# I3D:bio OMERO user training slides

**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 <mark>yellow-marked text</mark> 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:

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

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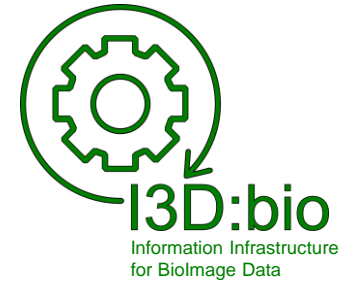
- The following slides are intended for reuse after ==substituting yellow-marked text== with the relevant information at your institute.
- Some content may not apply to the specific setup of the OMERO installation at your institute.

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

I3D:bio
Information Infrastructure
for BioImage Data

https://www.i3dbio.de

I3D:bio
Information Infrastructure
for BioImage Data

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

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

I3D:bio

Information Infrastructure
for BioImage Data
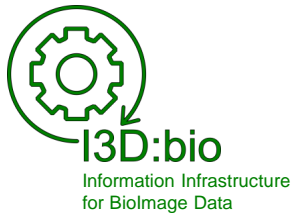
ADD AUTHOR / RESPONSIBLE PERSON FROM YOUR INSTITUE

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 Encephalomyelits"

## 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**

ADD LOGO SMALL

# Controlled vocabularies – *example: MeSH*

**Medical Subject Headings (MeSH)**

Controlled vocabulary in the form of a thesaurus

curated by the National Library of Medicine (US)

ADD LOGO SMALL

# 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

ADD LOGO SMALL

# 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

Browse different ontologies: https://bioportal.bioontology.org/

ADD LOGO SMALL

# 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.

ADD LOGO SMALL

# 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?**

ADD LOGO SMALL

# 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:*

| leukocyte | (a class, SubClassOf „…") |

isa

| lymphocyte | (a class, SubClassOf of „leukocyte") |

isa

| T cell | (a class, SubClassOf of „lymphocyte") |

isa

| B cell | (a class, SubClassOf of „lymphocyte") |

Property that defines the relationship

Here:
A `T cell` **Isa** `lymphocyte`

ADD LOGO SMALL
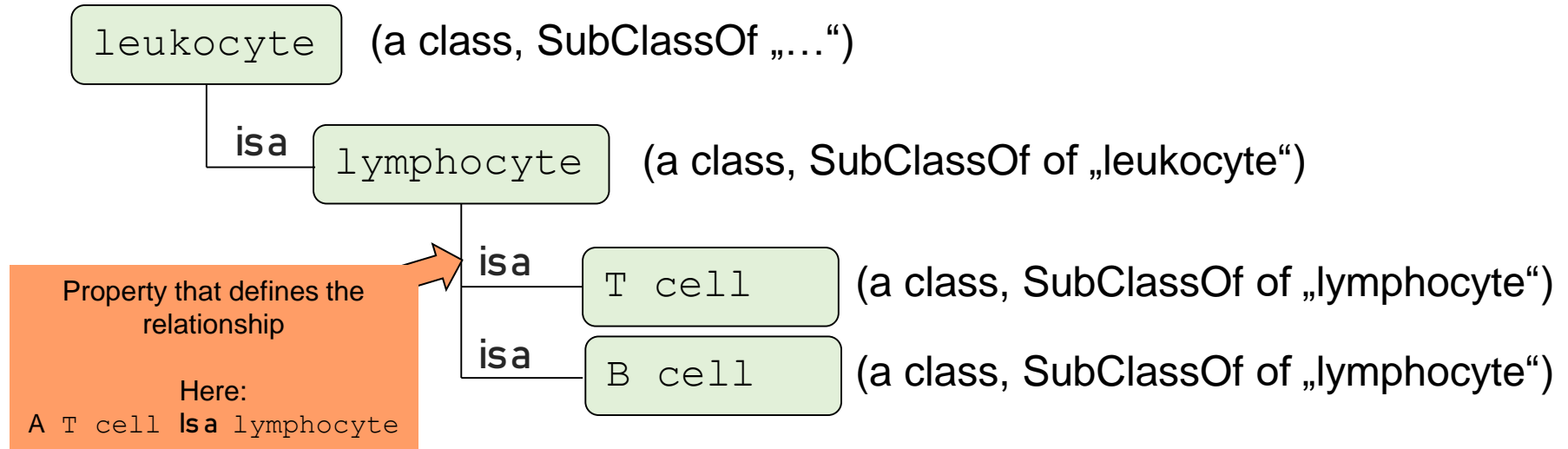
# Graph visualizations of different ontologies

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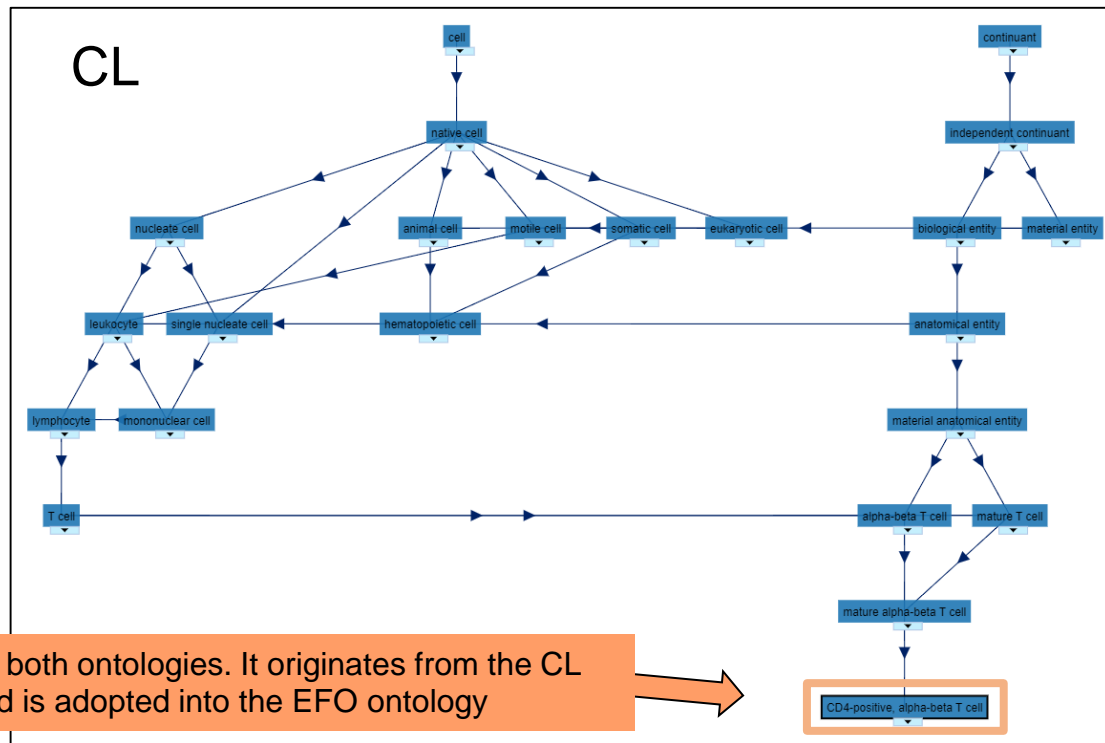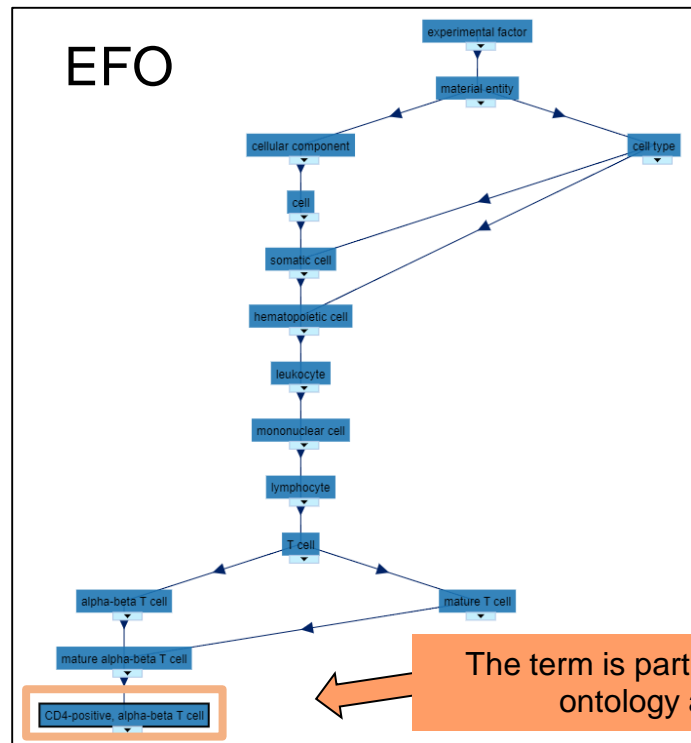


The term is part of both ontologies. It originates from the CL ontology and is adopted into the EFO ontology

ADD LOGO SMALL

# Mapping between ontologies

Terms are adopted from other ontologies, or synonyms in different ontologies are mapped to each other.
→ Semantic knowledge *across* domains!



EFO

CL

etc.

ADD LOGO SMALL
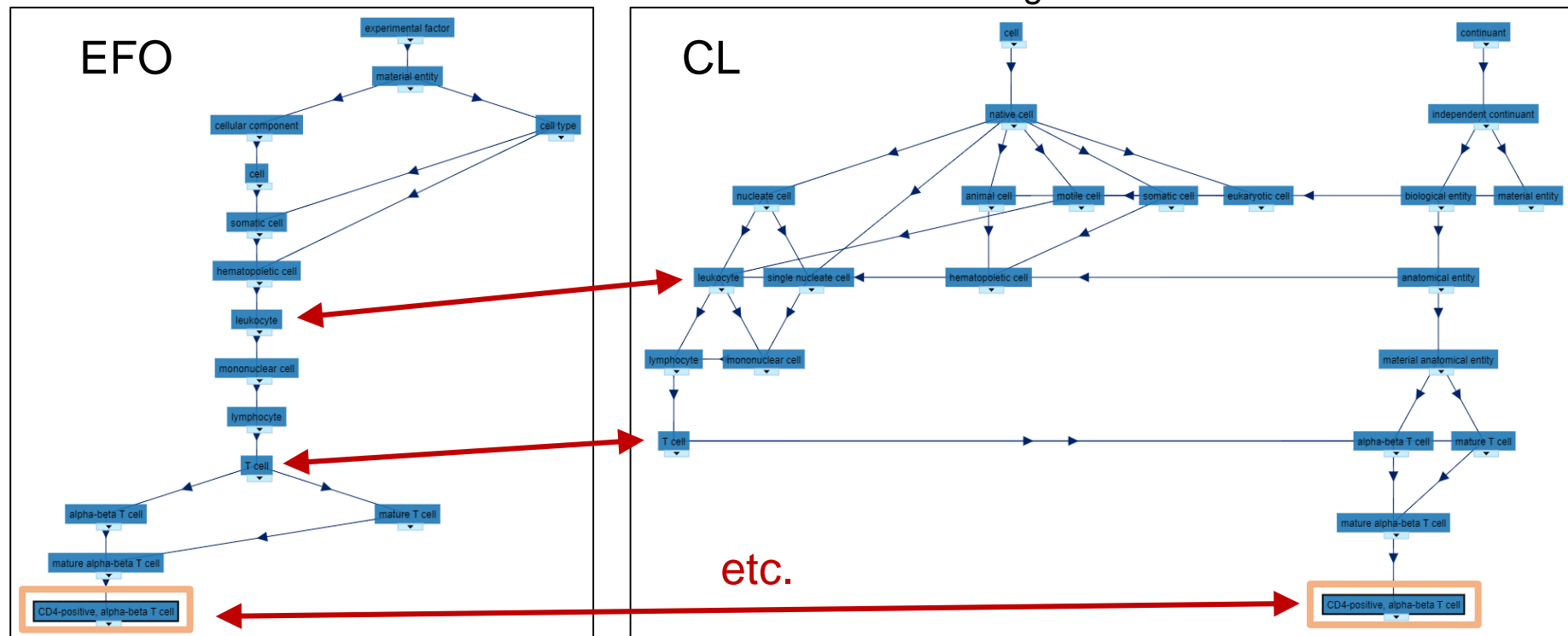
# The advantage of using ontologies



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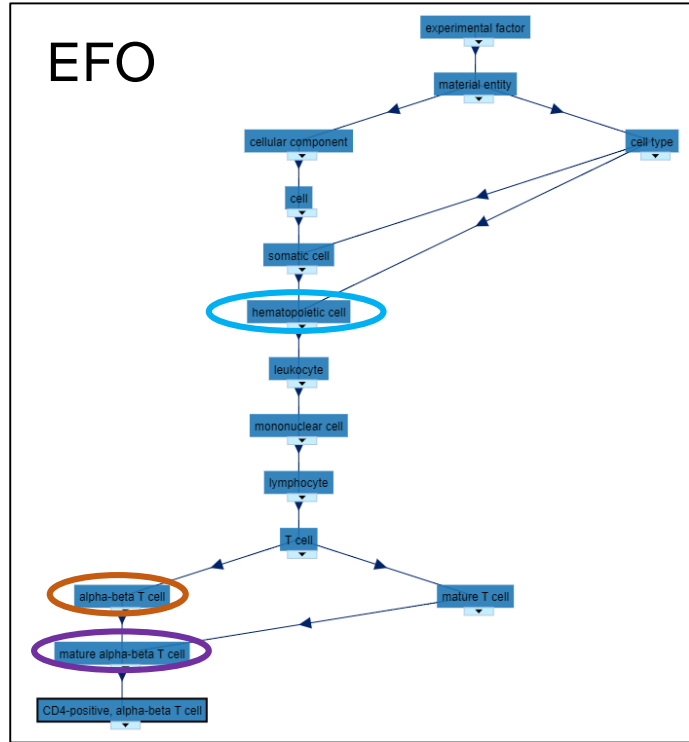
**A single Key-Value Pair can carry extended domain knowledge!**

„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 αβ-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!**

ADD LOGO SMALL

# 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 REMBI[1] items, and ISA-TAB[2]):*
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

ADD LOGO SMALL

# 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) Acadamy

ADD LOGO SMALL

# Getting started with ontologies – FAIR Cookbook



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

ADD LOGO SMALL

# Getting started with ontologies – BioPortal BioOntology

Check out:
- The Annotator
- The Recommender

ADD LOGO SMALL

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



Check out:
- The Related Tools

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

ADD LOGO SMALL

# Getting started with ontologies – ISA tools software suite



Software tools (outside of OMERO) for metadata annotation

https://isa-tools.org/ (last access: 2023-08-17)

ADD LOGO SMALL

# 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

ADD LOGO SMALL