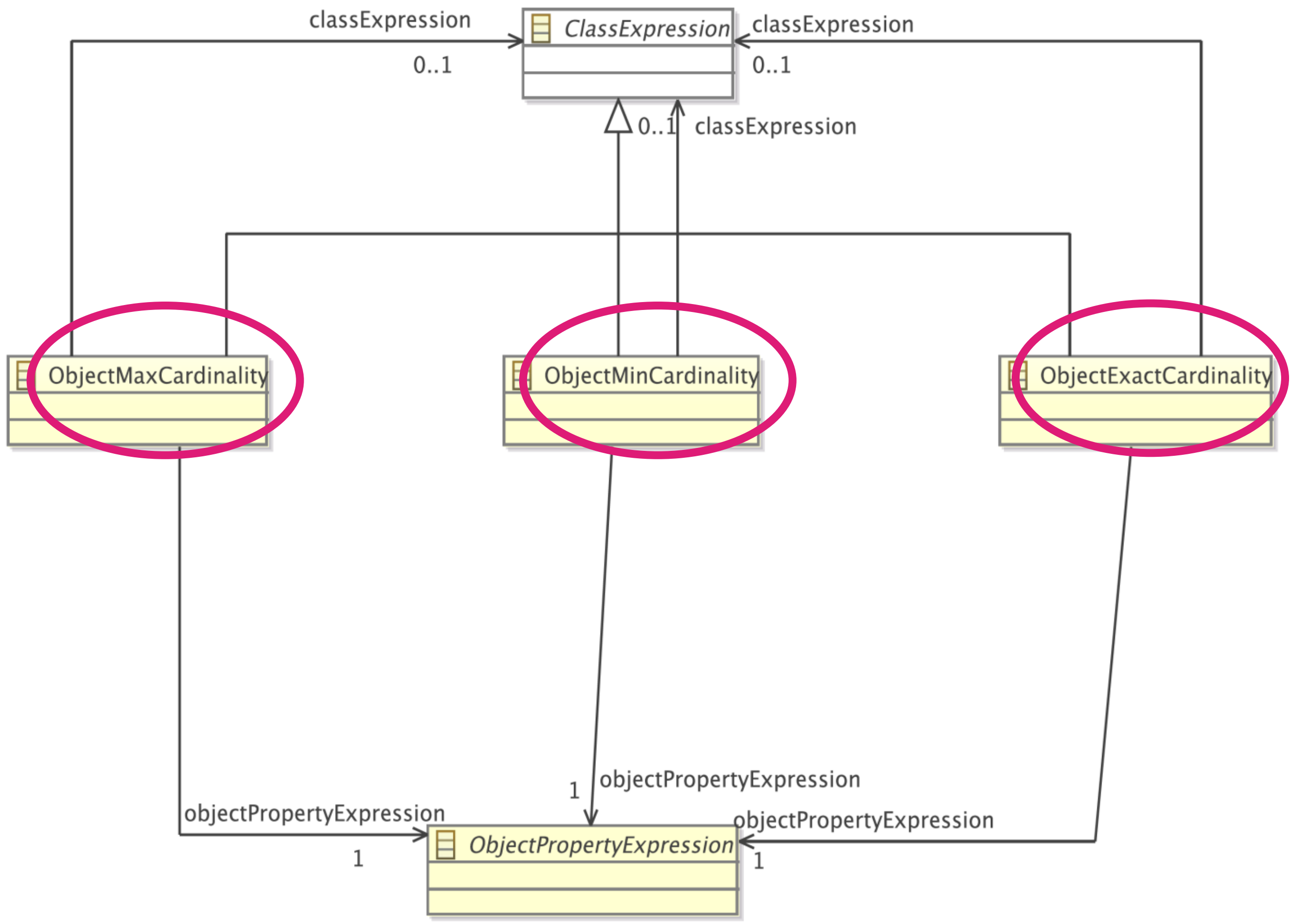
- OWL properties represent relations
  - Object properties between things
  - Data properties between things and literal values

:someone :isFacing :droplets (*isFacing* is an object property)

:aRainbow :hasNumberOfColours 7 (*hasNumberOfColours* is a data property)

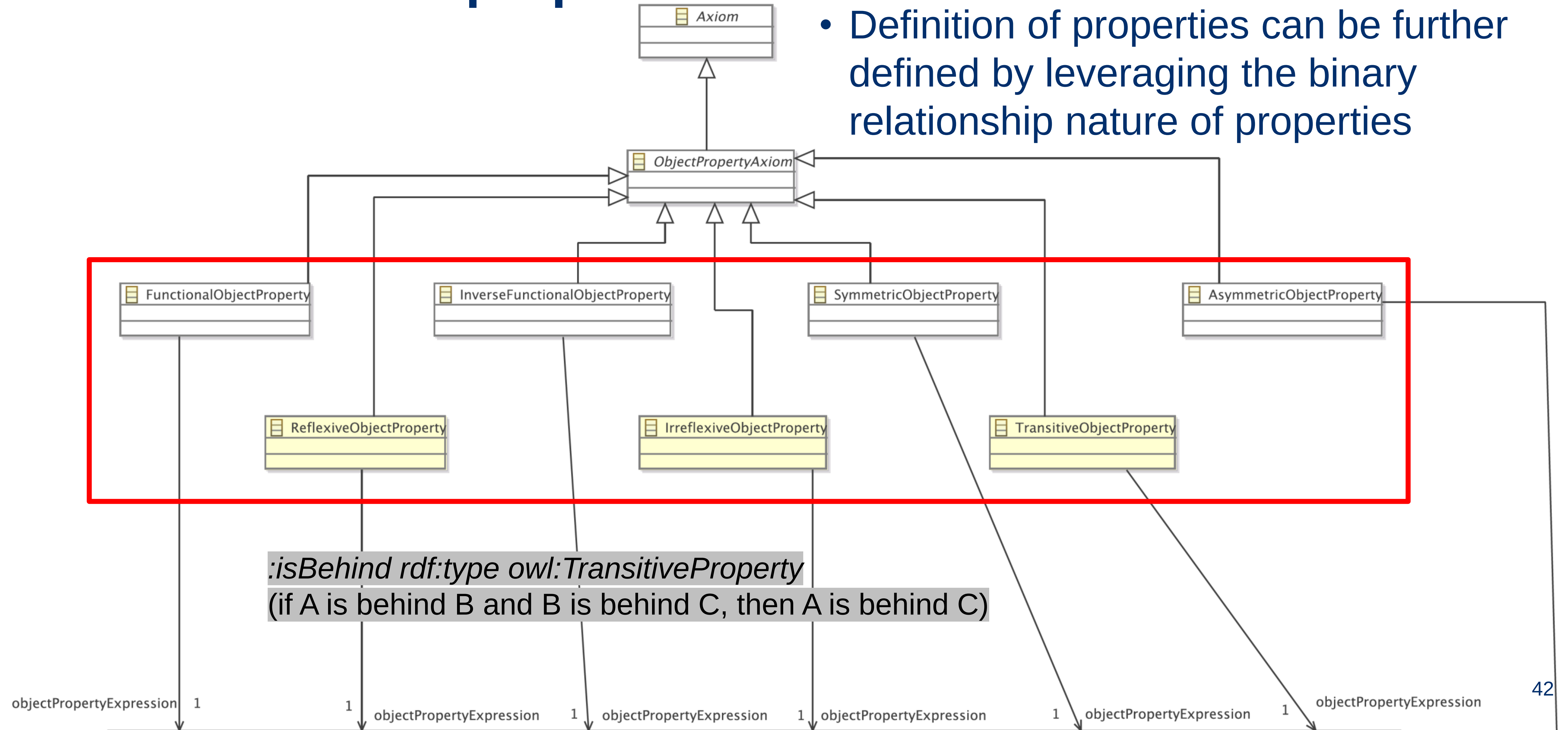
Features of OWL: class expressions from restrictions on properties

- Cardinality (from set theory)
- Class expressions can be defined further by placing cardinality restrictions on properties
- Restricts the range of the property

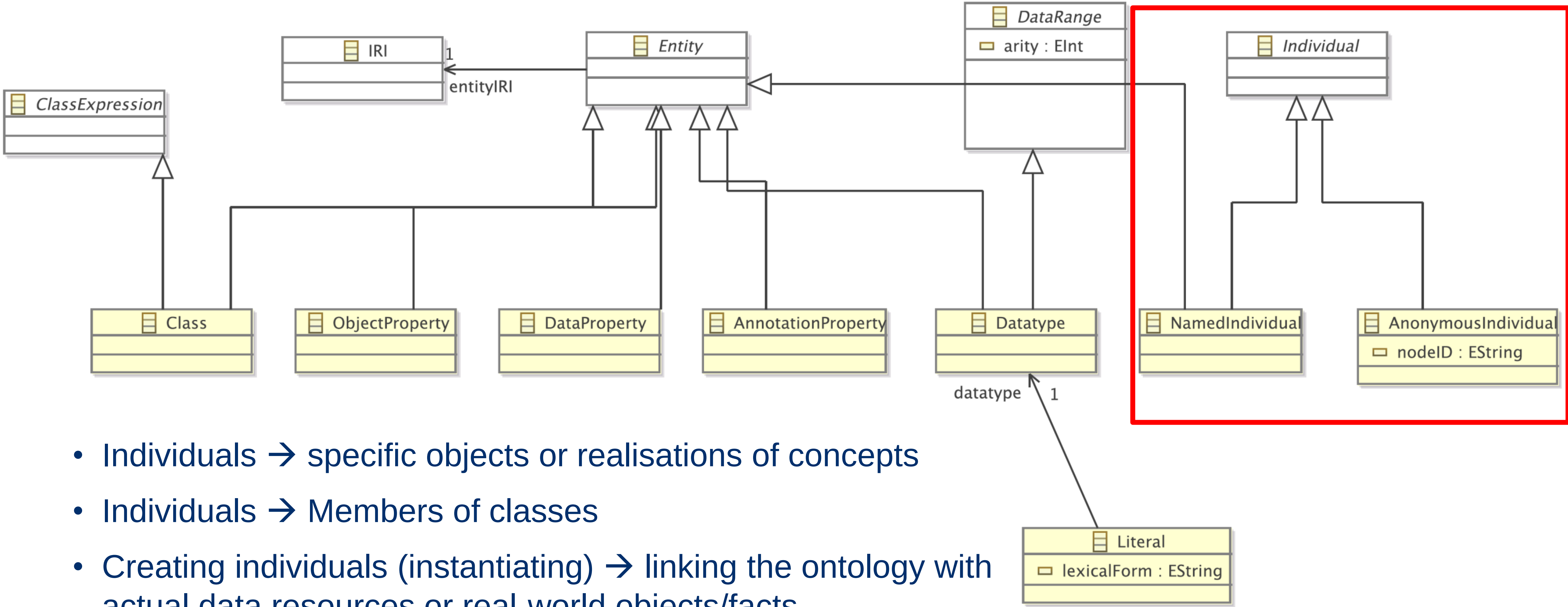


Features of OWL: class expressions from restrictions on properties

- Definition of properties can be further defined by leveraging the binary relationship nature of properties



Features of OWL: individuals



- Individuals → specific objects or realisations of concepts
- Individuals → Members of classes
- Creating individuals (instantiating) → linking the ontology with actual data resources or real-world objects/facts

NPL

National Physical Laboratory

npl.co.uk