



NFDI4
BIOIMAGE

Introduction to Ontologies and a suggestions on how to use in OMERO

Workshop: **FAIR data handling for microscopy: Structured metadata
annotation in OMERO**

April 29th & 30th, 2024, Day 1 – Session 5

Trainers: Tom Boissonnet, Vanessa Fuchs, **Christian Schmidt**



With the exception of logos and cited third-party content, the content of these slides is shared under the terms of the [Creative Commons Attribution License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/) unless the content is marked otherwise.

What is an image?

Microsoft Bing search for „image definition“

1. a representation of the external form of a person or thing in art:
"her work juxtaposed images from serious and popular art"
Ähnlich: likeness resemblance depiction portrayal representation
2. the general impression that a person, organization, or product presents to the public:
"she strives to project an image of youth"
Ähnlich: persona profile face identity front facade mask guise
3. a simile or metaphor:
"he uses the image of a hole to describe emotional emptiness"
Ähnlich: simile metaphor metonymy figure of speech trope

Asking a microscopist

https://en.wikipedia.org/wiki/Virtual_image

In *optics*, the *image* of an object is defined as the collection of *focus points* of *light rays* coming from the object. A *real image* is the collection of focus points made by *converging* rays, while a **virtual image** is the collection of focus points made by extensions of *diverging* rays.

Asking a research software engineer...

<https://docs.docker.com/guides/docker-concepts/the-basics/what-is-an-image/>

A container image is a standardized package that includes all of the files, binaries, libraries, and configurations to run a container.

Technical terms in science

Key: „cell type“

Value: „CD4+ T cell“

Key: „disease model“

Value: „Experimental Autoimmune Encephalomyelitis“

<input type="checkbox"/>	„cell type“	„type of cell“	„cell-type“	„cellular entity“	„cellular identity“
<input type="checkbox"/>	„CD4+ T cell“	„CD4-positive T-lymphocyte“	„naive, CD4-positive T cell“		
	„CD4-positive, alpha-beta T cell“	„Th0 cell“		„CD4+ T helper cell“	???
<input type="checkbox"/>	„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)

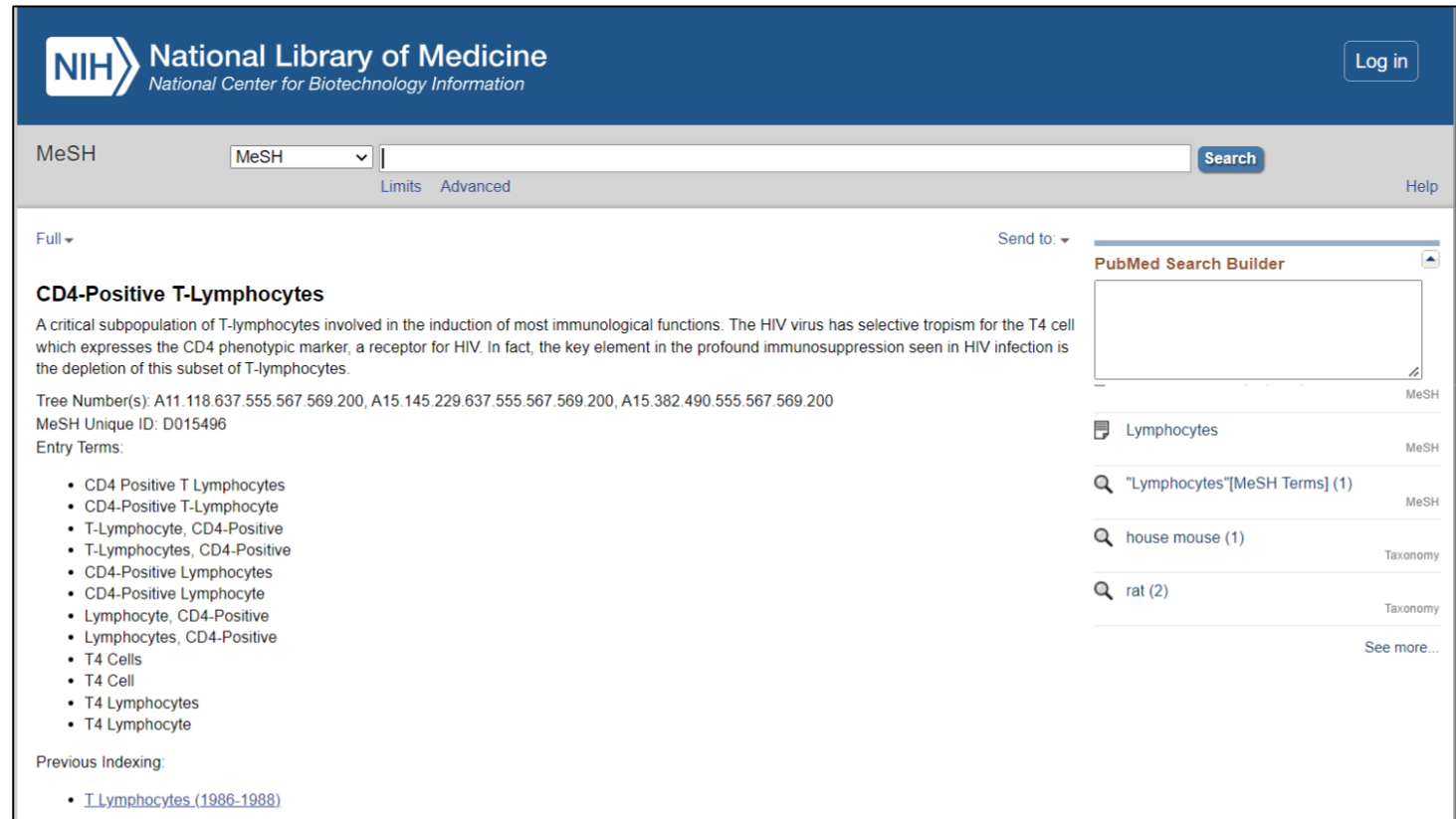
→ **Allows standardized usage of terms**

Controlled vocabularies examples

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 Help

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 \(1986-1988\)](#)

Lymphocytes MeSH

"Lymphocytes"[MeSH Terms] (1) MeSH

house mouse (1) Taxonomy

rat (2) Taxonomy

See more...

Controlled vocabularies examples

Systematized Nomenclature for Medicine – Clinical Terms

Curated by SNOMED International,
non-profit, membership fees, Germany
is a member

Delivering
SNOMED CT

SNOMED CT

SNOMED CT is currently the most comprehensive health terminology worldwide, a constantly growing ontology of preferred terms and their synonyms. The introduction of SNOMED CT in Germany is a building block for establishing semantic interoperability in the electronic exchange of health data.

→ [Read more](#)

https://www.bfarm.de/EN/Code-systems/Terminologies/_node.html

International Classification of Diseases (ICD), currently version 11

<https://en.wikipedia.org/wiki/ICD-11>

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), e.g., OWL, SKOS, OBO

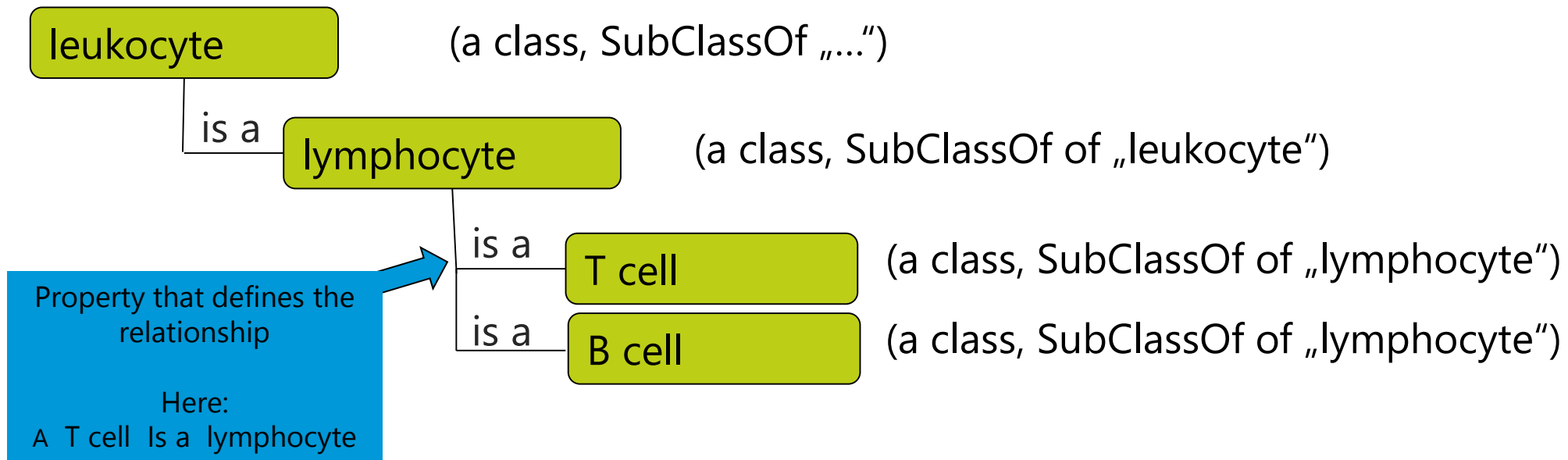
Examples of Ontologies:

- 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

Examples of potentially useful ontologies

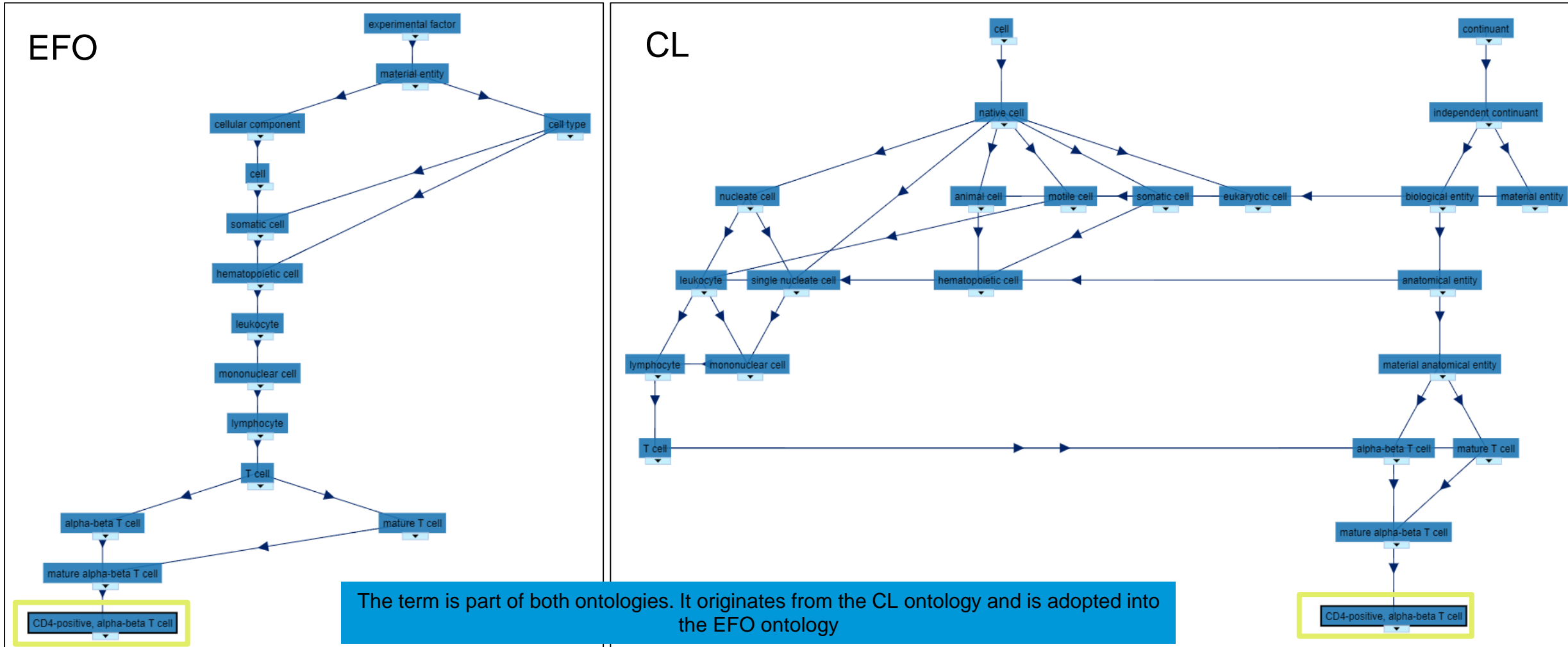
Different ontologies are designed to optimally **represent their respective domain knowledge** (for example, the relationship between terms)

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



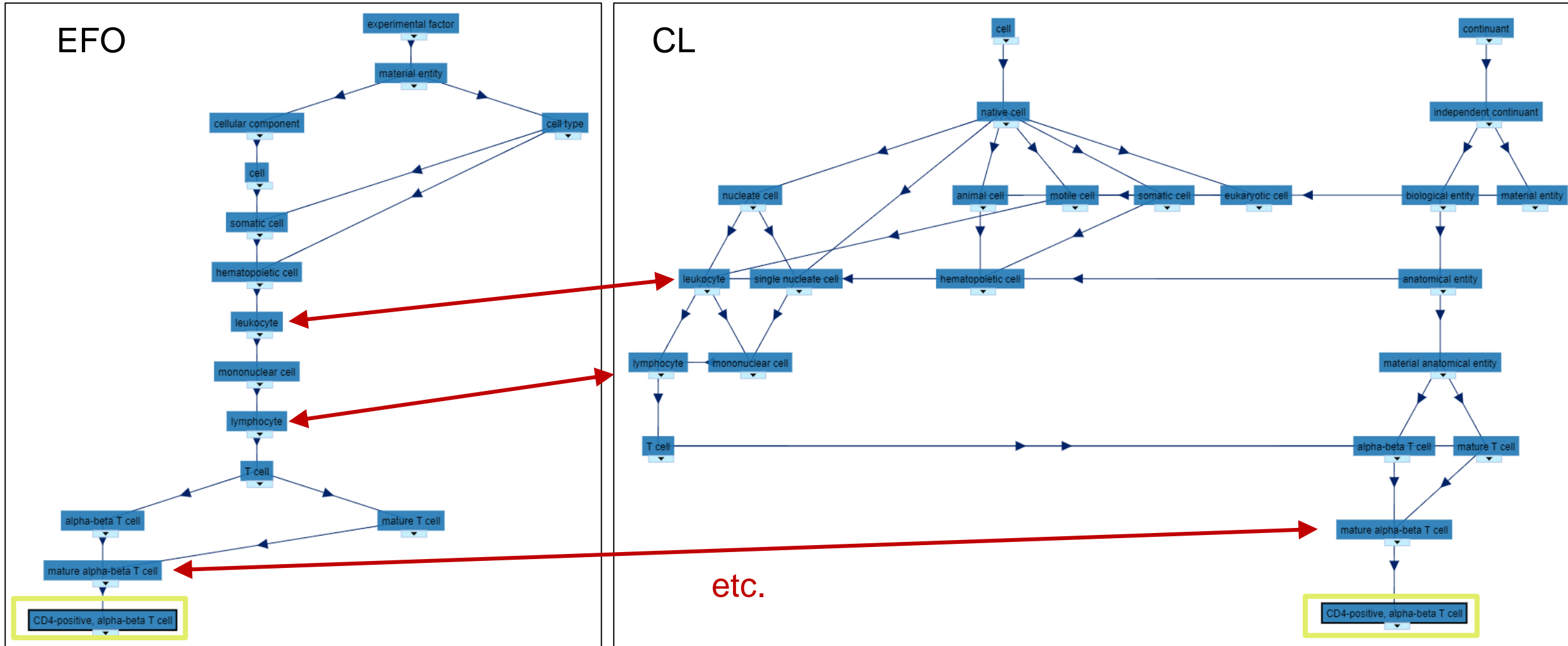
Examples of potentially useful ontologies

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

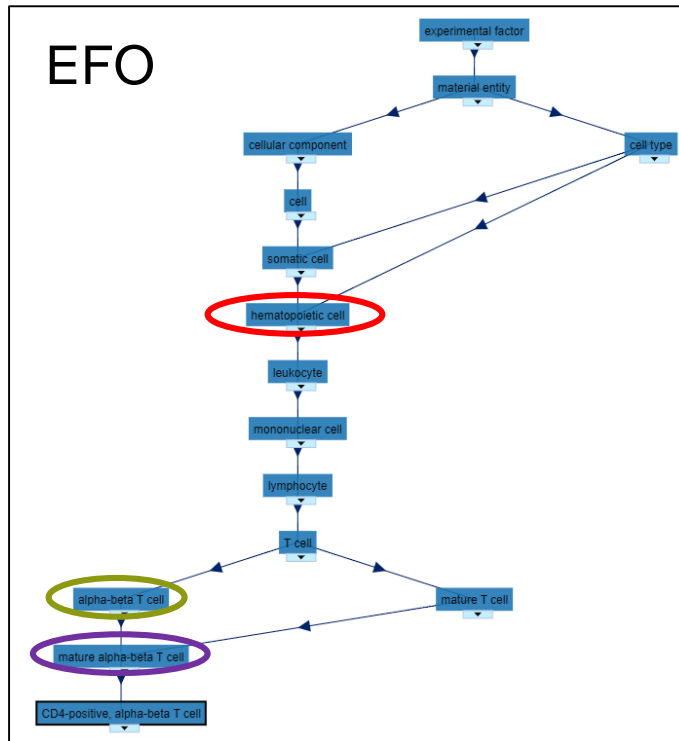


Adapted from: Schmidt C., Bortolomeazzi M., Boissonnet T., Fortmann-Grote C. *et al.* (2023). I3D:bio's OMERO training material: Re-usable, adjustable, multi-purpose slides for local user training. Zenodo. DOI: 10.5281/zenodo.8323588. If not stated otherwise, the content of this material (except for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

Mapping across ontologies



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!

Examples of potentially useful ontologies














BAO	BioAssays Ontology
EDAM (EDAM BioImaging)	Ontology of bioscientific data analysis and data management
EFO	Experimental Factor Ontology
CLO	Cell Line Ontology
CL	Cell Ontology
GO	Gene Ontology
UBERON	Uber Anatomy Ontology
FBbi	Biological Imaging Methods Ontology
ChEBI	Chemical Entities of Biological Interest

Inclusion and exclusion criteria for ontologies

- **Exclusion criteria:**

-  Absent licence or terms of use (*indicator of usability*)
-  Restrictive licences or terms of use with restrictions on redistribution and reuse
-  **Absence of term definitions**
-  Absence of sufficient class metadata (*indicator of quality*)
-  Absence of sustainability indicators (*absence of funding records*)

- **Inclusion criteria:**


-  Scope and coverage meets the requirements of the concept identified
-  Unique URI, textual definition and IDs for each term
-  Resource releases are versioned
-  Size of resource (*indicator of coverage*)
-  Number of classes and subclasses (*indicator of depth*)
-  Number of terms having definitions and synonyms (*indicator of richness*)
-  Presence of a help desk and contact point (*indicator of community support*)
-  Presence of term submission tracker/issue tracker (*indicator of resource agility and capability to grow upon request*)
-  Potential integrative nature of the resource (*as indicator of translational application potential*)
-  Licensing information available (*as indicator of freedom to use*)
-  Use of a top level ontology (*as indicator of a resource built for generic use*)
-  Pragmatism (*as indicator of actual, current real life practice*)
-  Possibility of collaborating: the resource accepts complaints/remarks that aim to fix or improve the terminology, while the resource organisation commits to fix or improve the terminology in brief delays (one month after receipt?)

These criteria are simply indicative and need to be modulated depending on the **contexts** described in the introduction, as specific constraints (e.g. regulatory requirements) may take precedence over some of the criteria listed here.


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

Ontologies (examples)

GENERAL INFORMATION



Experimental Factor Ontology (EFO)

 [10.25504/FAIRsharing.1gr4tz](https://doi.org/10.25504/FAIRsharing.1gr4tz)


Type Terminology artefact

Registry Standard


Description The Experimental Factor Ontology (EFO) is an application focused ontology modelling the experimental variables in multiple resources at the EBI and the Centre for Therapeutic Target Validation. The ontology has been developed to increase the richness of the annotations that are currently made in resources and to promote consistent annotation, to facilitate automatic annotation and to integrate external data. The ontology pulls together classes from reference ontologies such as disease, cell line, cell type and anatomy and adds axiomatisation as necessary to connect areas such as disease to phenotype.

Homepage <https://www.ebi.ac.uk/efo/>


Year of Creation 2010



GENERAL INFORMATION



BioAssay Ontology (BAO)

 [10.25504/FAIRsharing.mye76w](https://doi.org/10.25504/FAIRsharing.mye76w)


Type Terminology artefact


Registry Standard

Description The BioAssay Ontology (BAO) describes chemical biology screening assays and their results including high-throughput screening (HTS) data for the purpose of categorizing assays and data analysis.

Homepage <http://bioassayontology.org>

Year of Creation 2009

Maintainers [cchung](#) 



Demonstration and Group Task

Try out different ontology lookup services:

Ontology Lookup Service

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

Semantic Lookup Service

<https://semanticlookup.zbmed.de/ols/index>

BioPortal Bioontology

<https://bioportal.bioontology.org/>

Ontobee

<https://ontobee.org/>

Recommendation for Ontologies in OMERO

- Where possible, use terms that are derived from a useful ontology
- How to indicate the Ontology-compliant term choice:

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> or EFO

Why this style? Because it is already used by the ISA framework, hence, close to an annotation standard

- **Ontology compliance for *all* terms???**
 - Choose the essential keywords that represent your research
 - Identify a few ontologies that you can use sustainably

Acknowledgments

In cooperation with
Information Infrastructure for BioImage Data (I3D:bio)

<https://www.i3dbio.de/>

**German Cancer Research Center (DKFZ), Heidelberg
Department Enabling Technology**

Dr. Christian Schmidt, Project Coordinator

office@nfdi4bioimage.de