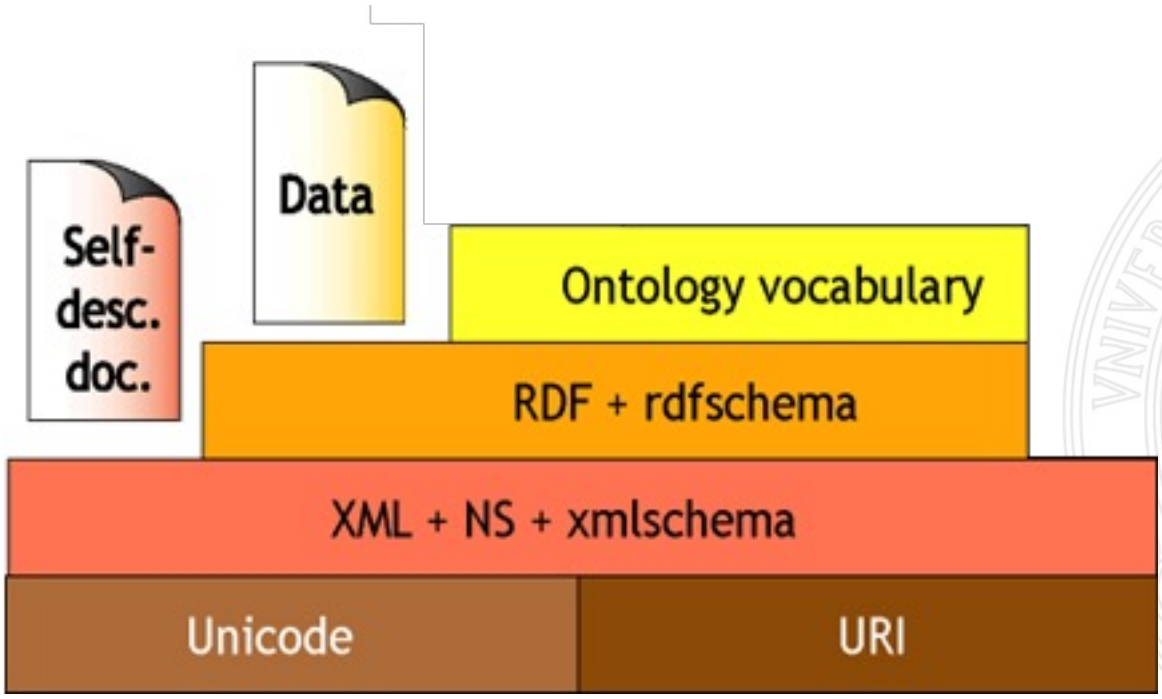
Introduction to bio-ontologies

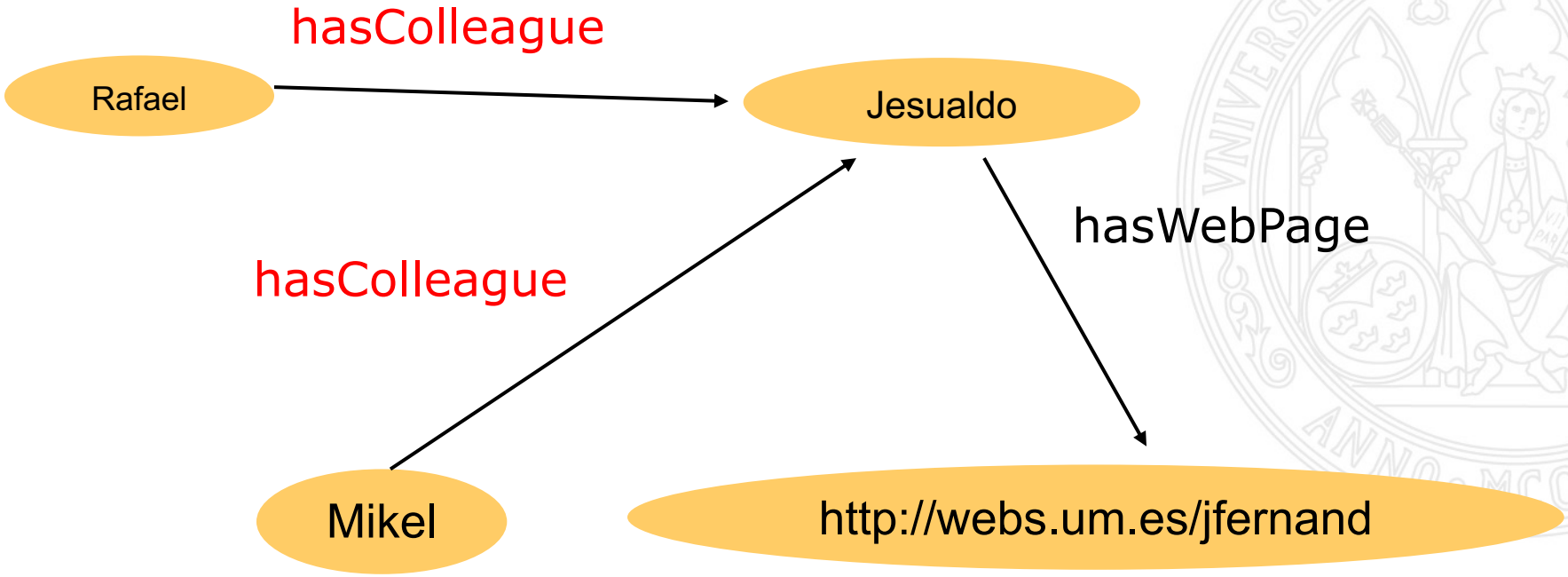
UNIVERSIDAD DE
MURCIA

(Part of) the Semantic Web Stack



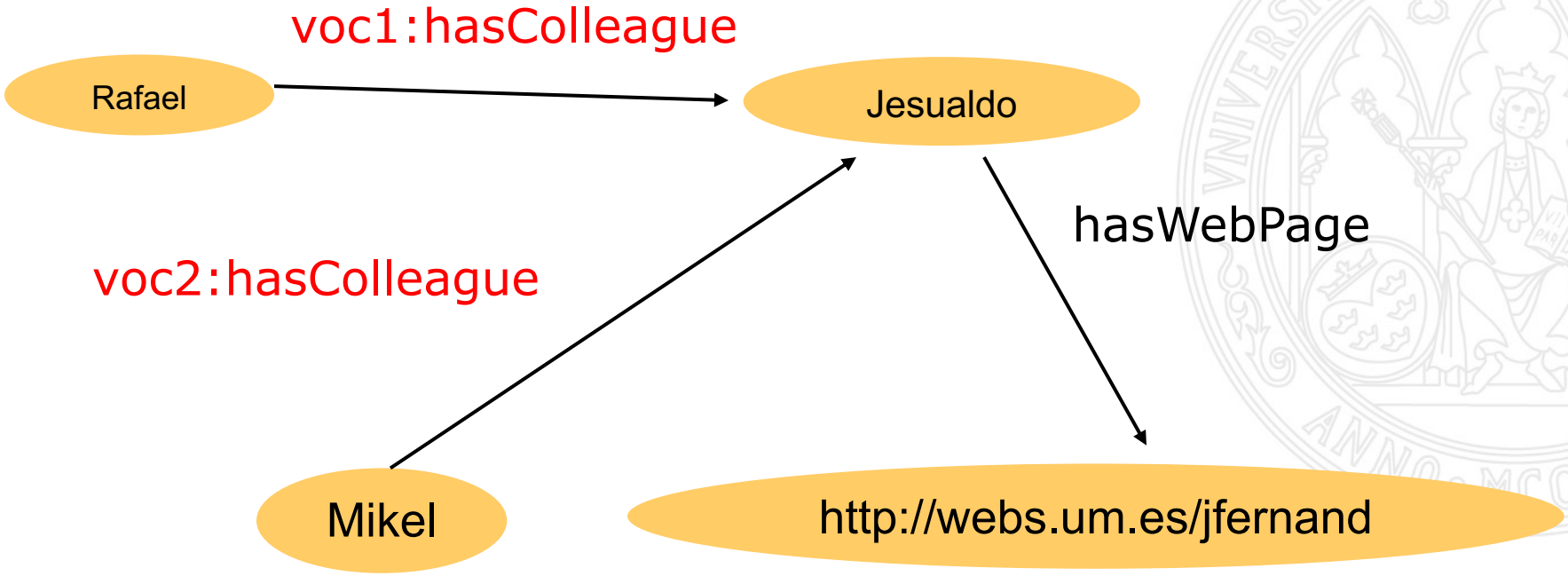
What if…

I think that RDF is OK but that there is too much freedom in the definition and use of properties



What if…

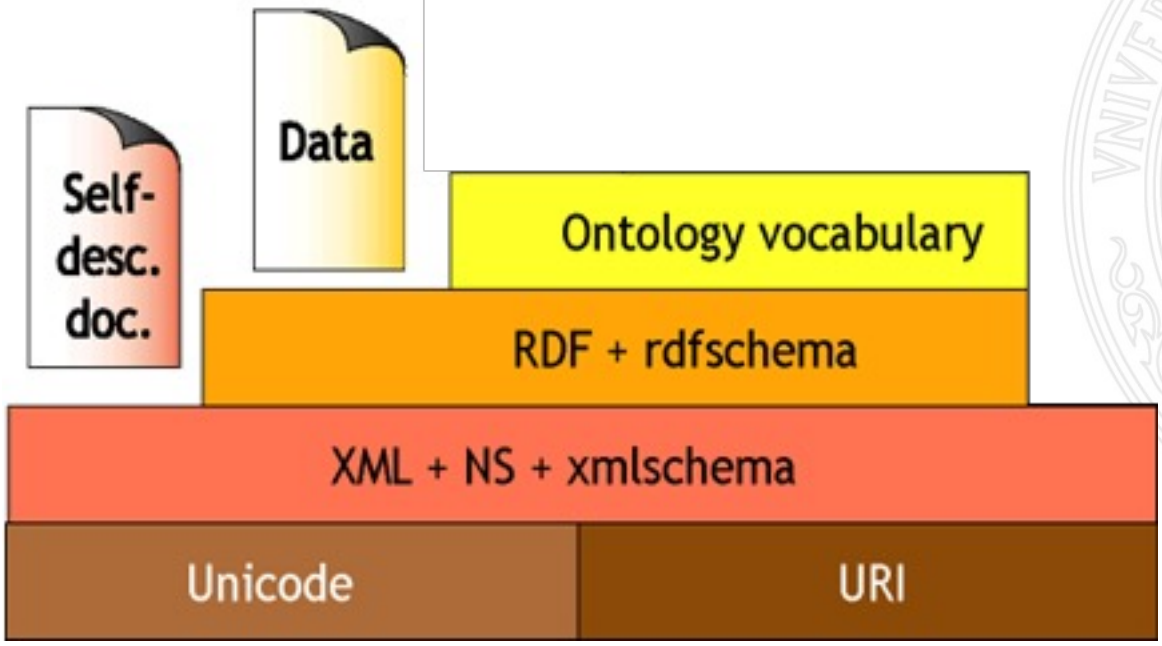
I think that RDF is OK but that there is too much freedom in the definition and use of properties



The Semantic Web Stack

- Ontology: **Formal representation of the domain knowledge**:

concepts+properties+restrictions



Types of ontologies

UNIVERSIDAD DE MURCIA

- Top-level ontologies: “domain-independent” knowledge: Processes, objects, physical entity,..
- Domain ontologies: gene, protein, volume, unit of measurement
- Application ontologies: ”my genes” (for my particular application)

Applications of ontologies in bioinformatics

UNIVERSIDAD DE MURCIA

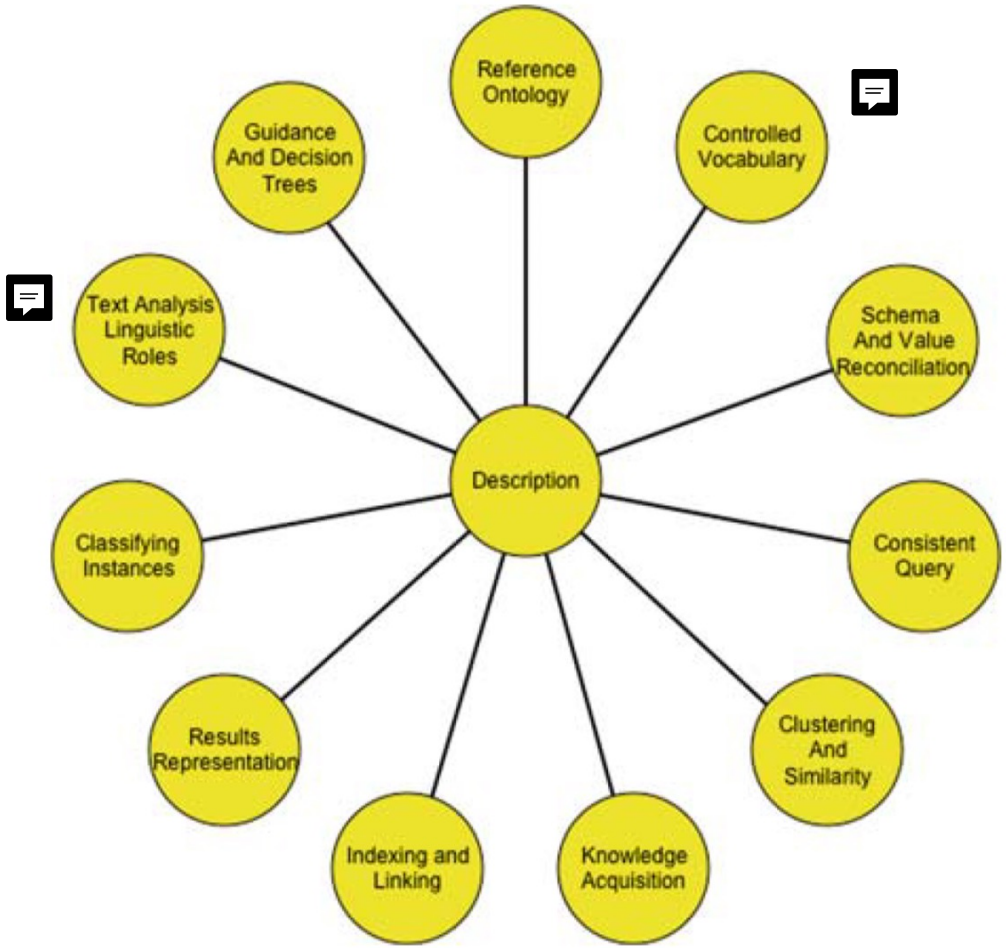


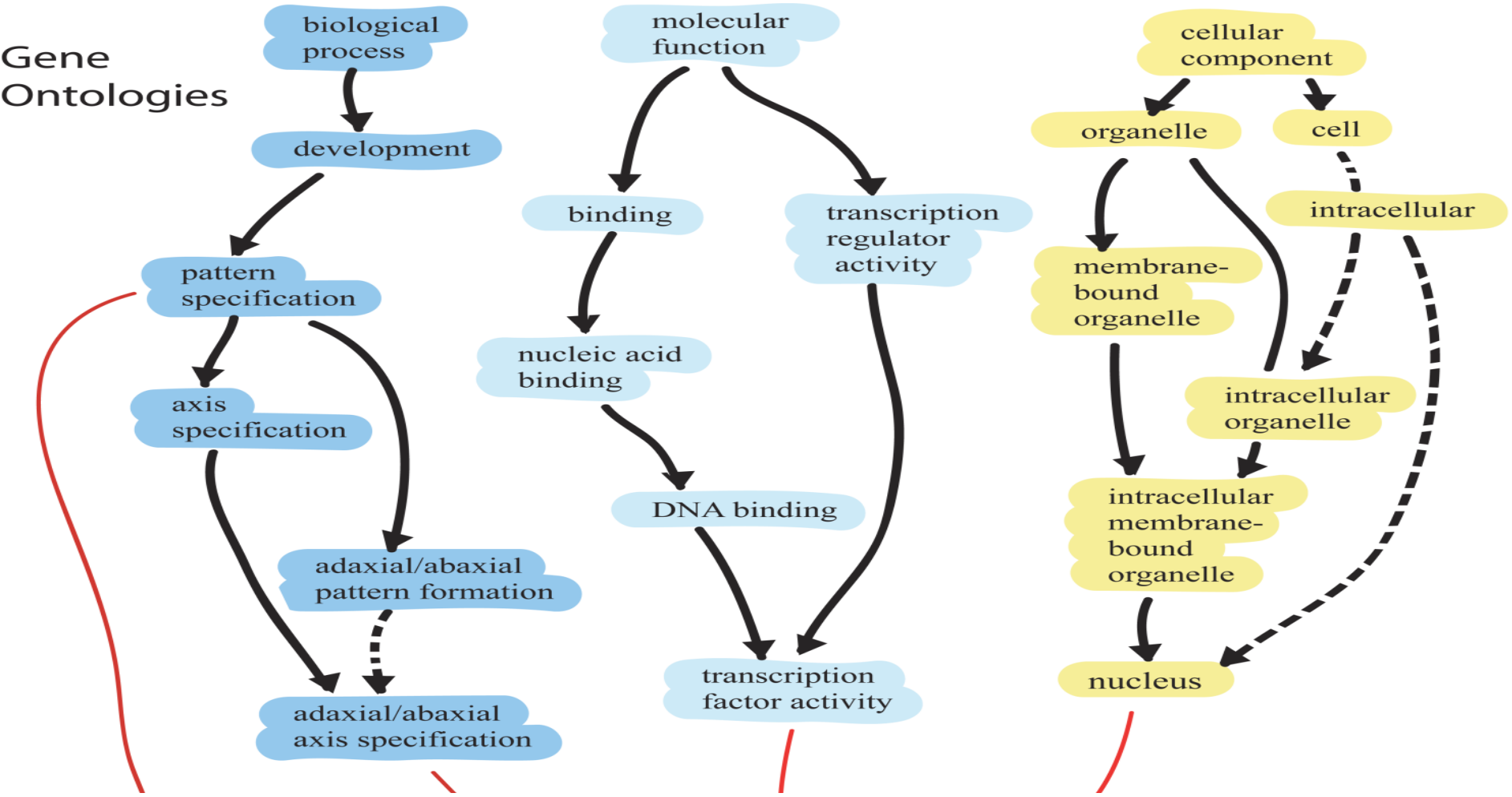
Fig. 1. A classification scheme for the uses of ontology and ontology-like artefacts within biology

Stevens, Robert, and Phillip Lord. "Application of ontologies in bioinformatics." *Handbook on ontologies*. Springer, Berlin, Heidelberg, 2009.

735-756.

The Gene Ontology (https://www.geneontology.org)

UNIVERSIDAD DE MURCIA



The Sequence Ontology (http://www.sequenceontology.org)

UNIVERSIDAD DE MURCIA

- sequence_attribute
- sequence_collection
- sequence_feature
  - junction
  - region
    - biological_region
    - biomaterial_region
    - experimental_feature
      - assembly
        - fragment_assembly
        - sequence_assembly
          - contig
          - expressed_sequence_as
          - golden_path
          - partial_genomic_sequer
          - tiling_path
          - virtual_sequence
      - assembly_component
      - consensus_region
      - conserved_region
      - experimentally_defined_bindi
      - high_identity_region

Term Only as OBO Format Export

sequence_assembly (CURRENT_SVN)	
SO Accession:	SO:0000353 (SOWiki)
Definition:	A sequence of nucleotides that has been algorithmically derived from an alignment of two or more different sequences.
Synonyms:	sequence assembly
DB Xrefs:	SO: ma
Parent:	assembly (SO:0001248)
Children:	partial_genomic_sequence_assembly (SO:0001876)
	expressed_sequence_assembly (SO:0001428)
	gap (SO:0000730)
	contig (SO:0000149)
	tiling_path (SO:0000472)
	virtual_sequence (SO:0000499)
	golden_path (SO:0000688)

In the image below graph nodes link to the appropriate terms. Clicking the image background will toggle the image between large and small formats.

Ontology repository

UNIVERSIDAD DE MURCIA

- Database of ontologies
- Different types of ontologies in a repository: domain, top-level, application ontologies
- Domain-specific repositories
- Quality control/assurance

Biomedical ontologies: Bioportal

UNIVERSIDAD DE MURCIA

http://bioportal.bioontology.org

Statistics

Ontologies	1,086
Classes	14,646,888

#	name	# visits
1	Medical Dictionary for Regulatory Activities Terminology (MedDRA) (MEDDRA)	23,031
2	SNOMED CT (SNOMEDCT)	10,222
3	RxNORM (RXNORM)	7,780
4	National Drug Data File (NDDF)	3,785
5	Foundational Model of Anatomy (FMA)	2,369
6	Read Codes, Clinical Terms Version 3 (CTV3) (RCD)	1,977
7	National Cancer Institute Thesaurus (NCIT)	1,766
8	Medical Subject Headings (MESH)	1,376
9	National Drug File - Reference Terminology (NDFRT)	1,348
10	Logical Observation Identifier Names and Codes (LOINC)	1,026
11	EDAM - Bioscientific data analysis ontology (EDAM)	982

Bioportal: Ontologies

UNIVERSIDAD DE MURCIA

Category

- All Organisms (31)
- Anatomy (71)
- Animal Development (16)
- Animal Gross Anatomy ...
- Arabidopsis (3)
- Biological Process (54)

Group

- BIBLIO (10)
- BIS (3)
- CGIAR (1)
- CTSA (6)
- OBO_Foundry (11)
- PSI (4)

Format

- OBO (100)
- OWL (638)
- SKOS (26)
- UMLS (31)

Bioportal: Ontologies

UNIVERSIDAD DE MURCIA

Medical Dictionary for Regulatory Activities Terminology (MedDRA) (MEDDRA)

notes 1 | projects 10 | classes 74,056

MedDRA is an international medical terminology with an emphasis on use for data entry, retrieval, analysis, and display

Uploaded: 9/25/20

SNOMED CT (SNOMEDCT)

notes 3 | projects 23 | classes 359,263

SNOMED Clinical Terms

Uploaded: 9/25/20

RxNORM (RXNORM)

projects 7 | classes 109,013

RxNorm Vocabulary

Uploaded: 9/25/20

Information about the ontology

UNIVERSIDAD DE MURCIA

Gene Expression Ontology

Last uploaded: December 16, 2015

Summary | Classes | Properties | Notes | Mappings | Widgets

Details

Acronym	GEXO
Visibility	Public
Description	An application ontology for the domain of gene expression. The ontology integrates fragments of GO and MI with data from GOA, IntAct, UniProt, NCBI, and orthology relations using ontological properties from RO and ISO.
Status	Beta
Format	OWL
Contact	Vladimir Mironov, vladimir.n.mironov@gmail.com
Categories	Biological Process

Submissions

Version	Released	Uploaded	Downloads
2015 (Parsed, Indexed, Metrics, Annotator)	12/16/2015	12/16/2015	OWL | CSV | RDF/XML
2015 (Archived)	09/08/2012	03/06/2015	OBO

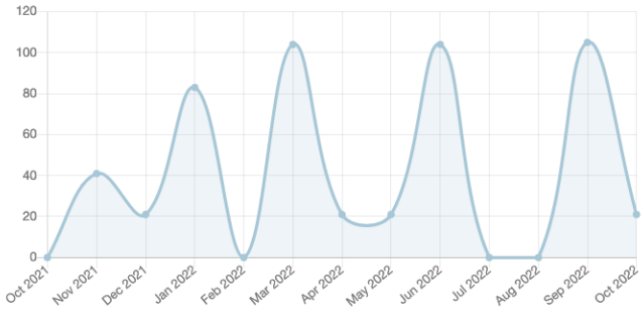
Views of GEXO

No views of GEXO available

Metrics

Classes	166,254
Individuals	157,102
Properties	12
Maximum depth	21
Maximum number of children	50,996
Average number of children	49
Classes with a single child	968
Classes with more than 25 children	150
Classes with no definition	0

Visits

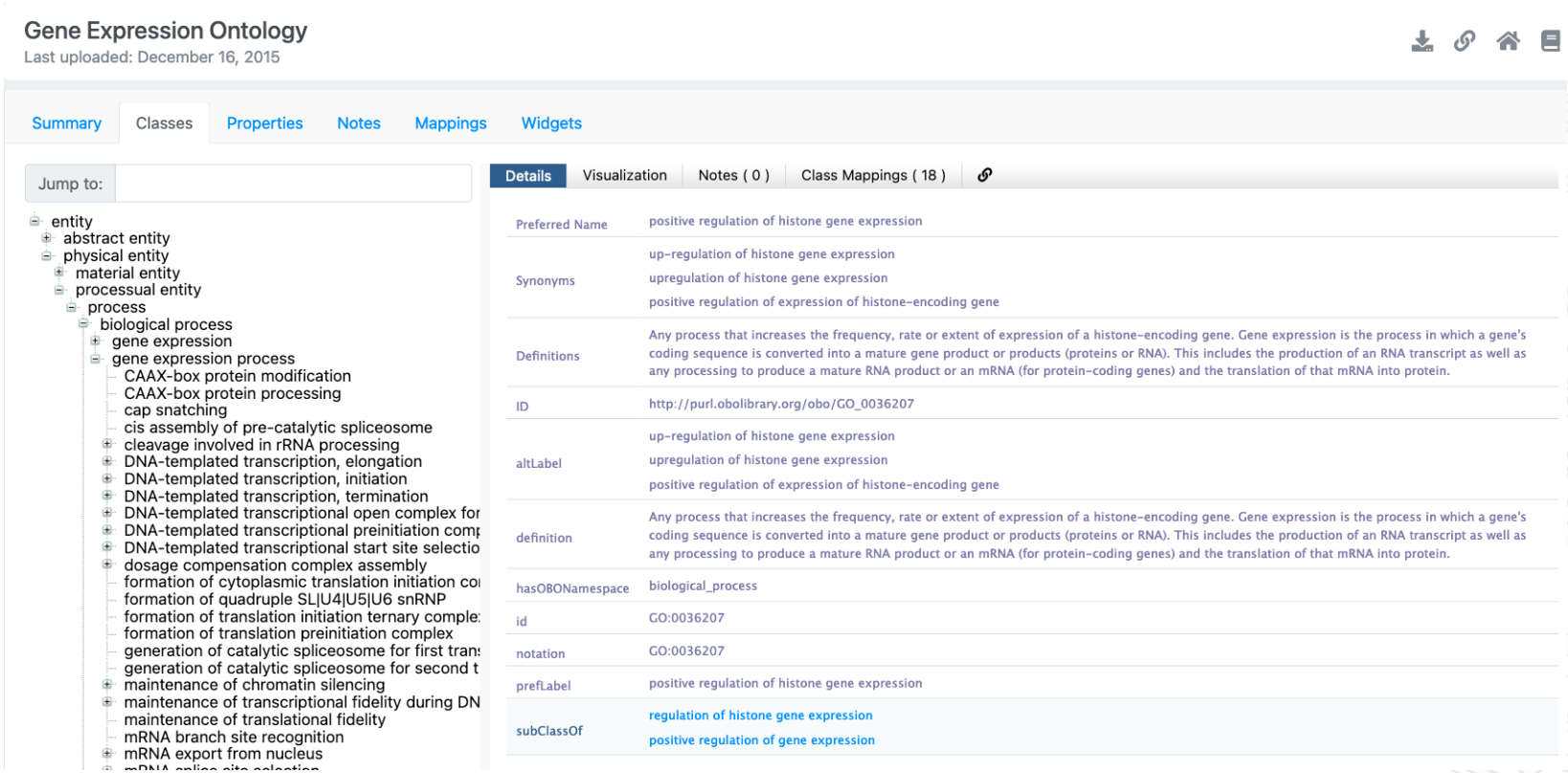


Projects using GEXO

No projects using GEXO

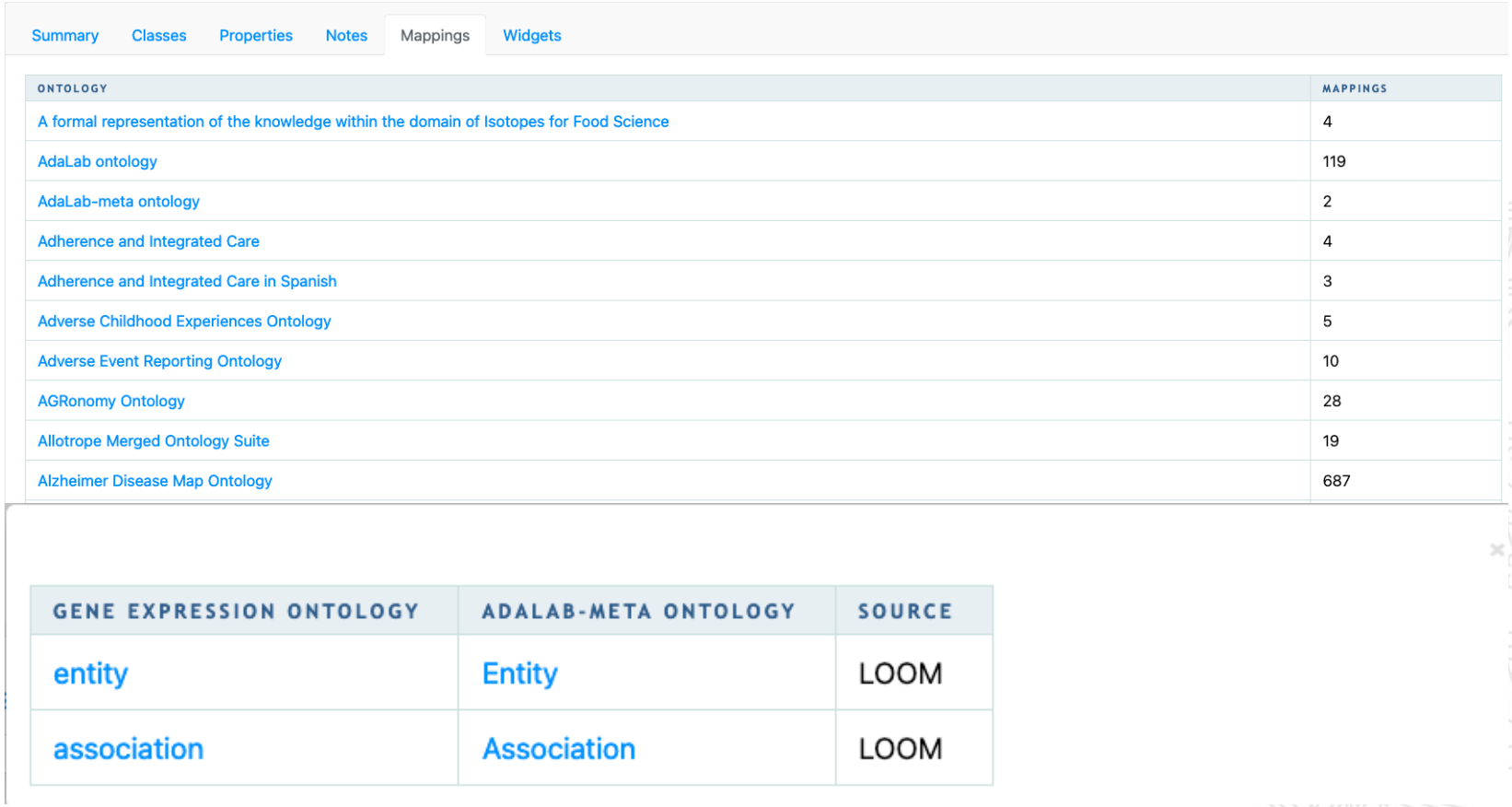
Accessing the ontology

UNIVERSIDAD DE MURCIA



Term mappings

UNIVERSIDAD DE MURCIA



Summary | Classes | Properties | Notes | Mappings | Widgets

ONTOLOGY	MAPPINGS
A formal representation of the knowledge within the domain of Isotopes for Food Science	4
AdaLab ontology	119
AdaLab-meta ontology	2
Adherence and Integrated Care	4
Adherence and Integrated Care in Spanish	3
Adverse Childhood Experiences Ontology	5
Adverse Event Reporting Ontology	10
AGRonomy Ontology	28
Allotrope Merged Ontology Suite	19
Alzheimer Disease Map Ontology	687

GENE EXPRESSION ONTOLOGY	ADALAB-META ONTOLOGY	SOURCE
entity	Entity	LOOM
association	Association	LOOM

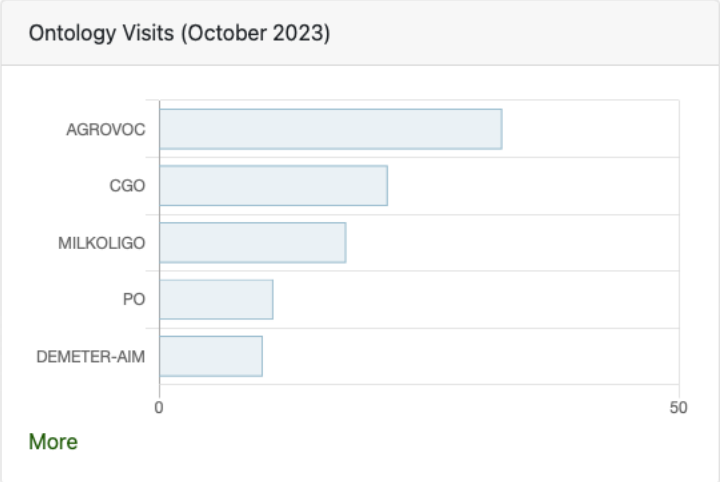
- These mappings do not mean equivalency

Agroportal

http://agroportal.lirmm.fr/

UNIVERSIDAD DE MURCIA

- It uses the BioPortal infrastructure for agriculture



AgroPortal Statistics	
Ontologies	158
Classes	688,240
Individuals	2,987,613
Projects	59
Users	390

Agroportal

http://agroportal.lirmm.fr/

UNIVERSIDAD DE MURCIA

- It uses the BioPortal infrastructure for agriculture

Agroportal

http://agroportal.lirmm.fr/

UNIVERSIDAD DE MURCIA

Category

- Agricultural Resear... (23)
- Animal Science an... (10)
- Biodiversity and Ec... (17)
- Breeding and Gene... (1)
- Farms and Farming... (7)
- Fisheries and Aqua... (2)
- Food Security (2)
- Food and Human ... (6)

Group

- AGBIODATA (18)
- AGROLD (5)
- CROP (37)
- INRAE (30)
- OBO-FOUNDRY (23)
- RICE (24)
- SEMANDIV (11)
- WHEAT (19)

Format

- OBO (12)
- OWL (100)
- SKOS (11)
- UMLS (2)

Ontology Content

- Notes (3)
- Reviews (2)
- Projects (110)
- Summary Only (0)

Is of Type

- Application Ontology (44)
- Core Ontology (2)
- Domain Ontology (53)
- Task Ontology (0)
- Upper Level Ontol... (6)
- Vocabulary (3)

Natural Language

- German (2)
- English (115)
- French (15)
- Italian (2)
- Portuguese (3)
- Spanish (4)

Formality Levels

- Classification sche... (2)
- Dictionary (0)
- Gazetteer (0)
- Glossary (0)
- List (0)
- Name authority list (0)
- Ontology (99)
- Semantic network (1)

OBO Foundry repository

http://www.obofoundry.org/

UNIVERSIDAD DE MURCIA

The Open Biological and Biomedical Ontology (OBO) Foundry

Community development of interoperable ontologies for the biological sciences

Learn about OBO best practices and community resources

- More about the OBO Foundry
- OBO Foundry principles
- OBO tutorial
- Ontology browsers, tutorials, and tools

Participate

- Join the OBO mailing list
- OBO Foundry Operations and Working Groups
- Submit bug reports or suggestions for improvement via GitHub
- Submit your ontology to be considered for inclusion in the OBO Foundry

OBO Library: find, use, and contribute to community ontologies

The table below lists current OBO ontologies (in alphabetical order, but with the ontologies that have been manually reviewed by the OBO Foundry listed first, and obsolete ontologies listed last).

Download table as: [YAML | JSON-LD | RDF/Turtle]

bfo	Basic Formal Ontology CC BY	The upper level ontology upon which OBO Foundry ontologies are built. Detail	
chebi	Chemical Entities of Biological Interest CC BY	A structured classification of molecular entities of biological interest focusing on 'small' chemical compounds. Detail	
doid	Human Disease Ontology PUBLICDOMAIN	An ontology for describing the classification of human diseases organized by etiology. Detail	

OBO Foundry repository

UNIVERSIDAD DE MURCIA

Gene Ontology

An ontology for describing the function of genes and gene products

Follow @news4go 2,064 followe

OntoBee AberOWL OLS AmiGO

The goal of the GeneOntology (GO) project is to provide a uniformway to describe the functions of gene products from organisms across all kingdoms of life and thereby enable analysis of genomic data

Products

go.owl	GO (OWL edition)	The main ontology in OWL. This is self contained and does not have connections to other OBO ontologies [page]
go.obo	GO (OBO Format edition)	Equivalent to go.owl, in obo format [page]
go.json	GO (JSON edition)	Equivalent to go.owl, in obograph json format [page]
go/extensions/go-plus.owl	GO-Plus	The main ontology plus axioms connecting to select external ontologies, with subsets of those ontologies [page]
go/go-base.owl	GO Base Module	The main ontology plus axioms connecting to select external ontologies, excluding the external ontologies themselves [page]
go/extensions/go-plus.json	GO-Plus	As go-plus.owl, in obographs json format [page]
go/go-basic.obo	GO-Basic, Filtered, for use with legacy tools	Basic version of the GO, filtered such that the graph is guaranteed to be acyclic and annotations can be propagated up the graph. The relations included are is a, part of, regulates, negatively regulates and positively regulates. This version excludes relationships that cross the

ID Space	go
PURL	http://purl.obolibrary.org/obo/go.owl
License	CC BY 4.0
Review Date	2010
Review Document	n/a
Homepage	http://geneontology.org/
Contact	Suzi Aleksander
Trackers	https://github.com/geneontology/go-ontology/issues/
Domain	biology
Taxon	All life
Dependencies	uberon cl ncbitaxon ro go/extensions/go-bridge-to-nifstd.owl

View Edit PURL

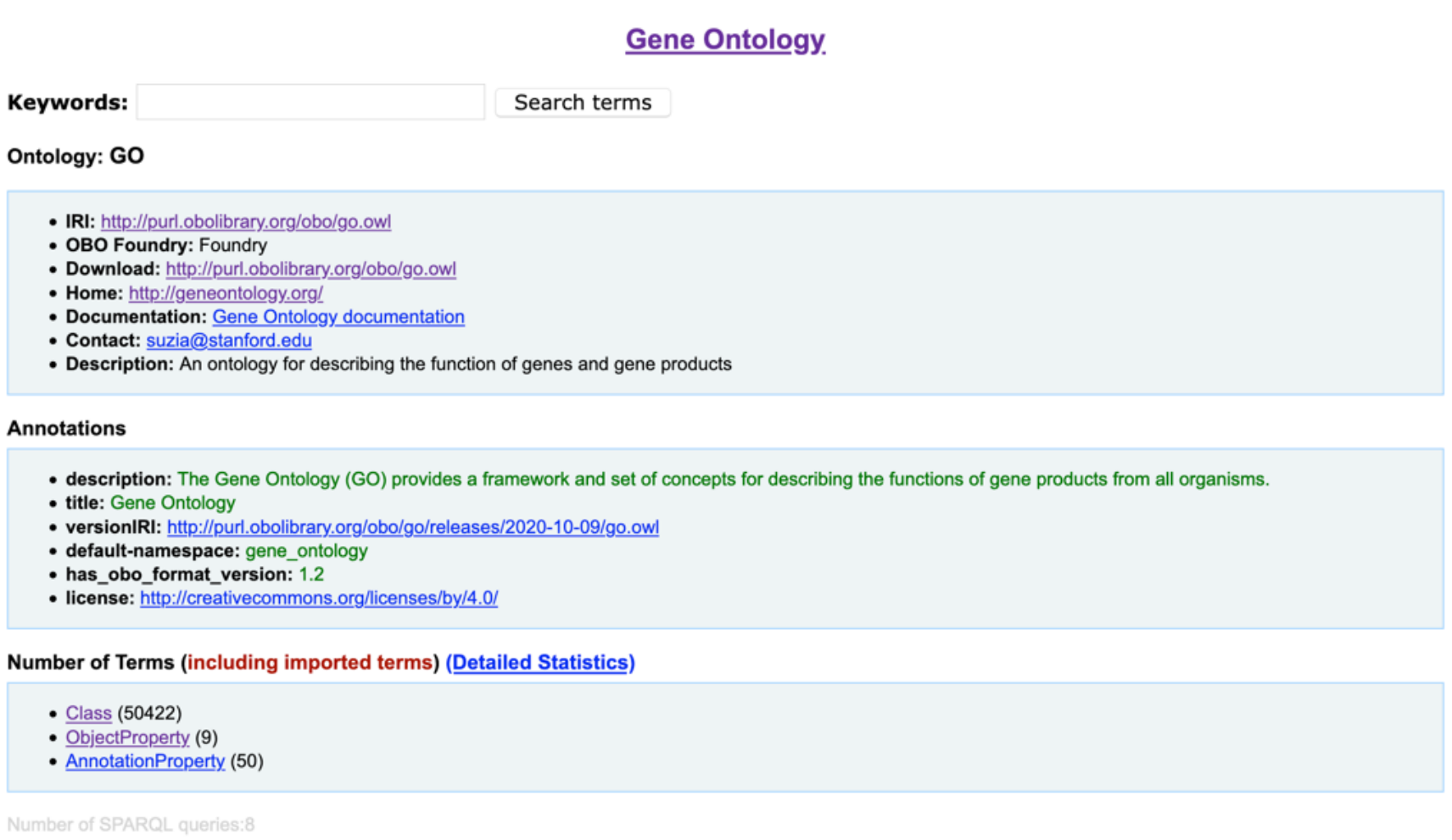
Generated by: _layouts/ontology_detail.html

See metadata guide

Edit the metadata for this page: go.md (GitHub will help you create a fork and pull request.)

Ontobee visualization

UNIVERSIDAD DE MURCIA

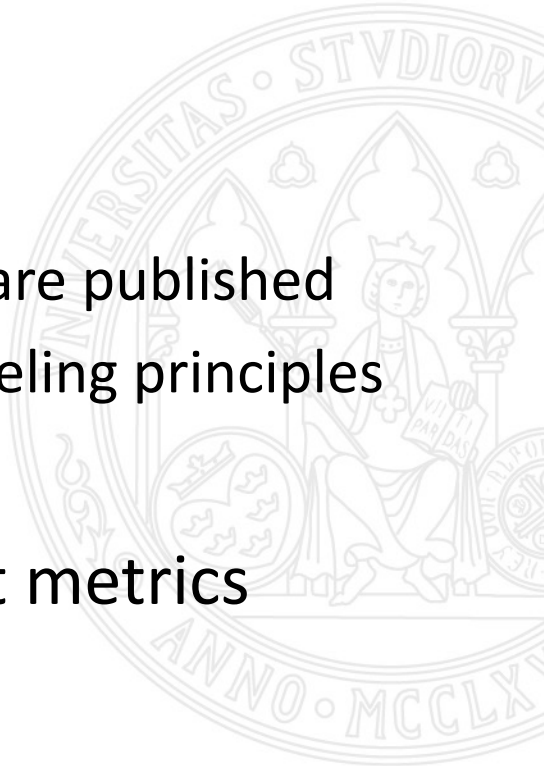


Check http://www.ontobee.org/ontology/GO?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FGO_0042995

Quality control in repositories

UNIVERSIDAD DE MURCIA

- BioPortal, AgroPortal:
  - Users can freely upload their ontologies
- OBO Foundry:
  - The consortium decides which ontologies are published
  - Ontologies are expected to adhere to modeling principles
- In general, not much information about metrics



OBO Foundry principles

UNIVERSIDAD DE MURCIA

http://www.obofoundry.org/principles/fp-000-summary.html

- Overview
- Open (principle 1)
- Common Format (principle 2)
- URI/Identifier Space (principle 3)
- Versioning (principle 4)
- Scope (principle 5)
- Textual Definitions (principle 6)
- Relations (principle 7)
- Documentation (principle 8)
- Documented Plurality of Users (principle 9)
- Commitment To Collaboration (principle 10)
- Locus of Authority (principle 11)
- Naming Conventions (principle 12)
- Maintenance (principle 16)

Ontology building

UNIVERSIDAD DE MURCIA

- Classes (Concepts): gene, protein (sets of individuals that share some properties)
- Individuals (Instances)
- Properties
  - Object properties (relations between instances of classes): regulates
  - Datatype properties (attributes, associations between instances of classes and data types): length
  - Annotations: label, synonym
- Axioms (restrictions, asserted conditions): gene codified in 1 organism

Ontology building

UNIVERSIDAD DE MURCIA

- **Terminological Knowledge (T-BOX)**
  - Axioms that describe the structure of the domain (conceptual schema, class level)

- **Knowledge about individuals (A-BOX)**
  - Axioms describing data, facts (instance level)

    HappyFather(John)
    hasChild(John, Mary)

Web Ontology Language Requirements

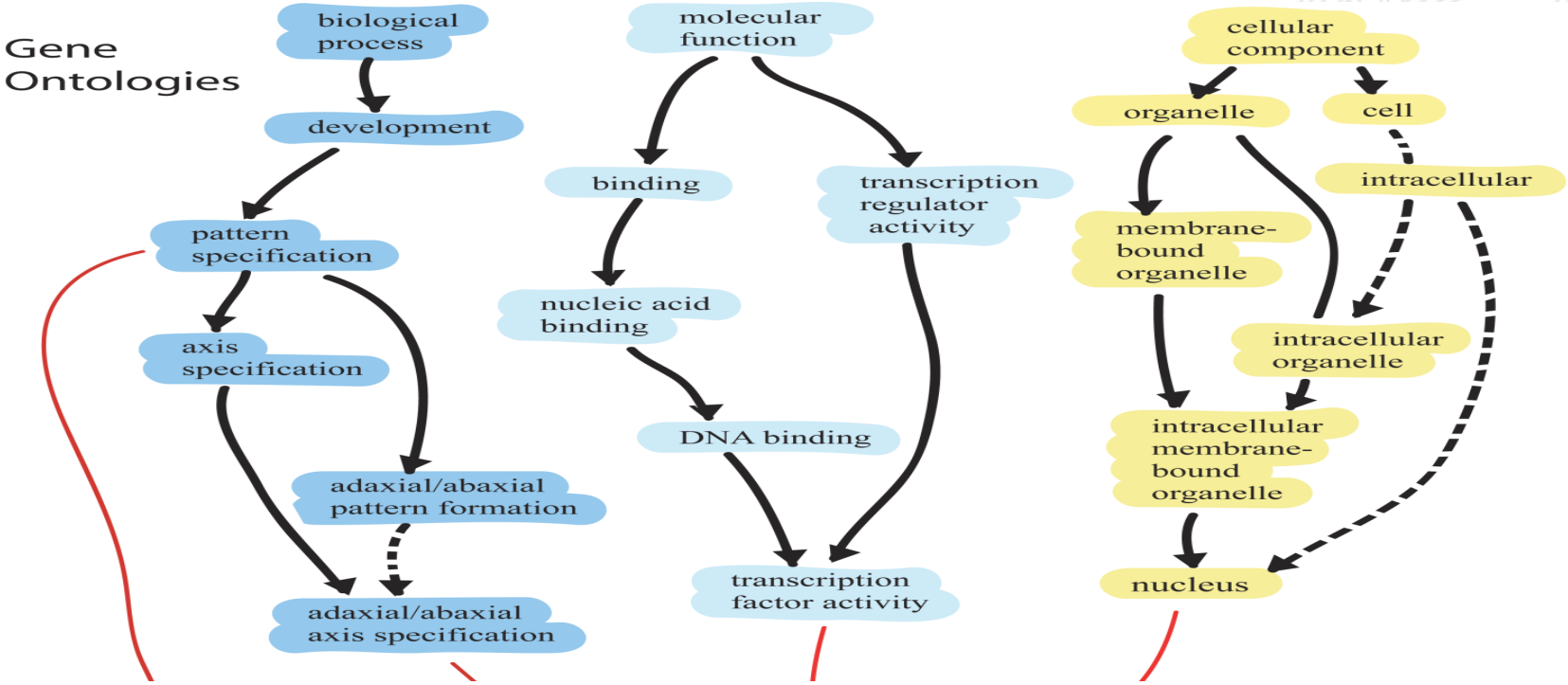
Desirable features identified for Web Ontology Language:

- Extends existing Web standards
  - Such as XML, RDF, RDFS
- Easy to understand and use
  - Should be based on familiar Knowledge Representation idioms
- Formally specified
- Of “adequate” expressive power
- **Possible to provide automated reasoning support**

Taxonomy (IS A, subClassOf)

UNIVERSIDAD DE MURCIA

- Hierarchical relation to define types of classes, from the most general to the most specific ones. These rules apply:
  - Rule “is-a”: Every subclass is a instance of the superclass
  - It is transitive (C is-a B and B is-a A then C is-a A)



OWL Class Constructors

UNIVERSIDAD DE MURCIA

- Build new classes from

Class, property and individuals

- Union

- Intersection

OWL Class Constructors

UNIVERSIDAD DE MURCIA

- Build new classes from

Class, property and individuals

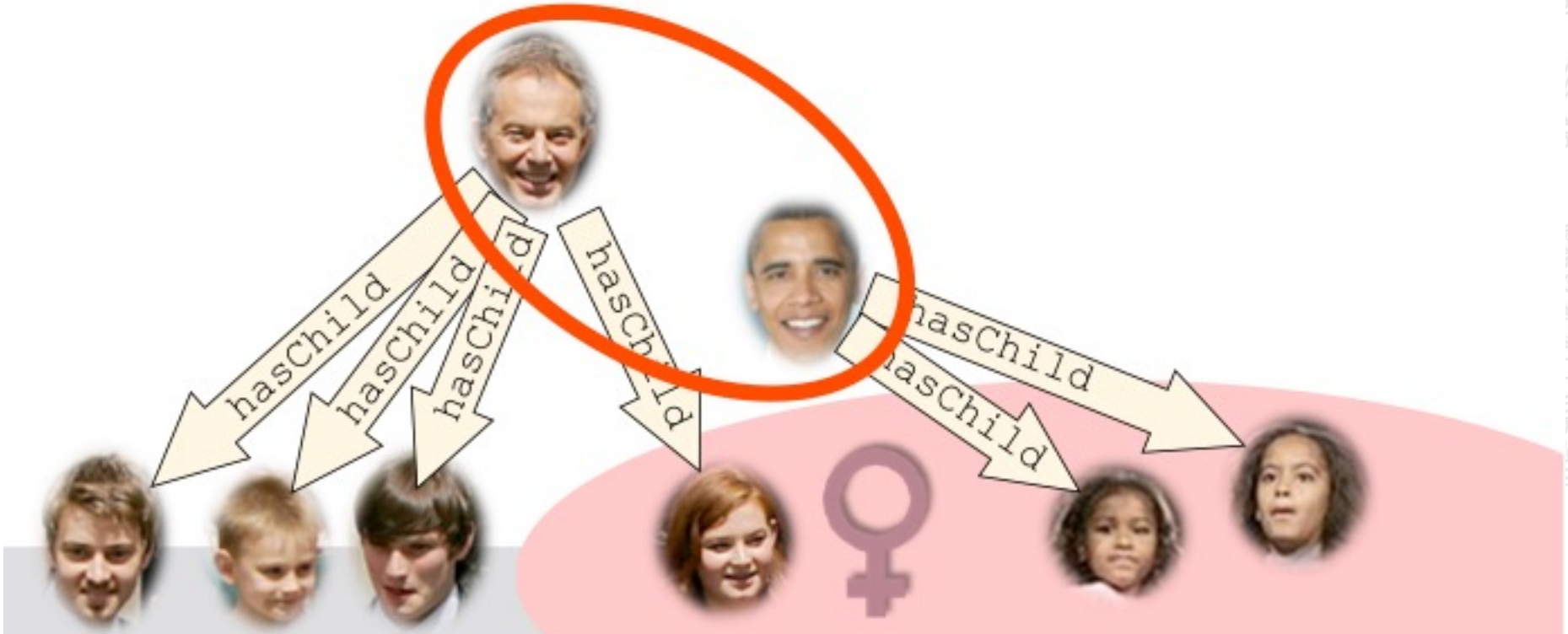
- Complement: not politician

- Closed classes

OWL Class Constructors

- build new classes from class, property and individual names
  - existential quantification: ∃hasChild.Female

```
<owl:Restriction>
    <owl:onProperty rdf:resource="hasChild"/>
    <owl:someValuesFrom rdf:resource="Female"/>
</owl:Restriction>
```

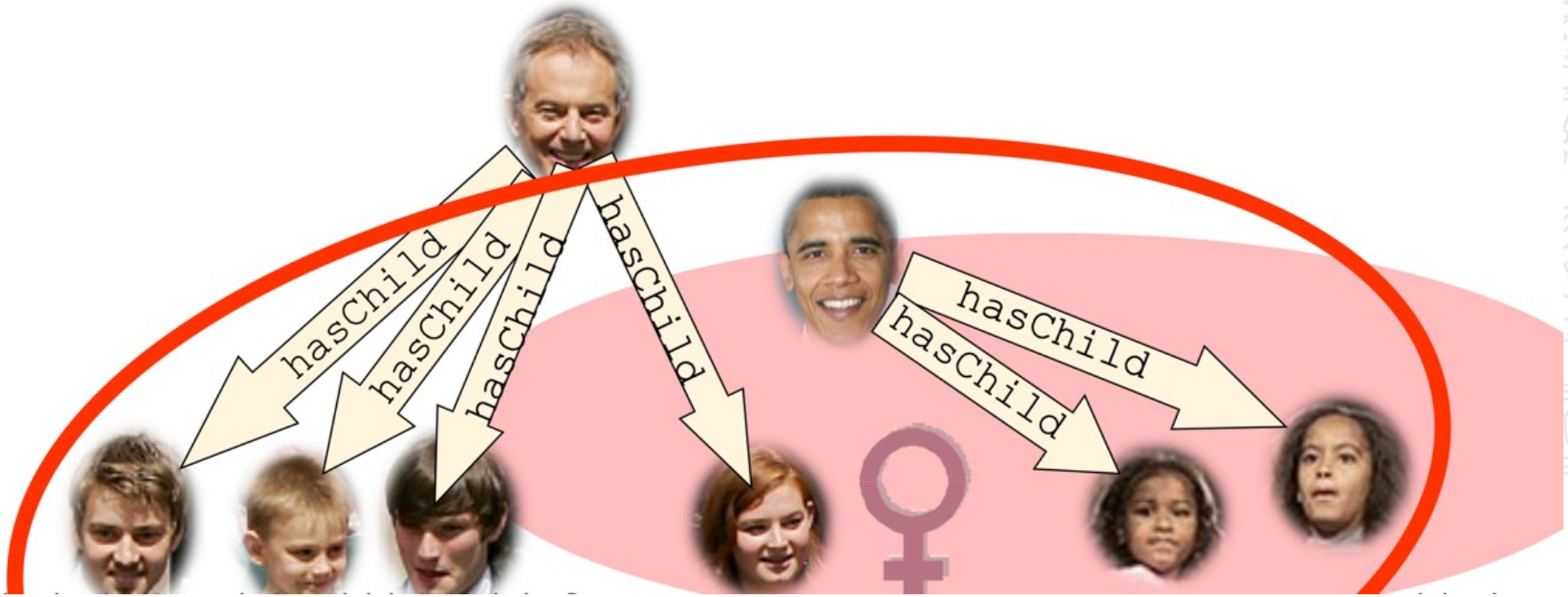


OWL Class Constructors

UNIVERSIDAD DE MURCIA

- build new classes from class, property and individual names
  - universal quantification: ∀hasChild.Female

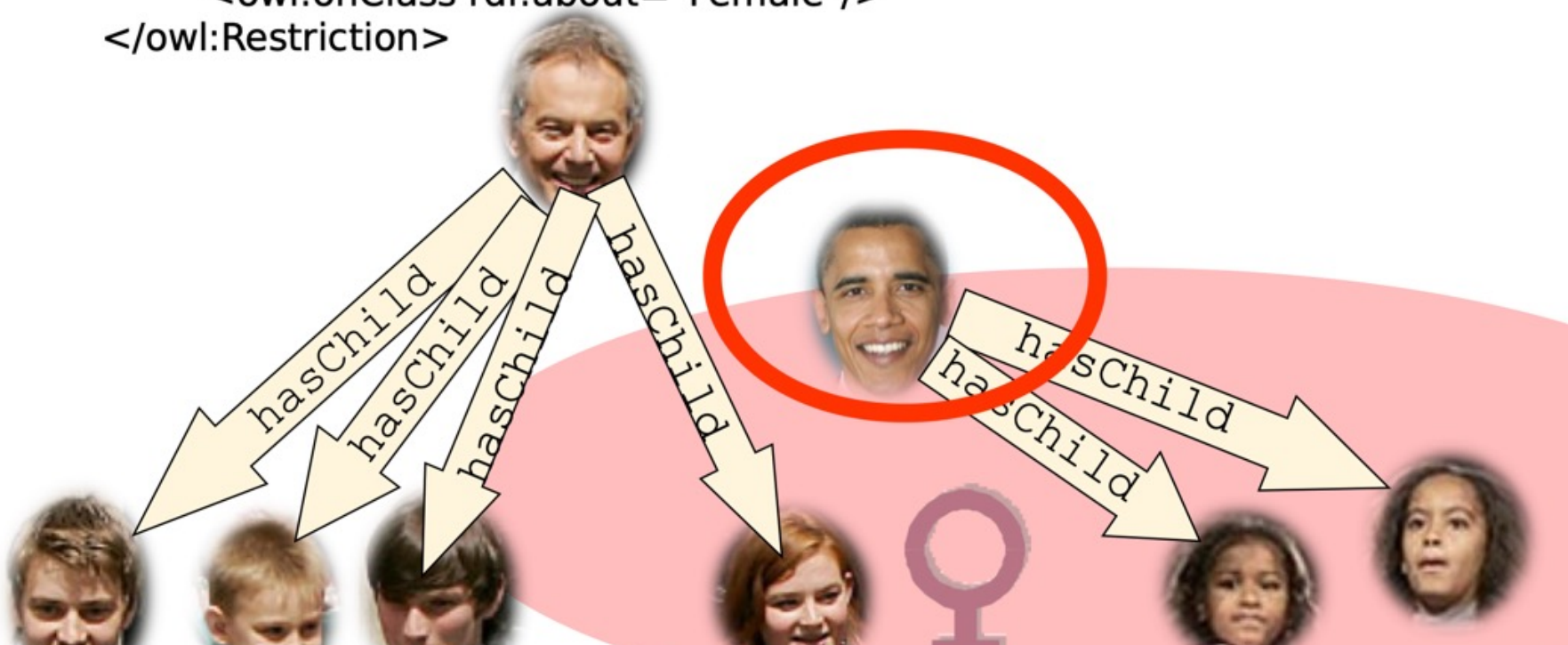
```
<owl:Restriction>
    <owl:onProperty rdf:resource="hasChild"/>
    <owl:allValuesFrom rdf:resource="Female"/>
</owl:Restriction>
```



OWL Class Constructors

- build new classes from class, property and individual names
  - cardinality restriction: ≥2hasChild.Female

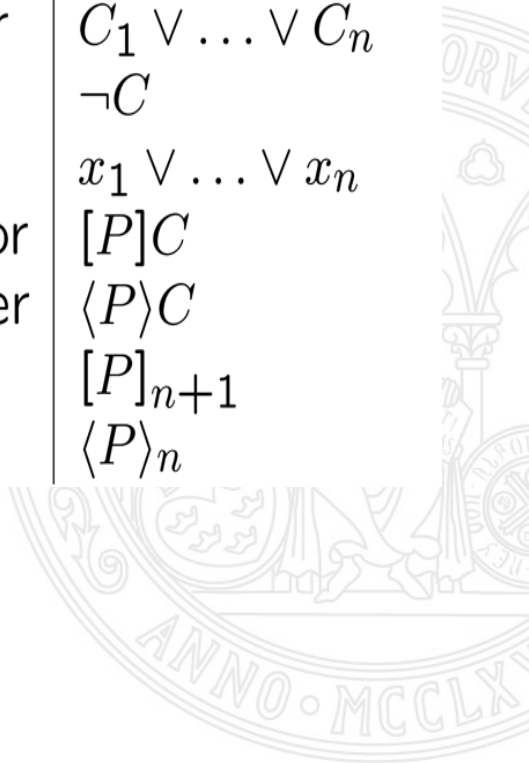
```
<owl:Restriction>
      <owl:minQualifiedCardinality rdf:datatype="&xsd;nonNegativeInteger“>
       2 </owl:minQualifiedCardinality>
      <owl:onProperty rdf:about="hasChild"/>
      <owl:onClass rdf:about=“Female"/>
</owl:Restriction>
```



OWL Class Constructors

UNIVERSIDAD DE MURCIA

Constructor	DL Syntax	Example	Modal Syntax
intersectionOf	$C_1 \sqcap \ldots \sqcap C_n$	Human $\sqcap$ Male	$C_1 \wedge \ldots \wedge C_n$
unionOf	$C_1 \sqcup \ldots \sqcup C_n$	Doctor $\sqcup$ Lawyer	$C_1 \vee \ldots \vee C_n$
complementOf	$\neg C$	$\neg$Male	$\neg C$
oneOf	$\{x_1\} \sqcup \ldots \sqcup \{x_n\}$	{john} $\sqcup$ {mary}	$x_1 \vee \ldots \vee x_n$
allValuesFrom	$\forall P.C$	$\forall$hasChild.Doctor	$[P]C$
someValuesFrom	$\exists P.C$	$\exists$hasChild.Lawyer	$\langle P \rangle C$
maxCardinality	$\leqslant nP$	$\leqslant$1hasChild	$[P]_{n+1}$
minCardinality	$\geqslant nP$	$\geqslant$2hasChild	$\langle P \rangle_n$



Special classes and properties

UNIVERSIDAD DE MURCIA

- Top class: class that contains all the individuals
  - owl:Thing
- Bottom class: empty class, no individuals
  - owl:Nothing
- Universal property: it links every individual to every individual
  - owl:topObjectProperty

Manchester OWL Syntax

UNIVERSIDAD DE MURCIA

- Closer to DL and closed to natural language

OWL	DL Symbol	Manchester OWL Syntax Keyword	Example
someValuesFrom	∃	**some**	hasChild **some** Man
allValuesFrom	∀	**only**	hasSibling **only** Woman
hasValue	∋	**value**	hasCountryOfOrigin **value** England
minCardinality	≥	**min**	hasChild **min** 3
cardinality	=	**exactly**	hasChild **exactly** 3
maxCardinality	≤	**max**	hasChild **max** 3

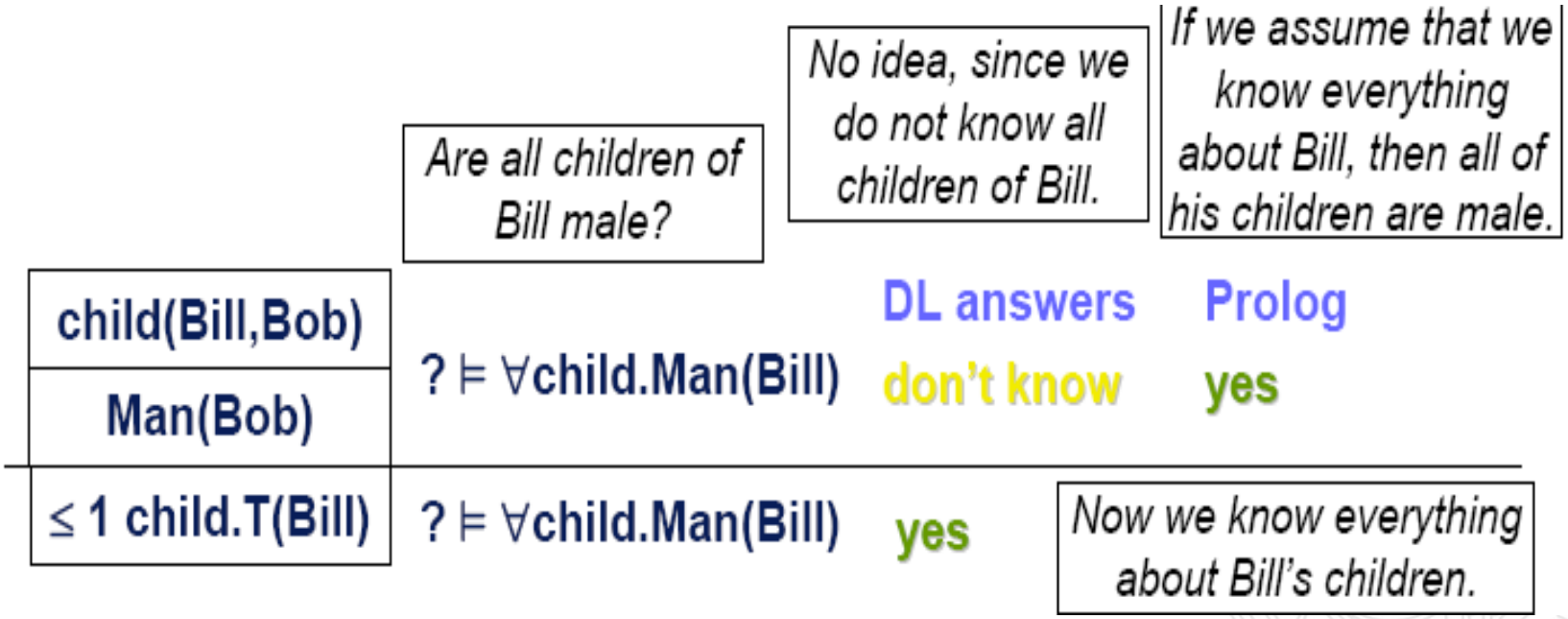
OWL	DL Symbol	Manchester OWL Syntax Keyword	Example
intersectionOf	⊓	**and**	Doctor **and** Female
unionOf	⊔	**or**	Man **or** Woman
complementOf	¬	**not**	**not** Child

Open Vs Closed World Assumption

UNIVERSIDAD DE MURCIA

- Open World Assumption (OWA)
  - The knowledge base may not contain all known individuals and facts
  - The existence of further individuals is possible if not explicitly excluded
- Closed World Assumption (CWA)
  - The knowledge base is assumed to contain all known individuals and facts
- Which one do you think is used in OWL?

Open Vs Closed World Assumption



Questions, comments…

UNIVERSIDAD DE MURCIA