

HOW TO USE THE SLIDE TEMPLATES

- To use your institution's slide design and logo, adjust the slides of this presentation using the „slide master“

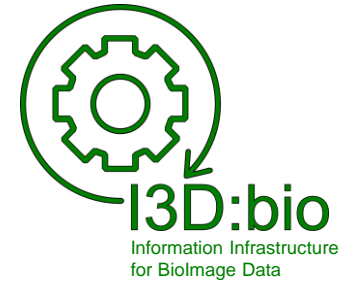
Note that these slides are optimized for 16:9 screen presentation layout

- Check the slides for **yellow-marked text** and insert the information according to your own institute's infrastructure.
- Feel free to use this material for videos, teaching, guidelines, etc., at your institute
- Please cite us (e.g., on page 1) when re-using this material or derivatives of it:

Adapted from: Schmidt C., Bortolomeazzi M., Boissonnet T., Fortmann-Grote C. *et al.* (2023). I3D:bio's OMERO training material: Re-usable, adjustable, multi-purpose slides for local user training. Zenodo. DOI: 10.5281/zenodo.8323588. If not stated otherwise, the content of this material (except for logos and the slide design) is published under [Creative Commons Attribution 4.0 license](#).

- If not stated otherwise, the content of this material (except for logos) is published under a [Creative Commons Attribution 4.0 license](#).
- This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 462231789 (Information Infrastructure for BioImage Data, I3D:bio)

Disclaimer



<https://www.i3dbio.de>

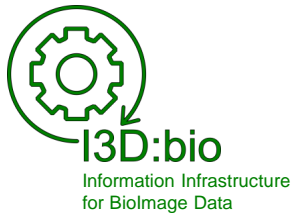
- The following slides are intended for reuse after **substituting yellow-marked text** with the relevant information at your institute.
- Some content may not apply to the specific setup of the OMERO installation at your institute.

The content reflects solely the authors' opinions and does not speak on behalf of the original software, its developers, or other cited community resources.

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project I3D:bio, grant number 462231789

Research Data Management for Bioimage Data at the **ADD INSTITUTE HERE**

Metadata Curation: What are ontologies? Why and how to use them?



ADD AUTHOR / RESPONSIBLE PERSON FROM YOUR INSTITUTE

Adapted from: Schmidt C., Bortolomeazzi M., Boissonnet T., Fortmann-Grote C. *et al.* (2023). I3D:bio's OMERO training material: Re-usable, adjustable, multi-purpose slides for local user training. Zenodo. DOI: 10.5281/zenodo.8323588
If not stated otherwise, the content of this material (except for logos and the slide design) is published under a [Creative Commons Attribution 4.0 license](https://creativecommons.org/licenses/by/4.0/).

**ADD LOGO
BIG**



Metadata details in form of Key-Value Pair annotation

Key-Value Pairs allow (standardized) annotation of detailed metadata

Consists of

- **Key:** Denotes a real-world object or an abstract concept that can be assigned a specific value (of different possible values)
- **Value:** Number or text string that specifies the object denoted under „Key“

Examples:

Key: „cell type“ **Value:** „CD4+ T cell“

Key: „disease model“ **Value:** „Experimental Autoimmune Encephalomyelitis“

Standardize Key-Value pairs?

Key: „cell type“

Value: „CD4+ T cell“

Key: „disease model“

Value: „Experimental Autoimmune Encephalomyelitis“



„cell type“ „type of cell“ „cell-type“ „cellular entity“ „cellular identity“



„CD4+ T cell“ „CD4-positive T-lymphocyte“ „naive, CD4-positive T cell“
„CD4-positive, alpha-beta T cell“ „Th0 cell“ „CD4+ T helper cell“



„Experimental Autoimmune Encephalomyelitis“ „EAE“ „Allergic Encephalomyelitis“

How to avoid ambiguity?

How to describe the data objectively?

How to make the metadata machine-interpretable?

Controlled vocabularies

A **controlled vocabulary** provides a list of terms.

- a definition of each term
- a unique identifier of each term
- different types exist, e.g.,
 - Alphabetical list
 - Thesaurus (a collection of synonyms)
 - Taxonomy (hierarchical or network-like list of terms)
 - (ontology)

→ **Allows standardized usage of terms**

Controlled vocabularies – example: MeSH

Medical Subject Headings (MeSH)

Controlled
vocabulary in the
form of a thesaurus

curated by the National
Library of Medicine (US)

NIH National Library of Medicine
National Center for Biotechnology Information

MeSH MeSH Search Limits Advanced

Full Send to: PubMed Search Builder

CD4-Positive T-Lymphocytes

A critical subpopulation of T-lymphocytes involved in the induction of most immunological functions. The HIV virus has selective tropism for the T4 cell which expresses the CD4 phenotypic marker, a receptor for HIV. In fact, the key element in the profound immunosuppression seen in HIV infection is the depletion of this subset of T-lymphocytes.

Tree Number(s): A11.118.637.555.567.569.200, A15.145.229.637.555.567.569.200, A15.382.490.555.567.569.200
MeSH Unique ID: D015496
Entry Terms:

- CD4 Positive T Lymphocytes
- CD4-Positive T-Lymphocyte
- T-Lymphocyte, CD4-Positive
- T-Lymphocytes, CD4-Positive
- CD4-Positive Lymphocytes
- CD4-Positive Lymphocyte
- Lymphocyte, CD4-Positive
- Lymphocytes, CD4-Positive
- T4 Cells
- T4 Cell
- T4 Lymphocytes
- T4 Lymphocyte

Previous Indexing:

- [T Lymphocytes \(1986-1988\)](#)

MeSH

Lymphocytes MeSH

"Lymphocytes"[MeSH Terms] (1) MeSH

house mouse (1) Taxonomy

rat (2) Taxonomy

See more...

Use of controlled vocabularies in practice

Key: „cell type“

Value: „CD4+ T cell“



„CD4+ T cell“

„**CD4-positive T-lymphocyte**“

„Th0 cell“

„naive, CD4-positive T cell“

Example of controlled vocabulary usage in Key-Value Pairs:

Key: cell type

Value: CD4-positive T-lymphocyte

Key: cell type term accession number

Value: <http://id.nlm.nih.gov/mesh/D015496>

- The term is defined
- Some hierarchical information is contained in MeSH see →
- Attributes / properties / relationships are missing

[All MeSH Categories](#)
[Anatomy Category](#)
[Cells](#)
[Blood Cells](#)
[Leukocytes](#)
[Leukocytes_Mononuclear](#)
[Lymphocytes](#)
[T-Lymphocytes](#)
CD4-Positive T-Lymphocytes
[T-Lymphocytes_Helper-Inducer](#)
[T_Follicular_Helper_Cells](#)
[Th1_Cells](#)
[Th17_Cells](#)
[Th2_Cells](#)
[T-Lymphocytes_Regulatory](#)

Ontologies

An **ontology** is a conceptual framework of how specific terms are used to represent *domain knowledge* in a (research) domain.

- Defines term attributes/properties, and relationships between the terms
- Terms with shared attributes are grouped into classes
- Terms in different ontologies are mapped to each other or adopted
- Can be extended over time with the evolving domain knowledge (i.e., an ontology is versioned)
- Formalized, i.e., ontologies can be expressed in ontology formats (machine-interpretable)

Examples of Ontologies:

- Experimental Factor Ontology (EFO) – curated by the EMBL EBI
- Biological Imaging Methods Ontology (FBbi) – curated by the Cell Image Library
- Cell Line Ontology (CLO) – community-based, curated at the University of Michigan

Ontologies consist of classes with attributes

Class

Represents a real-world object (e.g., „microscope objective lense“) or an abstract concept (e.g., „disease model“)

A class comprises subclasses or individual terms (instances) sharing attributes. Classes have specific relationships with each other.

Attribute

Specific property of a class (can be in form of Key-Value Pairs), e.g.:
Key: Definition Value: „This is the term definition (and a reference to a paper that first described it).“

Relationship

Relationship between classes

Note: The Key-Value Pairs in OMERO are not the same as the Key-Value Pairs for ontology class attributes. Both use the same concept independently.

Use of ontologies in practice

Key: „cell type“

Value: „CD4+ T cell“



„CD4+ T cell“

„CD4-positive T-lymphocyte“

„naive, CD4-positive T cell“

„CD4-positive, alpha-beta T cell“

„Th0 cell“

„CD4+ T helper cell“

Example of ontology usage in Key-Value Pairs:

Key: cell type

Value: CD4-positive, alpha-beta T cell

Key: cell type term accession number

Value: http://purl.obolibrary.org/obo/CL_0000624

Several ontologies can use the *same term*, e.g.:

- Experimental Factor Ontology (EFO)
- Cell Ontology (CL)
- Uber Anatomy Ontology (UBERON)
- others

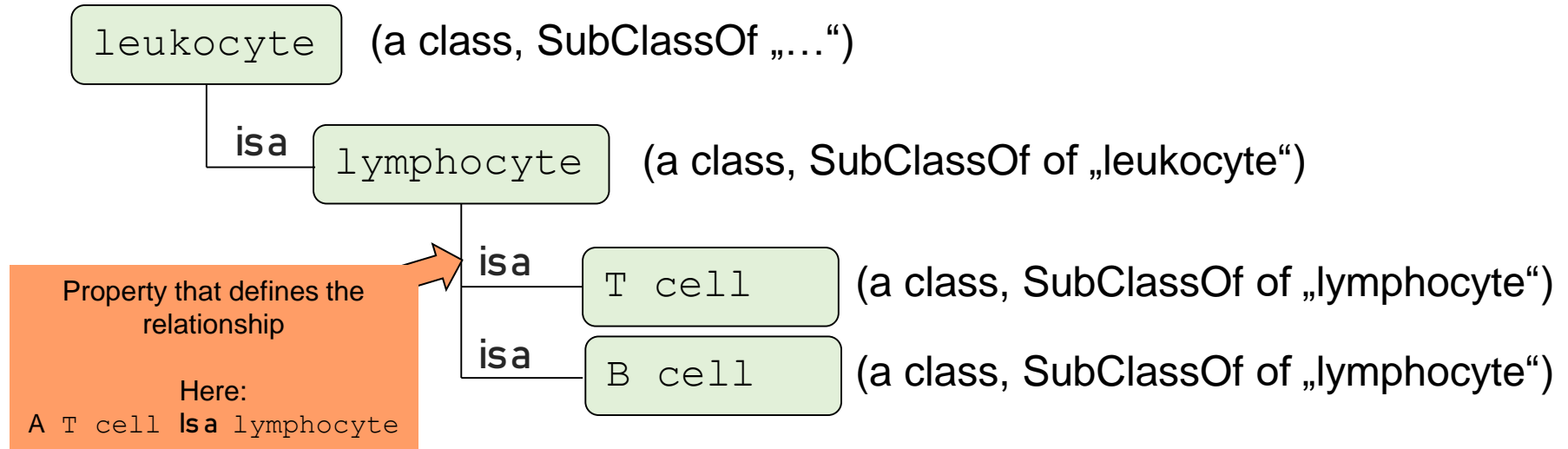
**Why are there many
different ontologies?**

Why are there so many ontologies?

Different ontologies are designed to optimally **represent their respective domain knowledge** (for example, the relationship between terms)

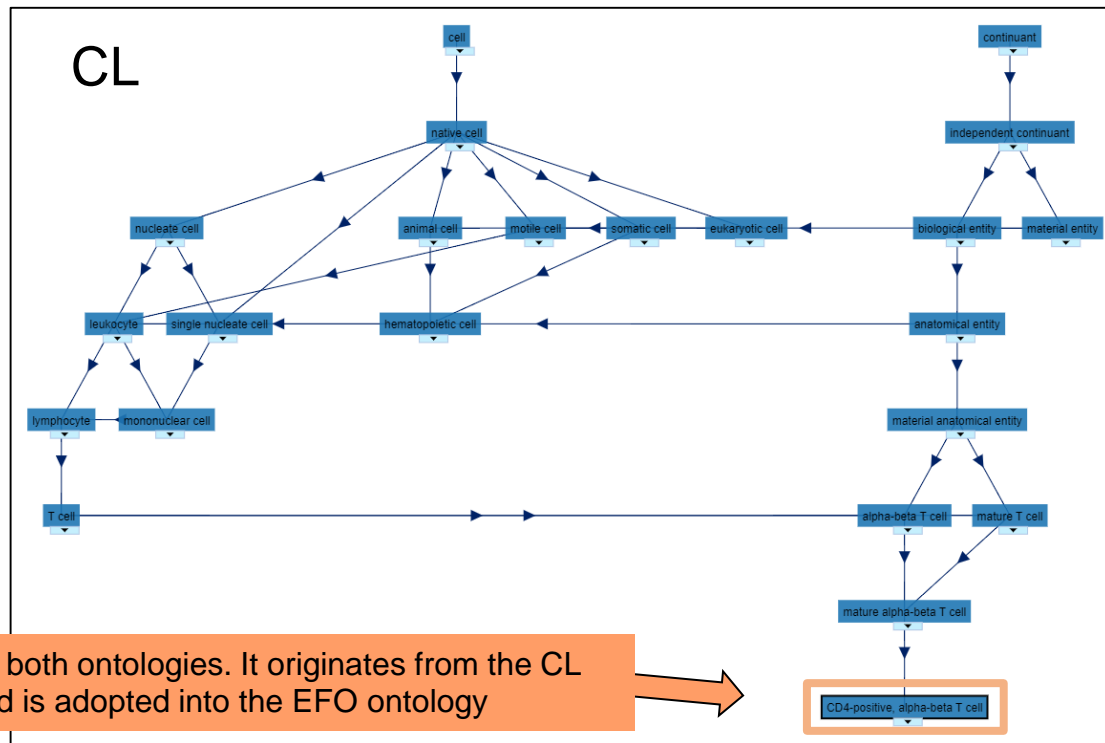
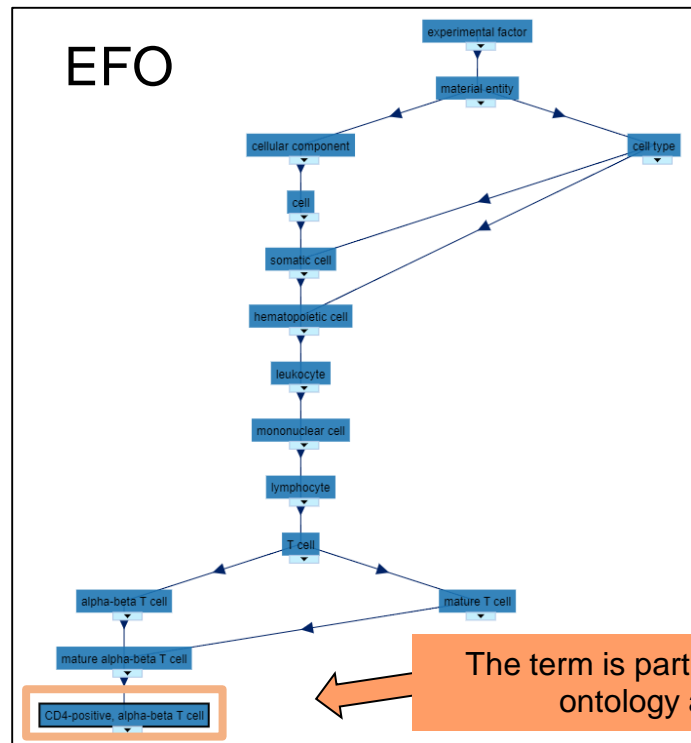
This knowledge can be represented as a tree structure or „knowledge graph“.

Example:



Graph visualizations of different ontologies

Term: CD4-positive, alpha-beta T cell; http://purl.obolibrary.org/obo/CL_0000624

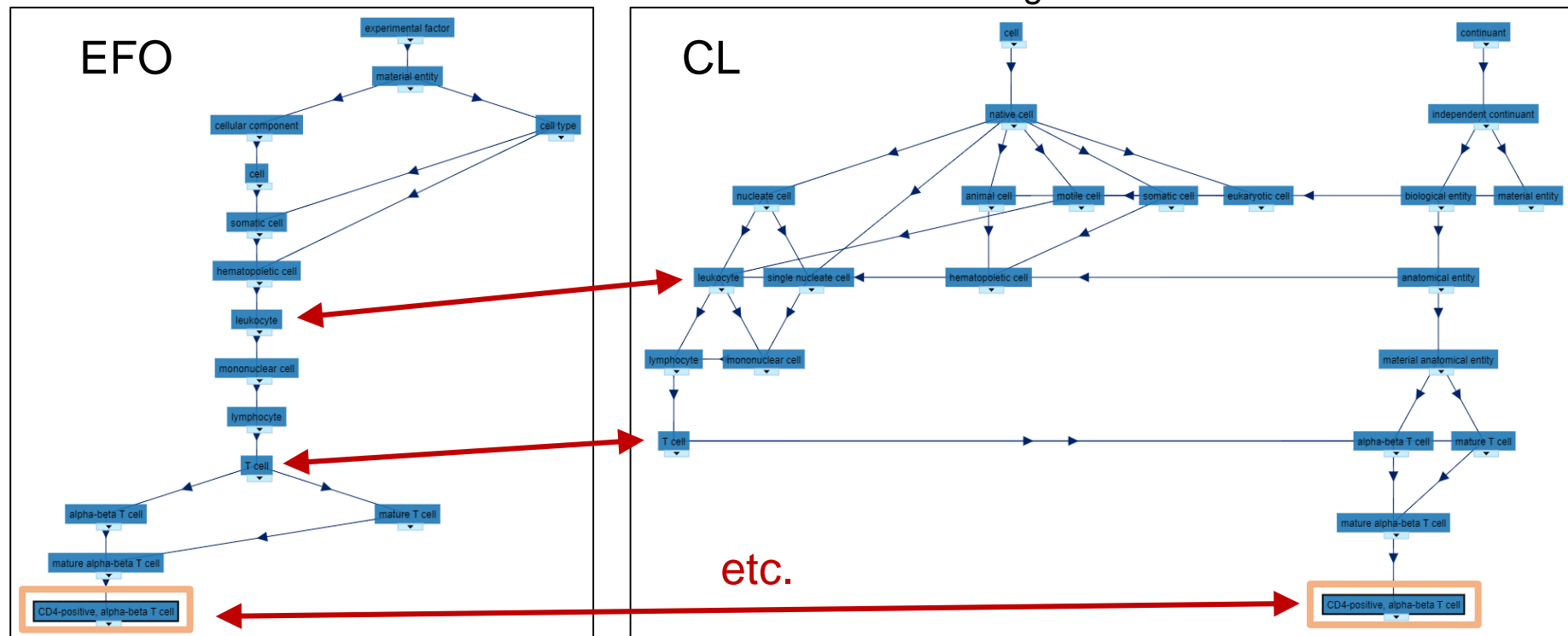


The term is part of both ontologies. It originates from the CL ontology and is adopted into the EFO ontology

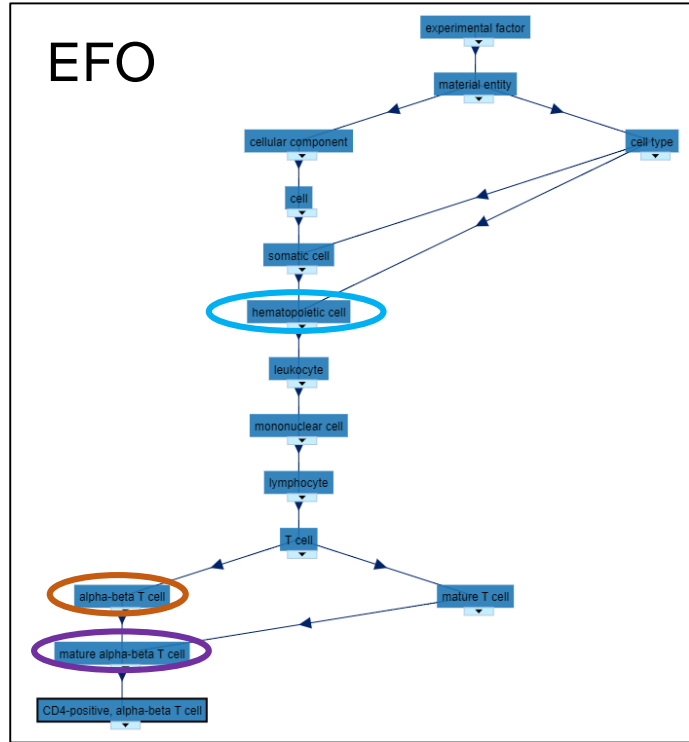
Mapping between ontologies

Terms are adopted from other ontologies, or synonyms in different ontologies are mapped to each other.

→ Semantic knowledge *across* domains!



The advantage of using ontologies



A single Key-Value Pair can carry extended domain knowledge!

„CD4-positive, alpha-beta T cell“ following an ontology (here: EFO) includes more information from the domain knowledge formalized in the ontology (and cross-domain knowledge formalized by mapping):

- Is carrying a T cell receptor with $\alpha\beta$ -chains
- Has completed thymic selection (i.e., is mature)
- Is a cell of the hematopoietic system
- etc...

Due to the ontology format, a computer can read the knowledge!

Using ontologies in OMERO 1/2

There is no unified standard for the use of ontologies in OMERO.

But we can start working with some recommendations.

Suggestion (based on REMBI¹ items, and ISA-TAB²):

To create machine-actionable metadata, make use of **ontology terms** and **ontology term source references**:

- Use the ontology-derived term as the Value for a specific Key
- Add the ontology term URL as the Value for a second Key using the **<Key> + „Term Accession Number“**

Example

Key: Biological entity

Value: CD4-positive, alpha-beta T cell

Key: Biological entity Term Accession Number

Value: http://purl.obolibrary.org/obo/CL_0000624

Using ontologies in OMERO 2/2

When and why to include the ontology source reference?

Ontologies allow for *cross-domain* referencing. I.e., a specific term in one ontology may be adopted from another ontology.

How do you know? Example:

A term was chosen from EFO ontology but the term ID implies CL ontology:

http://purl.obolibrary.org/obo/CL_0000624



Term ID points to CL (not EFO)

Solution? Include the ontology source URL:

Example

Key: Biological entity

Value: CD4-positive, alpha-beta T cell

Key: Biological entity Term Accession Number

Value: http://purl.obolibrary.org/obo/CL_0000624

Key: Biological entity Term Source REF

Value: <http://www.ebi.ac.uk/efo/efo.owl>

Benefits of using ontologies for data annotation


Pre-publication benefits:


- Avoid term ambiguity in collaborative research settings
- Use ontology identifiers for automated image analysis workflows
- Enable semantic search in your own dataset(s)


Post-publication benefits:

- Your publication & data are more likely to be found by others, which can increase citations and facilitate collaboration
- Your data is retrievable for semantic search across domains and can thus generate a higher scientific impact

Getting started with ontologies – Open Biological and Biomedical Ontologies (OBO) Academy

 OBO Semantic Engineering Training

 Search

 GitHub
☆ 59 🗨 32

OBO Semantic Engineering Training

About the course

Getting started with learning

Overview of lessons and tutorials

Contribute to the course

Courses >

Pathways >

Lessons >

Lessons in Development >

Tutorials >

How-to guides >

Reference >

Explanations >

Introduction to Ontologies

Logical axiomatization of classes & use of reasoning

Term Comments

Introduction to Annotation Properties

Writing good issues

Ontology Matching

Which biomedical ontologies should we use?

OWL, OBO, JSON? Base, simple, full, basic? What should you use, and why?

What are taxon constraints?

The logical building blocks of OWL

SubClassOf versus EquivalentTo

Existential restrictions

Introduction to ontologies

Based on [CL editors training](#) by David Osumi-Sutherland

Why do we need ontologies?

We face an ever-increasing deluge of biological data analysis. Ensuring that this data and analysis are Findable, Accessible, Interoperable, and Re-usable ([FAIR](#)) is a major challenge. Findability, Interoperability, and Reusability can all be enhanced by standardising metadata. Well-standardised metadata can make it easy to *find* data and analyses despite variations in terminology ('Clara cell' vs 'nonciliated bronchiolar secretory cell' vs 'club cell') and precision ('bronchial epithelial cell' vs 'club cell'). Understanding which entities are referred to in metadata and how they relate to the annotated material can help users work out if the data or analysis they have found is of interest to them and can aid in its re-use and interoperability with other data and analyses. For example, does an annotation of sample data with a term for breast cancer refer to the health status of the patient from which the sample was derived or that the sample itself comes from a breast cancer tumor?

We can't find what we're looking for

Given variation in terminology and precision, annotation with free text alone is not sufficient for findability. One very lightweight solution to this problem is to rely on user-generated keyword systems, combined with some method of allowing users to choose from previously used keywords. This can produce some degree of annotation alignment but also results in fragmented annotation and varying levels of precision with no clear way to relate annotations.

For example, trying to refer to feces, in NCBI BioSample:

Query	Records
Feces	22,592

Table of contents

Why do we need ontologies?

We can't find what we're looking for

We don't know what we're talking about

Controlled vocabulary (CV)

Key features

Example using wines

Hierarchical controlled vocabulary

Definition

Key features

Example using wines (Taxonomy of wine)

Support for grouping and varying levels of precision

From hierarchical CVs to ontologies

Synonyms

Polyhierarchy

Named relationships

What is an ontology?

Definition

Key features of well-structured ontologies:

Examples

Non-logical parts of ontologies

Identifiers

Using identifiers devoid of intrinsic meaning

IRIs? URIs? URLs?

Building scalable ontologies

Getting started with ontologies – FAIR Cookbook

← FAIRCOOKBOOK

GITHUB

FOREWORD

Introduction

Introducing the FAIR Principles

Reflecting on the ethical values of FAIR

Introducing our FAIRification framework

Prioritizing projects for FAIRification

Framing FAIR and the notion of metadata

Understanding the relation between FAIR and Knowledge Graphs

Training for FAIRification with open or synthetic biomedical datasets

Raising Awareness in Public Knowledge Graphs for Life Sciences

Reflecting on Practical Considerations for CROs to play FAIR

Data Protection Impact Assessment and Data Privacy Glossary

RECIPES AT A GLANCE

All Recipes In a Table

FAIR RECIPES

Findability

Accessibility

Interoperability

1. Registering SwissLipids identifiers in Wikidata

4. Introduction to terminologies and ontologies

Recipe Overview

🕒 Reading Time
15 minutes

▶ Executable Code
No

🔥 Difficulty
🔥🔥🔥

Introducing terminologies and ontologies

🖥️ Recipe Type
Survey / Review

👤 Audience
Data Curator, Data Manager, Data Scientist

📊 Maturity Level & Indicator
not applicable

Cite me with FCB019

4.1. Main objectives

The aim of this recipe is to provide a compact introduction about `controlled terminologies` and `ontologies`, why these resources are central to the preservation of knowledge and data mining and how such resources are developed.

4.2. Controlled terminology or ontology: what's the difference?

The need for `controlled vocabulary` often arises in situations where validation of textual information is necessary for operational requirements. The main initial driver for data entry harmonization is to increase query recall. In its most basic form, `keywords` may be used to perform indexation. However, if relying on user input alone, the chances of typographic errors increases with the number of users. These unavoidable events accumulate over time and end up hurting the accuracy of search results and this is the reason for offering sets of predefined values. It reduces the noise. However, this can come at the cost of precision, as the predefined terms may not cover the exact thing users may need to describe. Furthermore, term mis-selection by the user is not eliminated and introduces another type of error.

A `controlled terminology` is a `normative` collection of terms, the spelling of which is fixed and for which additional information may be provided such as a `definition`, a set of `synonyms`, an `editor`, a `version`, as well as a `license` determining the condition of use. The set of information about a specific controlled terminology term is designated as `term metadata`. In a controlled terminology, terms appear as a `flat list`, meaning that no relationship between any of the entities the controlled terminology represents is captured in any formal way. This is the main drawback and limitation of `controlled terminologies`, which are often developed to support a data model or an application.

<https://faircookbook.elixir-europe.org/content/recipes/interoperability/introduction-terminologies-ontologies.html>

CC BY

p 20

ADD LOGO SMALL

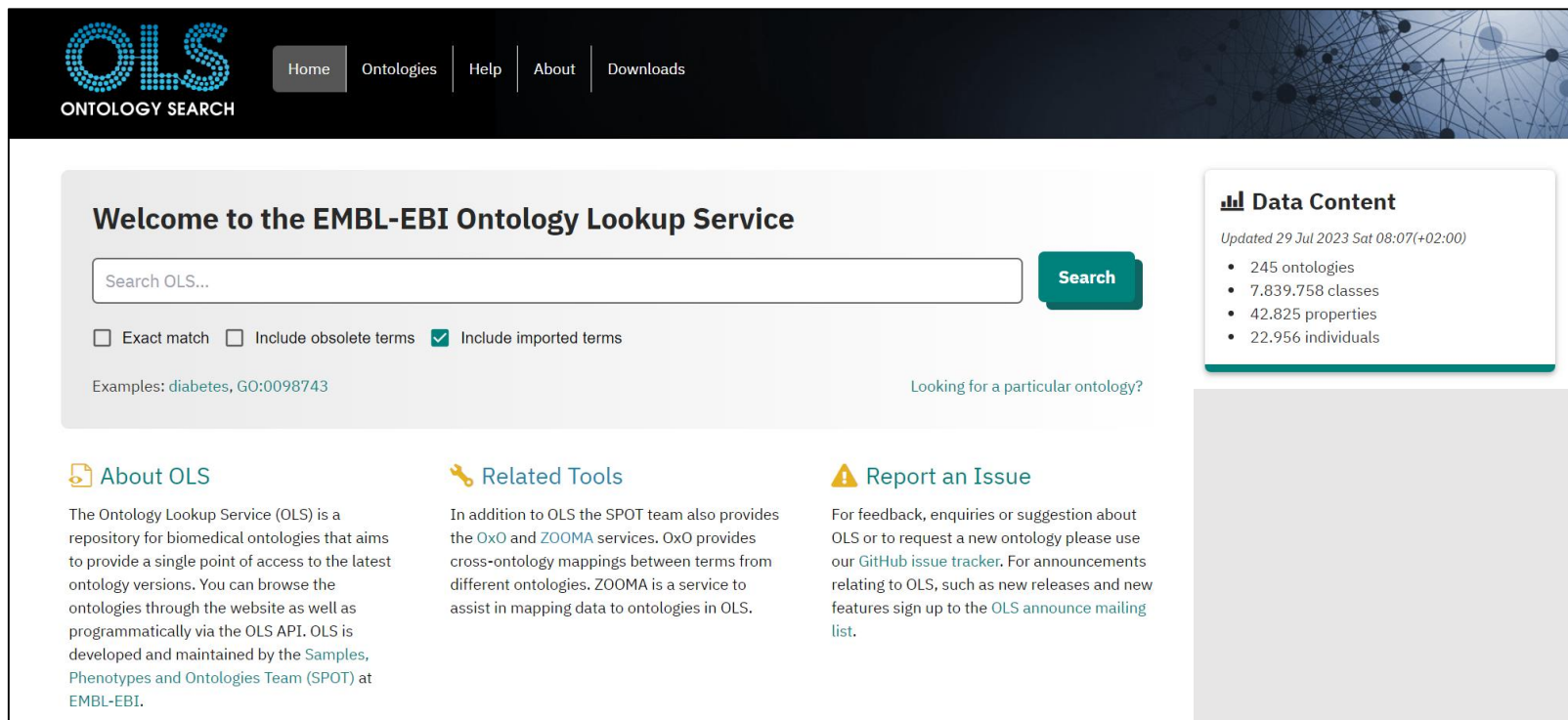
Getting started with ontologies – BioPortal BioOntology

- Check out:
- The Annotator
 - The Recommender

The screenshot displays the BioPortal homepage. At the top is a dark blue navigation bar with the BioPortal logo and links for Ontologies, Search, Annotator, Recommender, and Mappings. The main heading reads "Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies". Below this are two primary search sections: "Search for a class" and "Find an ontology". The "Search for a class" section includes a text input field with the placeholder "Enter a class, e.g. Melanoma", a blue search button, and a link to "Advanced Search". The "Find an ontology" section features a text input field with the placeholder "Start typing ontology name, then choose from list", a blue search button, and a "Browse Ontologies" button with a dropdown arrow. Below the search sections are two more panels. The "Ontology Visits (June 2023)" panel contains a horizontal bar chart showing visit counts for MEDDRA (~16,500), RXNORM (~14,000), SNOMEDCT (~7,000), LOINC (~5,000), and MESH (~4,000), with a "More" link below. The "BioPortal Statistics" panel is a table listing key metrics.

BioPortal Statistics	
Ontologies	1,062
Classes	15,915,705
Properties	36,286
Mappings	79,636,946

Getting started with ontologies – Ontology Lookup Service (by EMBL-EBI)



The screenshot shows the EMBL-EBI Ontology Lookup Service (OLS) homepage. At the top is a dark navigation bar with the OLS logo (a blue 'OLS' with a network diagram) and the text 'ONTOLOGY SEARCH'. To the right of the logo are links: Home, Ontologies, Help, About, and Downloads. The main content area has a light gray background. On the left, there's a 'Welcome to the EMBL-EBI Ontology Lookup Service' section with a search bar labeled 'Search OLS...' and a green 'Search' button. Below the search bar are three checkboxes: 'Exact match' (unchecked), 'Include obsolete terms' (unchecked), and 'Include imported terms' (checked). Below these are examples: 'diabetes, GO:0098743' and a link 'Looking for a particular ontology?'. To the right of the search bar is a 'Data Content' section with a bar chart icon, the title 'Data Content', and a timestamp 'Updated 29 Jul 2023 Sat 08:07(+02:00)'. Below this is a list of statistics: 245 ontologies, 7.839.758 classes, 42.825 properties, and 22.956 individuals. At the bottom of the main content area are three columns: 'About OLS' with a folder icon, 'Related Tools' with a wrench icon, and 'Report an Issue' with a warning icon. Each column contains a brief description of the service or tool.

Welcome to the EMBL-EBI Ontology Lookup Service

Search OLS...

☐ Exact match ☐ Include obsolete terms ☒ Include imported terms

Examples: diabetes, GO:0098743

Looking for a particular ontology?

Data Content

Updated 29 Jul 2023 Sat 08:07(+02:00)

- 245 ontologies
- 7.839.758 classes
- 42.825 properties
- 22.956 individuals

About OLS

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. You can browse the ontologies through the website as well as programmatically via the OLS API. OLS is developed and maintained by the [Samples, Phenotypes and Ontologies Team \(SPOT\)](#) at EMBL-EBI.

Related Tools

In addition to OLS the SPOT team also provides the [OxO](#) and [ZOOMA](#) services. OxO provides cross-ontology mappings between terms from different ontologies. ZOOMA is a service to assist in mapping data to ontologies in OLS.





Report an Issue

For feedback, enquiries or suggestion about OLS or to request a new ontology please use our [GitHub issue tracker](#). For announcements relating to OLS, such as new releases and new features sign up to the [OLS announce mailing list](#).

Check out:
- The Related Tools


<https://www.ebi.ac.uk/ols4>


Getting started with ontologies – ISA tools software suite


 [Software Suite](#) [Model and Formats](#) [Blog](#) [Support](#) [User Community and Publications](#)   


The open source **ISA framework** and tools help to manage an increasingly diverse set of life science, environmental and biomedical experiments that employing one or a combination of technologies.


Built around the 'Investigation' (the project context), 'Study' (a unit of research) and 'Assay' (analytical measurement) data model and serializations (tabular, JSON and RDF), the ISA framework helps you to provide rich description of the experimental metadata (i.e. sample characteristics, technology and measurement types, sample-to-data relationships) so that the resulting data and discoveries are reproducible and reusable.


**Collect and curate, following standards**
Describe the experimental steps using community-defined minimum reporting requirements and ontologies, where possible.

**Store and browse, locally or publicly**
Create your own repository to search and browse the experimental description and associated data, hosted openly or privately.

**Submit to public repositories**
When required, reformat experiments for submission to supported public repositories or directly export to those already using ISA formats.

**Analyse with existing tools**
Upload experimental descriptions and associated data to a growing number of well-known analysis systems that ISA formats connect with.

**Release, reason and nanopublish**
Explore and reason over your experiments, open them to the linked data universe, or publish nano-statements of your discoveries.

**Publish data alongside your article**
Directly export your experiments to a new generation of data journals that are accepting submissions in ISA formats.

Software tools (outside of OMERO) for metadata annotation

MDE.mic (OMERO.mde) for ontology-compliant annotation

Intermediate step during the data import to OMERO:

Review and Annotate metadata using **OMERO.mde**, a metadata editor.

It allows to edit:

- metadata of individual files,
- metadata the import queue in batch,
- and is supported by standardized, but configurable metadata fields and ontology term look-up

