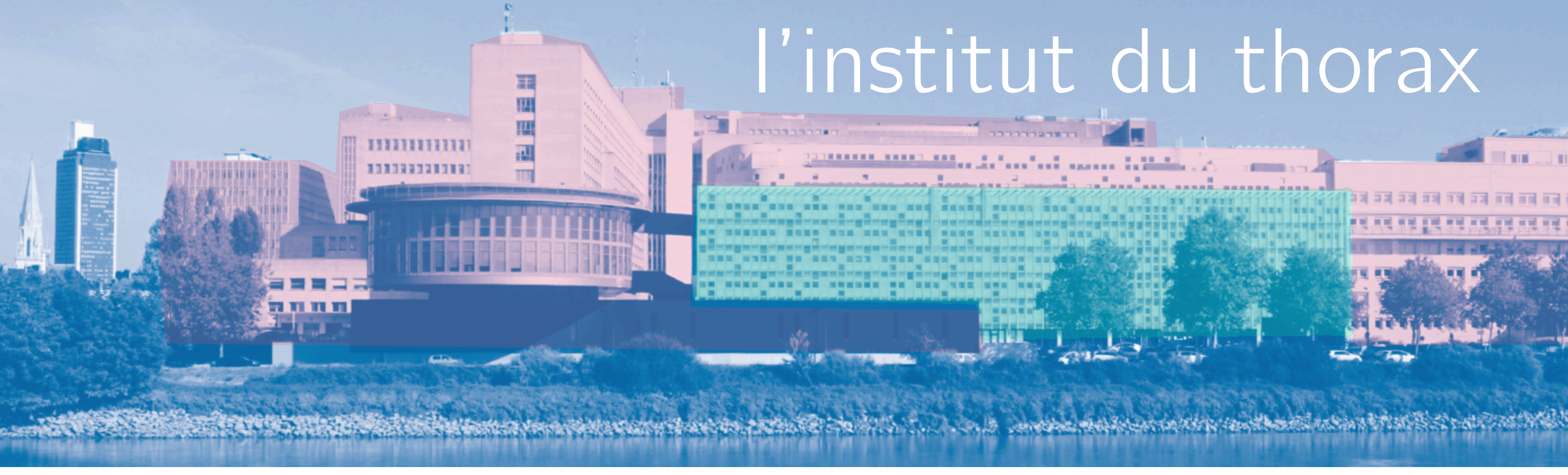
How to implement and verify FAIR principles ?

Alban Gaignard
CNRS, institut du thorax, Nantes, France

November 24, 2024

ANF "Workflow et reproductibilité" 2025, Paris

Who am I?



l'institut du thorax

Better understanding of cardio-vascular and metabolic diseases

Gene ←→ function

associations

Translational medicine

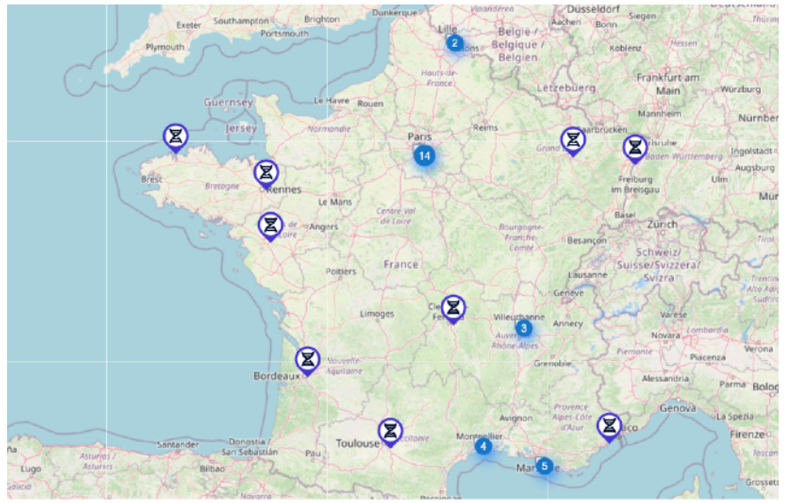
university hospital ⊕ reseach lab

Bioinformatics

- Massive production of genomic sequence & health data
  → Workflows + HPC
- Integration of multi-modal and multi-scale data
- Predictive models

l'institut du thorax

French Institute for Bioinformatics



A national research infrastructure for Bioinformatics providing:

Compute & Storage

Tools & Workfows, Databases, Training, Open Sciences

Communities: health, agronomy, biodiversity, microbiology

Actively involved in

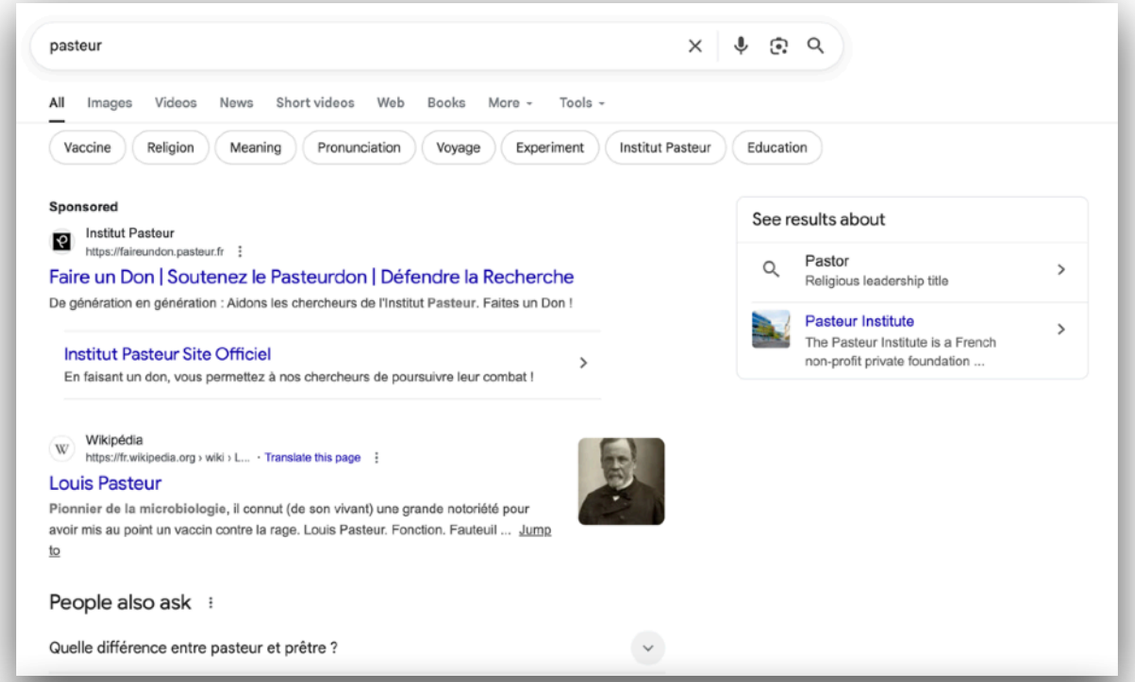
- **Open sciences & interoperability**: FAIR-Checker, metadata standards, ontologies (Bioschemas, EDAM)
- **Health community**: genomic data discoverability & sharing (Beacon, FEGA) + data integration (Knowledge Graphs)

A practical overview on FAIR assessment, in 2 hours ...

- 📚 **Quick intro on FAIR principles**
- **Knowledge graphs and semantic metadata**
  - 📚 What are Knowledge Graphs, Computational Ontologies ?
  - ✏️ How to create Knowledge Graphs ? (team work, "pen & paper")
  - ✏️ How to write machine-readable semantic metadata ? (demo, python code)
  - ✏️ How to query knowledge graphs with SPARQL (demo, python code)
- **Tools and resources for increased FAIRness**
  - 📚 FAIR-Checker
  - 📚 Tools and workflow registries
  - ✏️ Using FAIR-Checker and interpreting the FAIR assement results
  - ✏️ Using the FAIR-Checker API and processing multiple resources

FAIR principles

Being **findable** by both human and machines



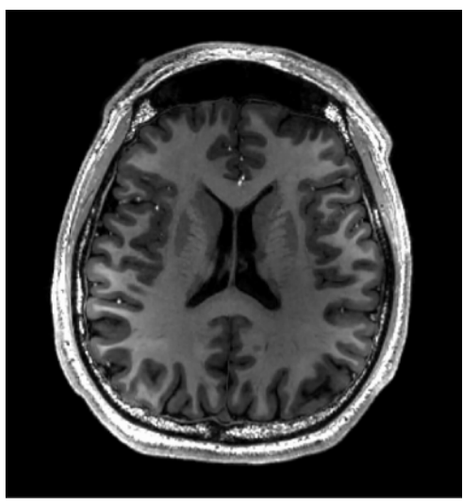
- "Pasteur" is not a good **identifier**
  → used to designate many "Humans"
  → used to designate many locations or institutions

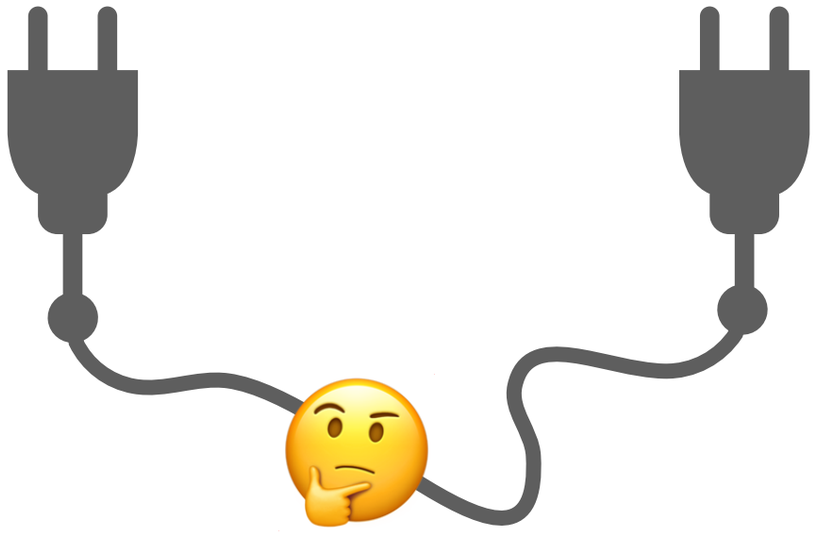
  We need **Persistent Globally Unique Identifiers**

- **Rich metadata** is key to identify **relevant** data
  - Entity kind/nature
  - Relationships with other entities

Being **interoperable**

```
@HWI-ST534_129:2:24:20503:16510:CGATGT
CTGAGAGCCGGGGAAGCCGGCGGAGCCGGGGGACTGGCGAGCCGGAACAT
+
HHHHHHHHHHEFDDGDDFBFGG>7D4<9;<&?:;<DC>CCDD@?=?A###
@HWI-ST534_129:2:42:2118:9580:CGATGT
GGCGGAGCCGGGTGACTGGCGAGCCGGAACATCAGGCGCCGCCGCAGAGA
+
GEECGGGBGIDF6FFFFEF=IDEFBEE8E8E?EEB@6=9B##########
@HWI-ST534_129:2:2:12654:80229:CGATGT
CGGAGCCGGGTGACTGGCGAGCCGGAACATCAGGCGCCGCCGCAGAGAAG
+
GGEGFCDCBBAEEEEGGFGFG;EGEEGFFBBEDEBDFGFCFF;DF2D<DD
@HWI-ST534_129:2:48:12356:179714:CGATGT
GAGCCGGGTGACTGGCGAGCCGGAACATCAGGCGCCGCCGCAGAGAAGAA
+
E=GHFHEGHHBCGDDBEEBBCBDDDE@EGBD=ABDCB?EC;@@8@EEB;E
@HWI-ST534_129:2:44:8225:39540:CGATGT
GGGTGACTGGCGAGCCGGAACATCAGGCGCCGCCGCAGAGAAGAACTATG
+
HHHEHHHHHHHHHGHHHHFHHHHHHHHDFDHHBHFDDFEFEFF>G<CCCE
```

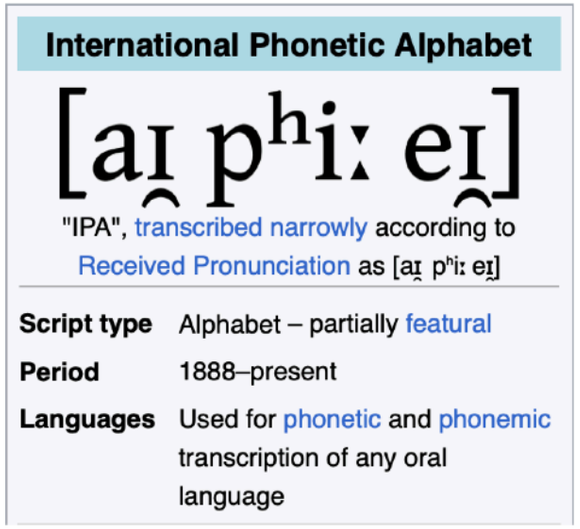




- **Technical**: exchange protocols compatible with different systems, e.g., HTTPS
- **Syntactic**: data and metadata structure and syntax, e.g., format, read/written by different systems
- **Semantic**: data and metadata can be understood/actioned upon by different systems
  $\rightarrow$ ontologies and vocabularies

Being **accessible** by both human and machines

- **Open (and free) communication protocols** (e.g. web standards) for everyone (human and software agents) ; CSV is a non-proprietary format



- Ability to specify **access control**

Image source: Wikipedia, http://www.flaticon.com

Being **re-usable**

- Clearly define a **usage license**
  $\rightarrow$ legal condition for (re)use

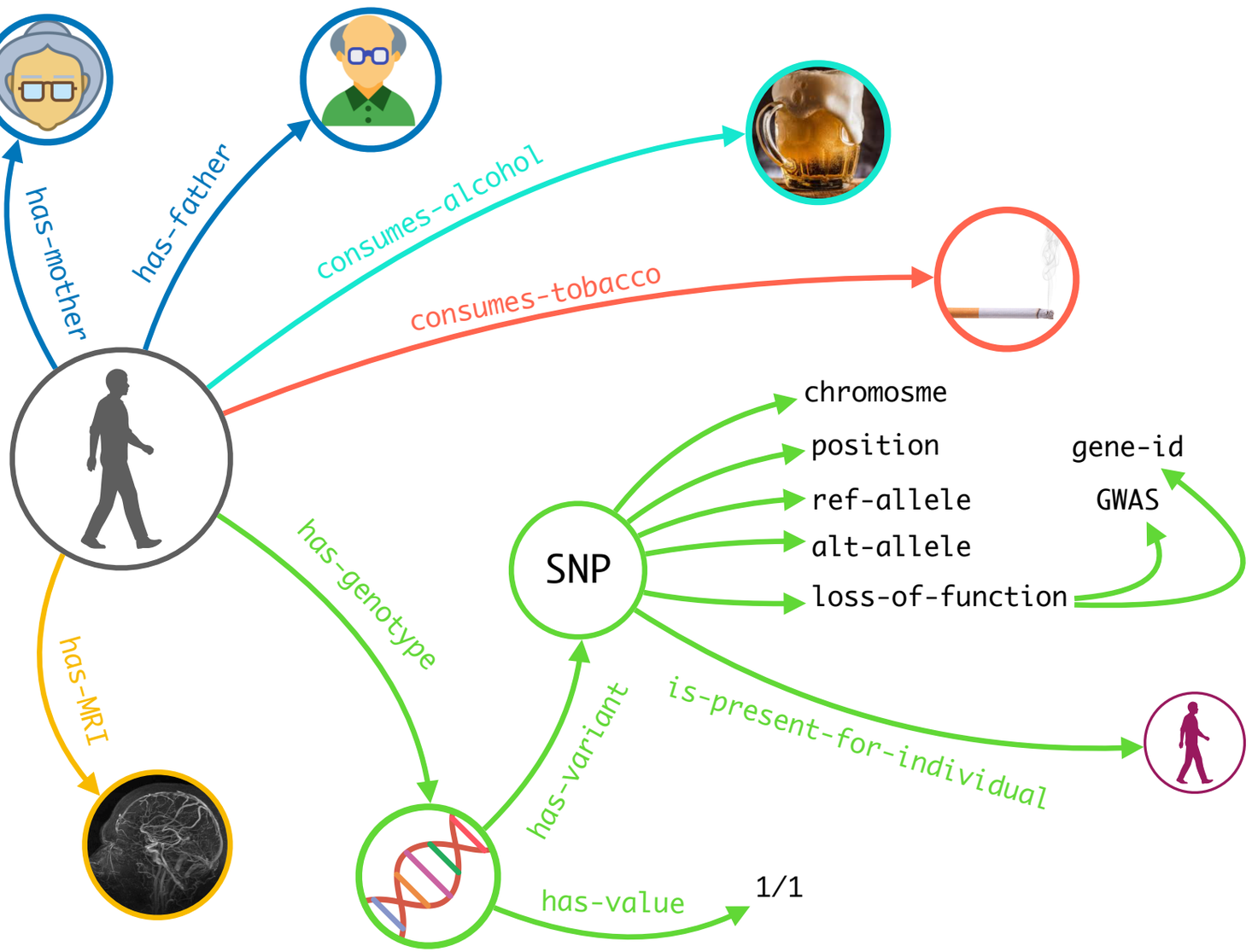
- Cumulative sciences need **provenance**
  - crediting
    - funders
    - creators
  - fine-grained documentation of **data production**
    fine-grained documentation of **software components**

- **Community standards**

Knowledge Graphs

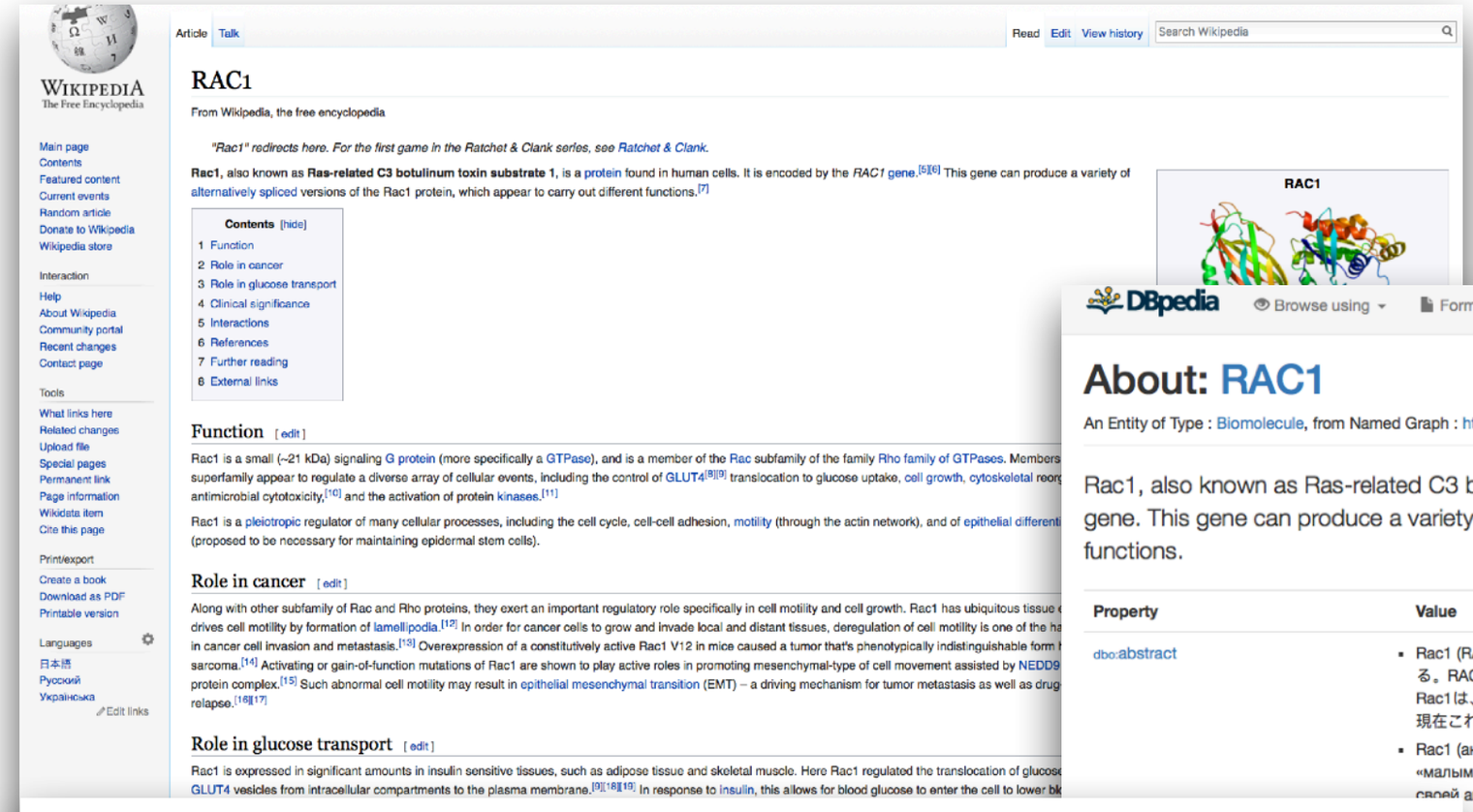
❶ Link data with Knowledge Graphs

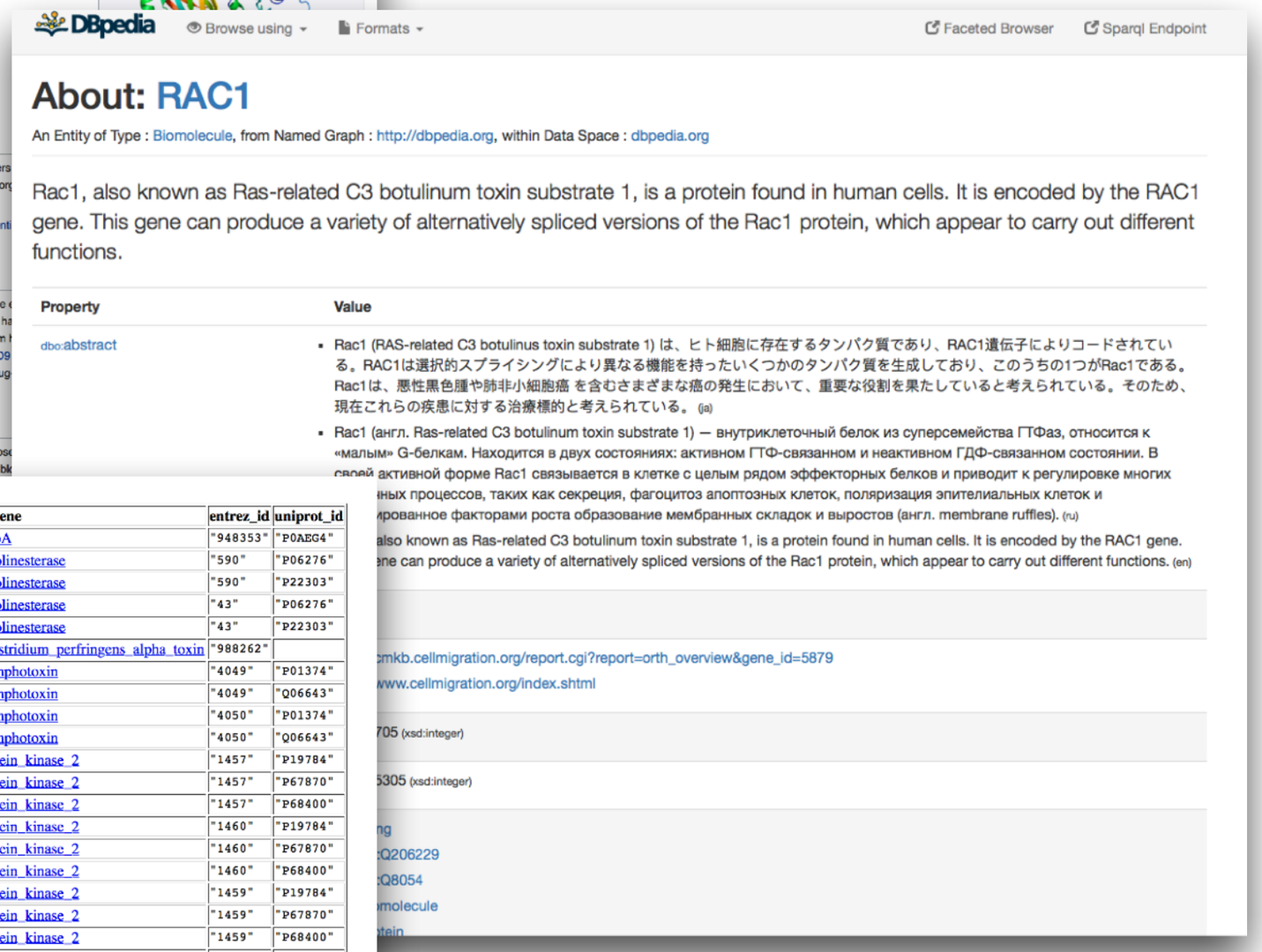
« a collection of interlinked descriptions of things
(real-word objects, abstract concepts, events, etc.) »



- ✓ a **database** to store and retrieve information
- ✓ a **graph** to represent multiple relationships and to perform network / community analysis
- ✓ a **knowledge base** with formal semantics to perform logical reasoning, inferences ...

Wikipedia → DBpedia Knowledge Graph





http://dbpedia.org/sparql

Virtuoso SPARQL Query Editor

About | Namespace Prefixes | Inference rules | RDF views | iSPARQL

Default Data Set Name (Graph IRI)

http://dbpedia.org

Query Text

```
SELECT DISTINCT ?gene ?entrez_id ?uniprot_id WHERE {
    ?gene dbo:abstract ?abstract .
    FILTER (regex(?abstract, "toxin")).
    ?gene dbo:entrezgene ?entrez_id .
    OPTIONAL {?gene dbo:uniprot ?uniprot_id} .
}
```

(Security restrictions of this server do not allow you to retrieve remote RDF data, see details.)

Results Format: HTML

Execution timeout: 30000 milliseconds *(values less than 1000 are ignored)*

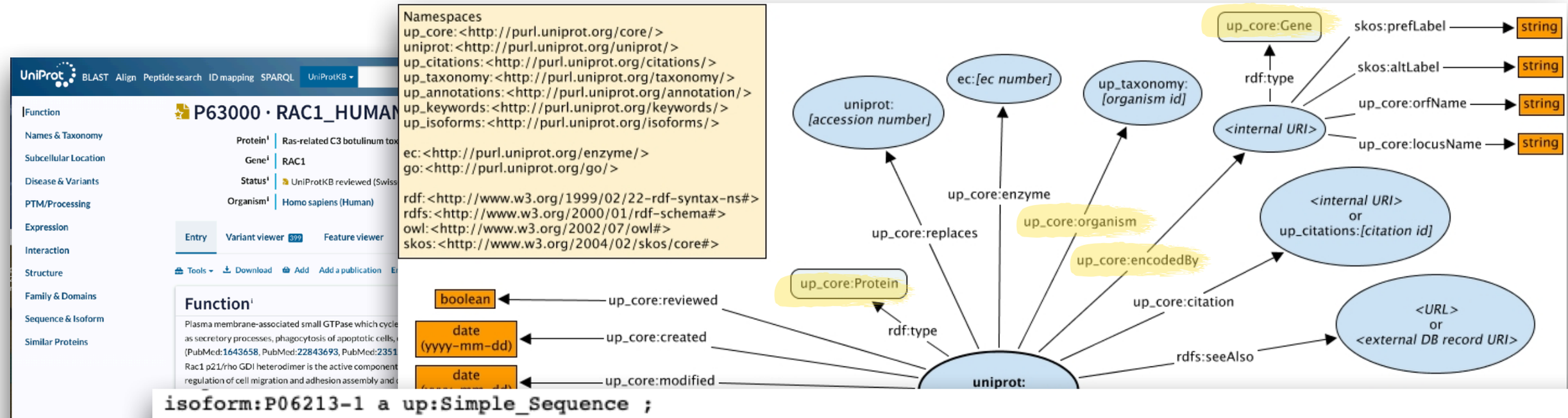
Options:
- [x] Strict checking of void variables
- [ ] Log debug info at the end of output (has no effect on some queries and output formats)
- [ ] Generate SPARQL compilation report (instead of executing the query)

(The result can only be sent back to browser, not saved on the server, see details)

Run Query Reset

gene	entrez_id	uniprot_id
http://dbpedia.org/resource/DsbA	"948353"	"P0AEG4"
http://dbpedia.org/resource/Cholinesterase	"590"	"P06276"
http://dbpedia.org/resource/Cholinesterase	"590"	"P22303"
http://dbpedia.org/resource/Cholinesterase	"43"	"P06276"
http://dbpedia.org/resource/Cholinesterase	"43"	"P22303"
http://dbpedia.org/resource/Clostridium_perfringens_alpha_toxin	"988262"	
http://dbpedia.org/resource/Lymphotoxin	"4049"	"P01374"
http://dbpedia.org/resource/Lymphotoxin	"4049"	"Q06643"
http://dbpedia.org/resource/Lymphotoxin	"4050"	"P01374"
http://dbpedia.org/resource/Lymphotoxin	"4050"	"Q06643"
http://dbpedia.org/resource/Casein_kinase_2	"1457"	"P19784"
http://dbpedia.org/resource/Casein_kinase_2	"1457"	"P67870"
http://dbpedia.org/resource/Casein_kinase_2	"1457"	"P68400"
http://dbpedia.org/resource/Casein_kinase_2	"1460"	"P19784"
http://dbpedia.org/resource/Casein_kinase_2	"1460"	"P67870"
http://dbpedia.org/resource/Casein_kinase_2	"1460"	"P68400"
http://dbpedia.org/resource/Casein_kinase_2	"1459"	"P19784"
http://dbpedia.org/resource/Casein_kinase_2	"1459"	"P67870"
http://dbpedia.org/resource/Casein_kinase_2	"1459"	"P68400"
http://dbpedia.org/resource/Collagenase	"4317"	"P03956"
http://dbpedia.org/resource/Collagenase	"4317"	"P22894"
http://dbpedia.org/resource/Collagenase	"4312"	"P03956"
http://dbpedia.org/resource/Collagenase	"4312"	"P22894"
http://dbpedia.org/resource/Guanylin	"2980"	"Q02747"
http://dbpedia.org/resource/Macrophage_inflammatory_protein	"6348"	"P10147"
http://dbpedia.org/resource/Macrophage_inflammatory_protein	"6348"	"P13236"
http://dbpedia.org/resource/Macrophage_inflammatory_protein	"6351"	"P10147"
http://dbpedia.org/resource/Macrophage_inflammatory_protein	"6351"	"P13236"

Uniprot Knowledge Graph




```
isoform:P06213-1 a up:Simple_Sequence ;
  up:modified "2010-10-05"^^xsd:date ;
  up:version 4 ;
  up:precursor true ;
  up:mass 156333 ;
```

There are 217,505,202,099 triples is this release. All triples are available in the default graph. There are 22 named graphs corresponding to specific datasets.

Graph	Documentation	Triples	Distinct subjects	Distinct predicates	Distinct classes	Distinct objects	License
uniparc	Documentation	160,189,731,200	40,455,837,024	29	6	46,916,767,863	http://creativecommons.org/licenses/by/4.0/
uniprot	Documentation	44,256,643,227	9,441,439,078	124	121	8,462,262,751	http://creativecommons.org/licenses/by/4.0/
uniref	Documentation	10,224,623,630	1,393,813,725	14	3	1,409,539,937	http://creativecommons.org/licenses/by/4.0/
obsolete	Documentation	2,102,255,458	277,358,373	10	3	286,609,935	http://creativecommons.org/licenses/by/4.0/
citationmapping	Documentation	625,262,380	123,810,071	12	4	29,448,749	http://creativecommons.org/licenses/by/4.0/
taxonomy	Documentation	60,041,721	26,918	21	4	4,698,602	http://creativecommons.org/licenses/by/4.0/
citations	Documentation	31,212,544	419,769	19	5	8,870,230	http://creativecommons.org/licenses/by/4.0/
proteomes	Documentation	8,984,258	1,999,807	33	11	3,777,324	http://creativecommons.org/licenses/by/4.0/
chebi	Documentation	3,419,539	221,830	24	6	1,828,527	http://creativecommons.org/licenses/by/4.0/
rhea	Documentation	1,962,186	138,720	67	3	540,446	http://creativecommons.org/licenses/by/4.0/

Uniprot Knowledge Graph

All protein sequences associated to Escherichia coli K-12 ?

UniProt SPARQL Downloads Documentation/Help

Your SPARQL query

Add common prefixes

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX taxon: <http://purl.uniprot.org/taxonomy/>
PREFIX up: <http://purl.uniprot.org/core/>
SELECT ?protein ?organism ?isoform ?sequence
WHERE
{
    ?protein a up:Protein .
    ?protein up:organism ?organism .
    # Taxon subclasses are materialized, do not use rdfs:subClassOf+
    ?organism rdfs:subClassOf taxon:83333 .
    ?protein up:sequence ?isoform .
    ?isoform rdf:value ?sequence .
}
```

Submit Query

Examples

1. Select all taxa from the UniProt taxonomy **Use**
2. Select all bacterial taxa and their scientific name from the UniProt taxonomy **Use**
3. Select all UniProtKB entries, and their organism and amino acid sequences (including isoforms), for *E. coli K12* and all its strains **Use**
4. Select the UniProtKB entry with the mnemonic 'A4_HUMAN' **Use**
5. Select a mapping of UniProtKB to PDB entries using the UniProtKB cross-references to the PDB database **Use**
6. Select all cross-references to external databases of the category '3D structure databases' of UniProtKB entries that are classified with the keyword 'Acetoin biosynthesis (KW-0005)' **Use**
7. Select reviewed UniProtKB entries (Swiss-Prot), and their recommended protein name, that have a preferred gene name that contains the text 'DNA' **Use**
8. Select the preferred gene name and disease annotation of all human UniProtKB entries that are known to be involved in a disease **Use**
9. Select all human UniProtKB entries with a sequence variant that leads to a 'loss of function' **Use**
10. Select all human UniProtKB entries with a sequence variant that leads to a tyrosine to phenylalanine substitution **Use**
11. Select all UniProtKB entries with annotated transmembrane regions and the regions' begin and end coordinates on the canonical sequence **Use**
12. Select all UniProtKB entries that were integrated on the 30th of November 2010 **Use**
13. Was any UniProtKB entry integrated on the 9th of January 2013 **Use**
14. Construct new triples of the type 'HumanProtein' from all human UniProtKB entries **Use**
15. Select the average number of cross-references to the PDB database of UniProtKB entries that have at least one cross-reference to the PDB database **Use**
16. More examples

About

This SPARQL endpoint contains all UniProt data. It is free to access and supports the SPARQL 1.1 Standard.

There are 217,505,202,099 triples in this release (2025_04). The query timeout is 45 minutes. All triples are available in the default graph. There are 22 named graphs.

Documentation

1. Classes and predicates defined by the UniProt consortium
2. Tutorial on using SPARQL with UniProt
3. Statistics and diagrams
4. Example queries

News

Forthcoming changes

Table of contents * [Reorganizing the protein space in ...

UniProt release 2025_04

The (RPs) provided by UniProt aim to ...

UniProt release 2025_03

Cross-references have been added to the CARD database, The Comprehensive Antibiotic Resistance Database. CARD is ...

News archive

Massive (and diverse) data already available in the form of **knowledge graphs** ...

... how can we ensure machines/humans speak the **same language** ?

What about **semantic interoperability** ?

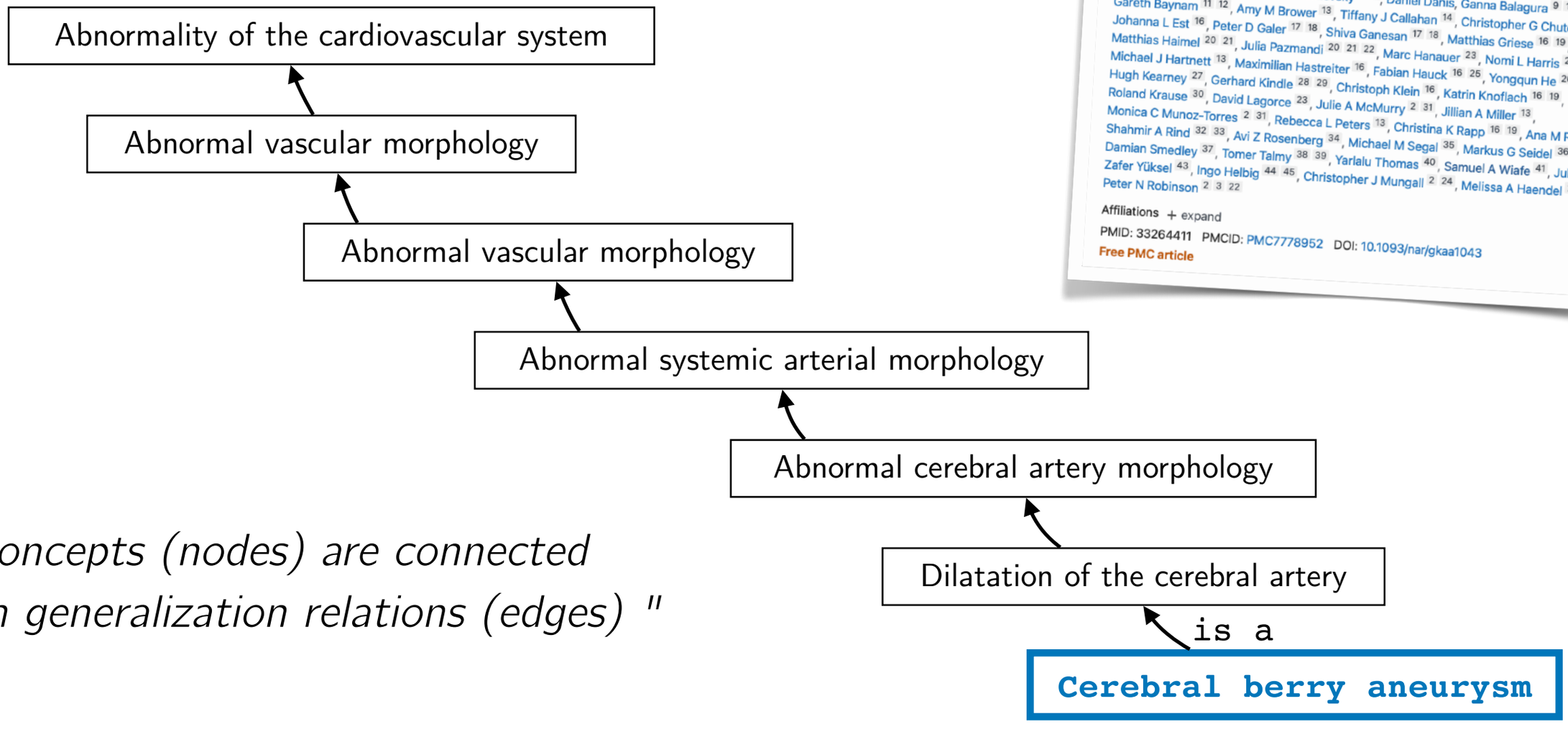
❷ Uniformly describe **what** is observed with data ?

Computational ontology

« a formal specification of a shared conceptualization » (Borst, 1997)

→ 1,049 life science ontologies registered in BioPortal (2023)

Human Phenotype Ontology



> Nucleic Acids Res. 2021 Jan 8;49(D1):D1207-D1217. doi: 10.1093/nar/gkaa1043.

The Human Phenotype Ontology in 2021

Sebastian Köhler [1 2], Michael Gargano [2 3], Nicolas Matentzoglu [2 4 5], Leigh C Carmody [2 3], David Lewis-Smith [6 7], Nicole A Vasilevsky [2 8], Daniel Danis, Ganna Balagura [9 10], Gareth Baynam [11 12], Amy M Brower [13], Tiffany J Callahan [14], Christopher G Chute [15], Johanna L Est [16], Peter D Galer [17 18], Shiva Ganesan [17 18], Matthias Griese [16 19], Matthias Haimel [20 21], Julia Pazmandi [20 21 22], Marc Hanauer [23], Nomi L Harris [2 24], Michael J Hartnett [13], Maximilian Hastreiter [16], Fabian Hauck [16 25], Yongqun He [26], Tim Jeske [16], Hugh Kearney [27], Gerhard Kindle [28 29], Christoph Klein [16], Katrin Knoflach [16 19], Roland Krause [30], David Lagorce [23], Julie A McMurry [2 31], Jillian A Miller [13], Monica C Munoz-Torres [2 31], Rebecca L Peters [13], Christina K Rapp [16 19], Ana M Rath [23], Shahmir A Rind [32 33], Avi Z Rosenberg [34], Michael M Segal [35], Markus G Seidel [36], Damian Smedley [37], Tomer Talmy [38 39], Yarlalu Thomas [40], Samuel A Wiafe [41], Julie Xian [17 42], Zafer Yüksel [43], Ingo Helbig [44 45], Christopher J Mungall [2 24], Melissa A Haendel [2 8 31], Peter N Robinson [2 3 22]

Affiliations + expand

PMID: 33264411 PMCID: PMC7778952 DOI: 10.1093/nar/gkaa1043
Free PMC article

" Concepts (nodes) are connected with generalization relations (edges) "

BioPortal

BioPortal Ontologies Search Annotator Recommender Mappings Login

Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies

Search for a class

Enter a class, e.g. Melanoma

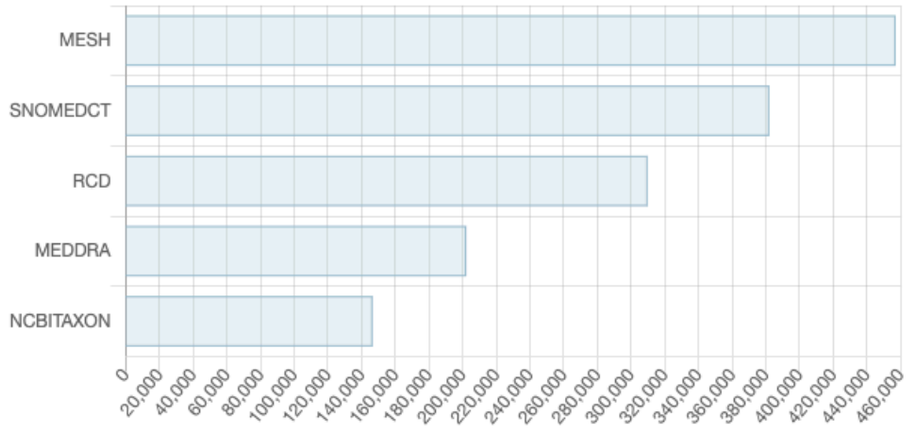
Advanced search

Find an ontology

Start typing ontology name, then choose from list

Browse ontologies

Ontology visits (October 2025)



More

Statistics

Ontologies	1,233
Classes	17,558,240
Properties	36,286
Mappings	92,868,606

Environment Ontology

Last uploaded: October 22, 2025

Jump to

- biological_process
- entity
  - continuant
    - generically dependent continuant
    - independent continuant
      - anatomical entity
      - immaterial entity
      - material entity
        - anthropogenic litter
        - astronomical body part
          - abyssal clay
          - acid dune sand
          - alpine tree line ecotone
          - aquatic ecosystem
          - aquatic natural environment
          - area of attached mussel assemblages
          - area of drift ice
          - area of open water
          - area of pack ice
          - area of perennial ice or snow
          - area of sea ice
          - beach sand
          - biome
            - alpine biome
              - **alpine tundra biome**
            - aquatic biome

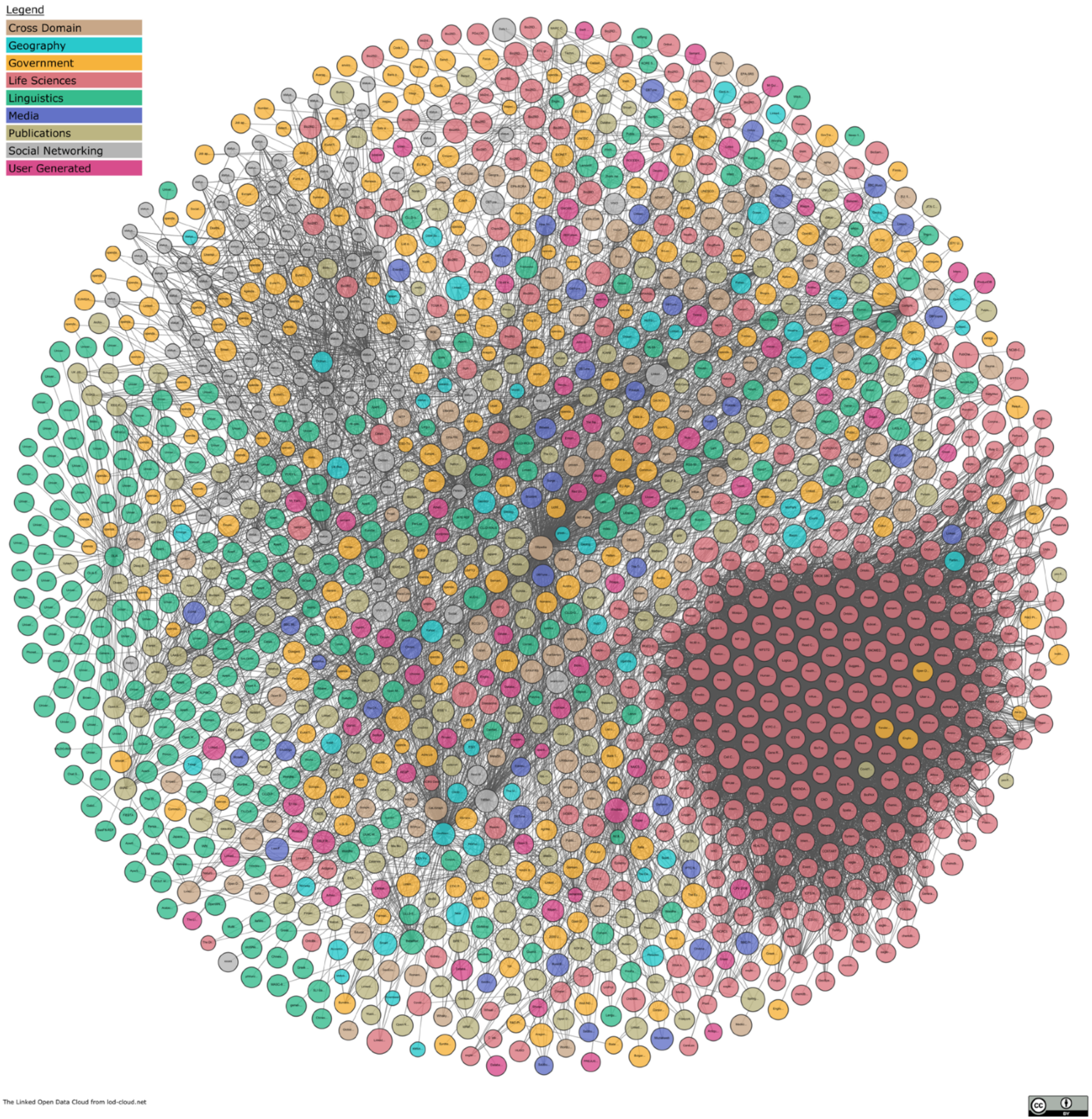
Details Visualization Notes (0) Mappings (8)

Add a proposal

Id	http://purl.obolibrary.org/obo/ENVO_01001505
Preferred Name	alpine tundra biome
Definitions	A tundra biome which exists at high altitudes and where vegetation - dominated by a few species of dwarf shrubs, a few grasses, sedges, lichens, and mosses - is stunted due to low temperatures and high winds. The absence of trees in this biome is primarily due to high altitude rather than high latitude. On Earth, it lies roughly between the summer isotherm of 10 degrees Centigrade and the snow line. Primary productivity is low in this biome because of the extremes of climate.
Synonyms	mountain tundra
Type	http://www.w3.org/2002/07/owl#Class

All Properties

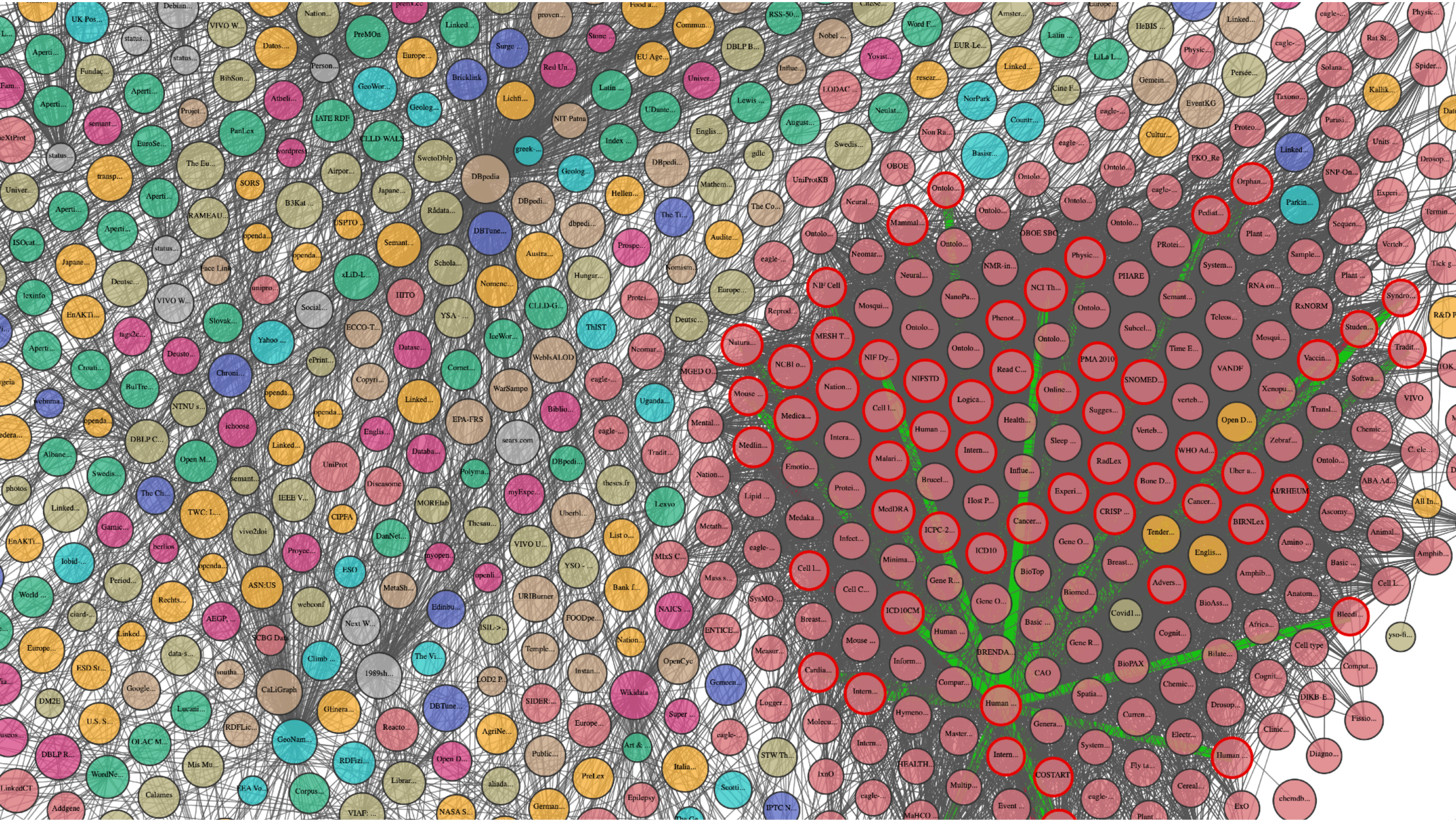
definition	A tundra biome which exists at high altitudes and where vegetation - dominated by a few species of dwarf shrubs, a few grasses, sedges, lichens, and mosses - is stunted due to low temperatures and high winds.
label	alpine tundra biome
comment	The absence of trees in this biome is primarily due to high altitude rather than high latitude. On Earth, it lies roughly between the summer isotherm of 10 degrees Centigrade and the snow line. Primary productivity is low in this biome because of the extremes of climate.
prefLabel	alpine tundra biome
database_cross_reference	SPIRE:Tundra http://sweetontology.net/realmCryo/AlpineTundra
in_subset	envoPolar



Linked Open Data Cloud → 1357 interlinked RDF datasets in 2025

https://lod-cloud.net/versions/2025-09-02/lod-cloud.svg

Life-science semantic resources are massively interconnected



Knowledge Graphs are instrumental for FAIR principles

- By-design, built for being both **human and machine-readable**

  → Interoperability ✅

- Semantic Web technologies provide open and **standard protocols**: URLs / HTTP / RDF format / SPARQL query language (W3C standards)

  → Findability ✅ (URIs for identifying things on the web)

  → Accessibility ✅

  → Interoperability ✅

- Ontologies are **community**-aggreed controlled vocabularies

  → Interoperability ✅

  → Reuse ✅

Creating machine-readable metadata with "mini Knowledge Graphs"

RDF to link data

Definitions

(1) An RDF **statement** expresses a **relationship** between two resources (things)

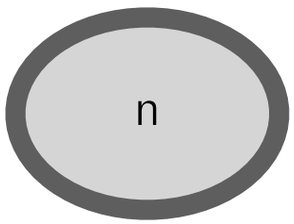
(2) The **subject** and the **object** represent the two resources being related ; the **predicate** represents the nature of their relationship

(3) The relationship is phrased in a **directional** way (from subject to object) and is called in RDF a **property**

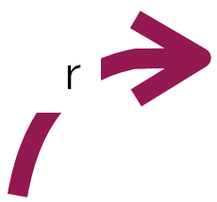
RDF triples are simple "subject | verb | object" sentences:

```
<RAC1> <is a> <human gene> .
<RAC1> <has_label> "Rac Family Small GTPase 1" .
<seq1> <is a variant of> <RAC1> .
```

Graphical syntax

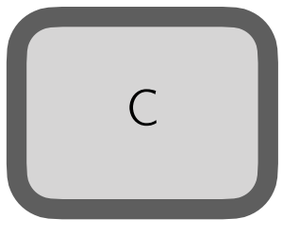


: a node in the knowledge graph



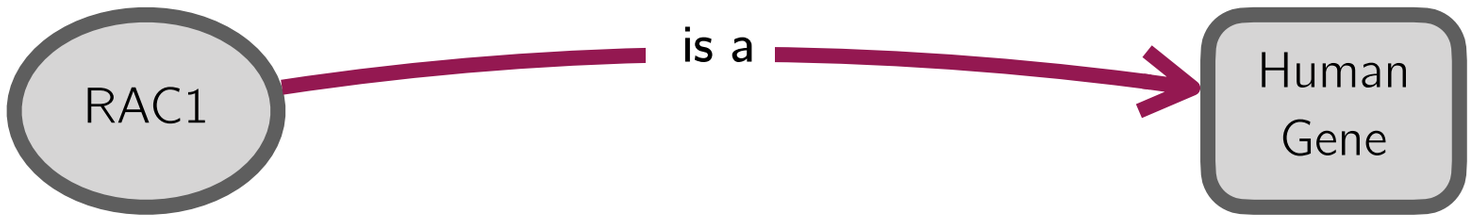
: a property/relation/edge in the knowledge graph

"…" : a literal → simple textual, numerical, boolean, date value

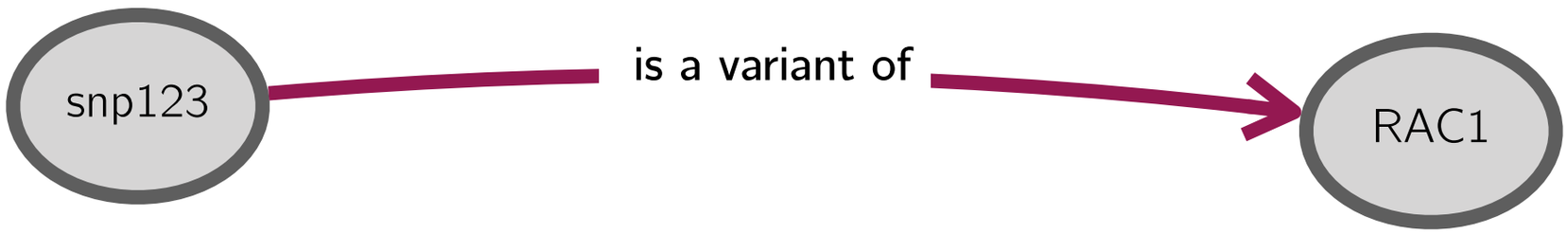


: a special node, denoting an ontology concept/class

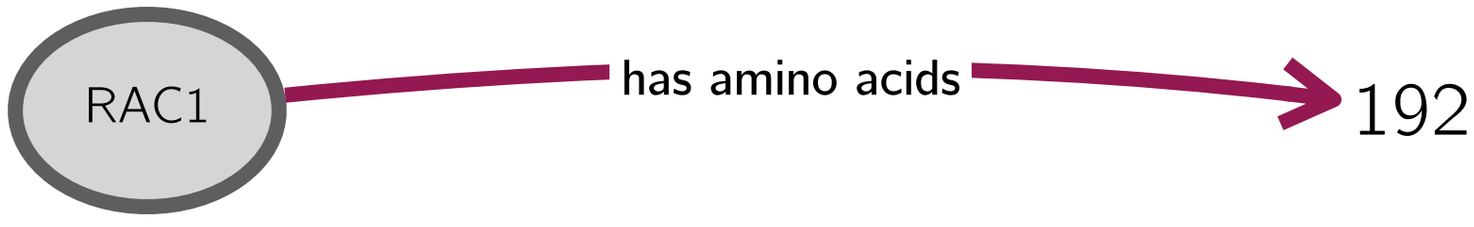
Examples




```
<http://RAC1> <http://is_a> <http://Human_Gene> .
```




```
<http://snp123> <http://is_a_variant_of> <http://RAC1> .
```



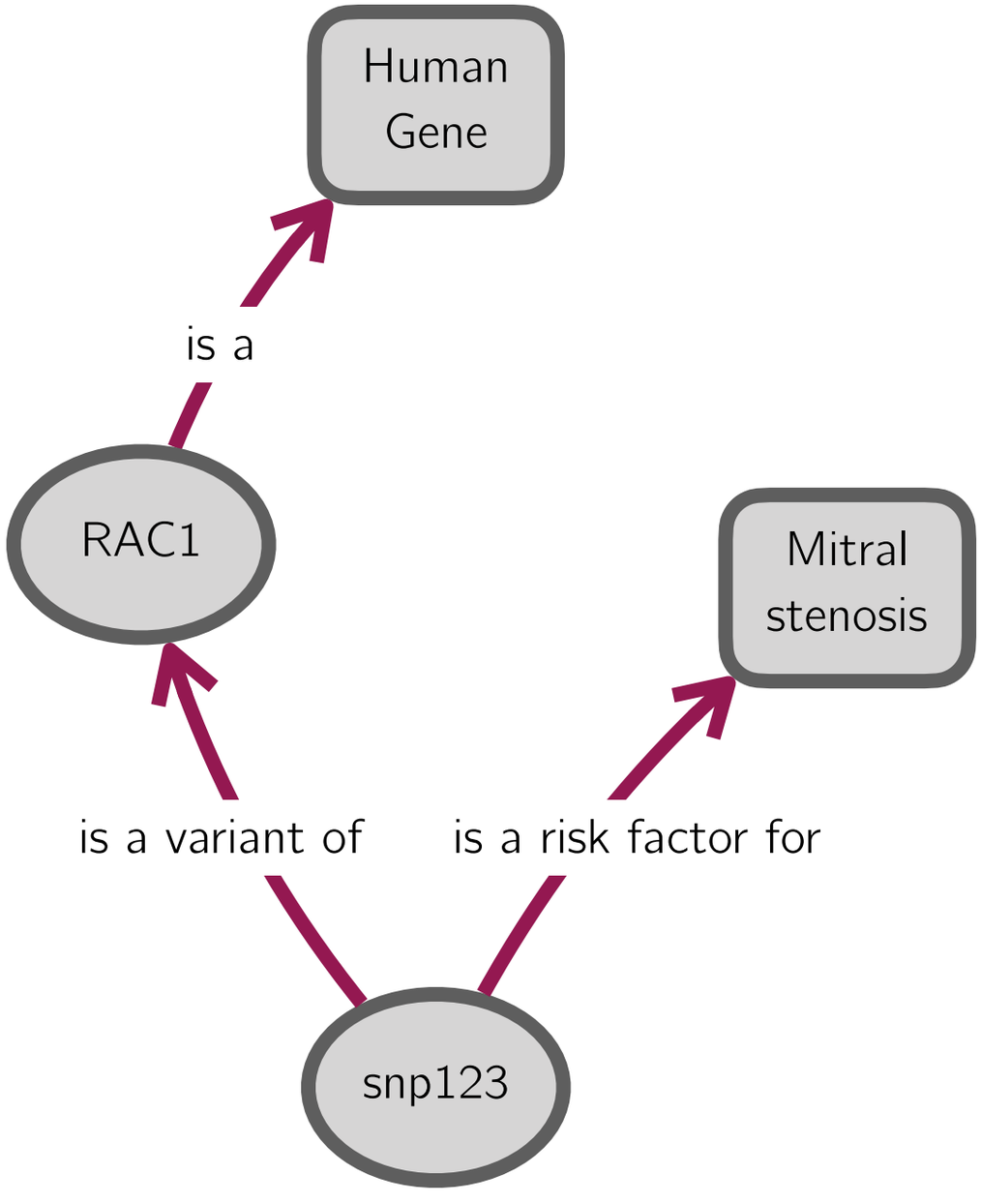

```
<http://RAC1>  <http://has_amino_acids>  192 .
```

RDF graphs

Definitions

(1) A **graph** structure is formed with a set of **nodes** (resources) and **edges**(relationships between resources)

(2) A set of RDF triples is called an RDF graph. RDF is a **directed**, **labeled graph** data format for representing information/ knowledge on the Web.



Writing RDF graphs with the Turtle syntax

Definitions

(1) One line per triple, each element separated by **space**, each triple endswith a **.**

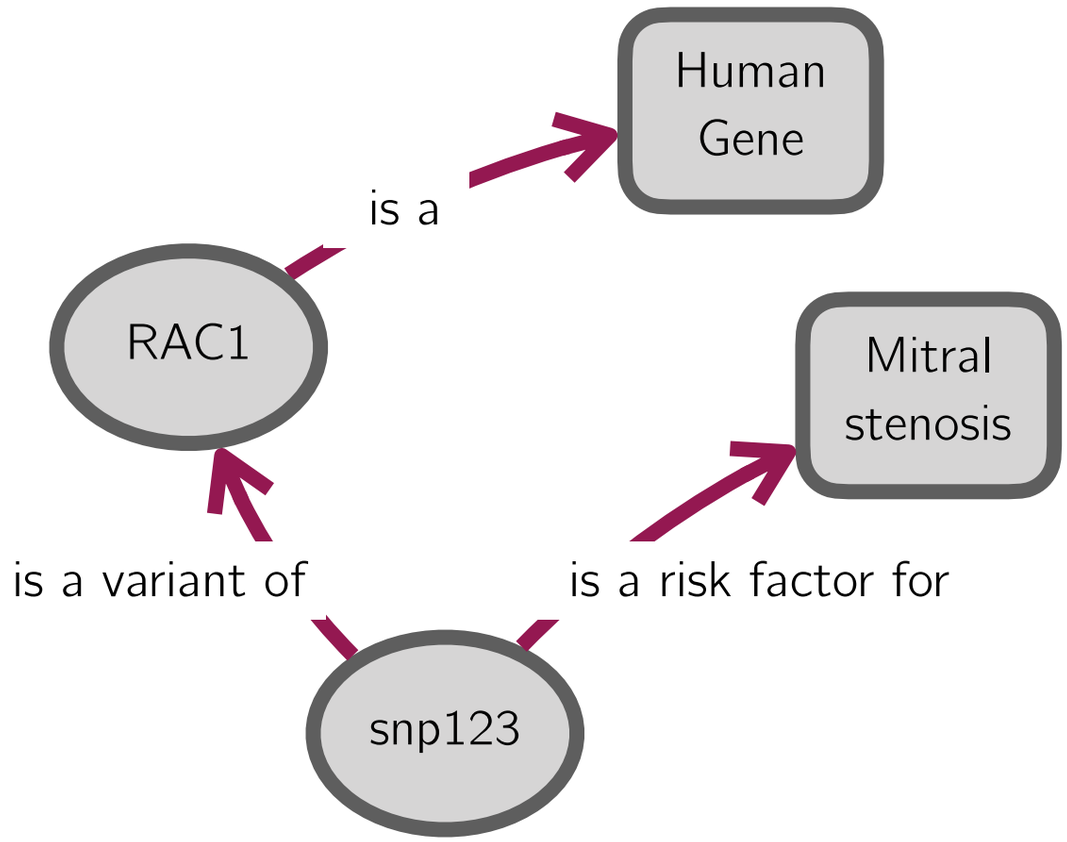
S P O .

(2) If two triples describe the same subject, you can reuse it:

S P_1 O_1 ;
P_2 O_2 .

(3) If two triples describe the same subject and predicate, you can reuse it:

S P O_1 , O_2 .




```
@prefix ns: <http://my/namespace/> .


ns:RAC1     rdf:type ns:Human_gene .

ns:snp123   ns:is_a_variant_of      ns:RAC1 ;

            ns:is_a_risk_factor_for ns:Mitral_stenosis .
```

Hands-on session: from text to KG ✏️

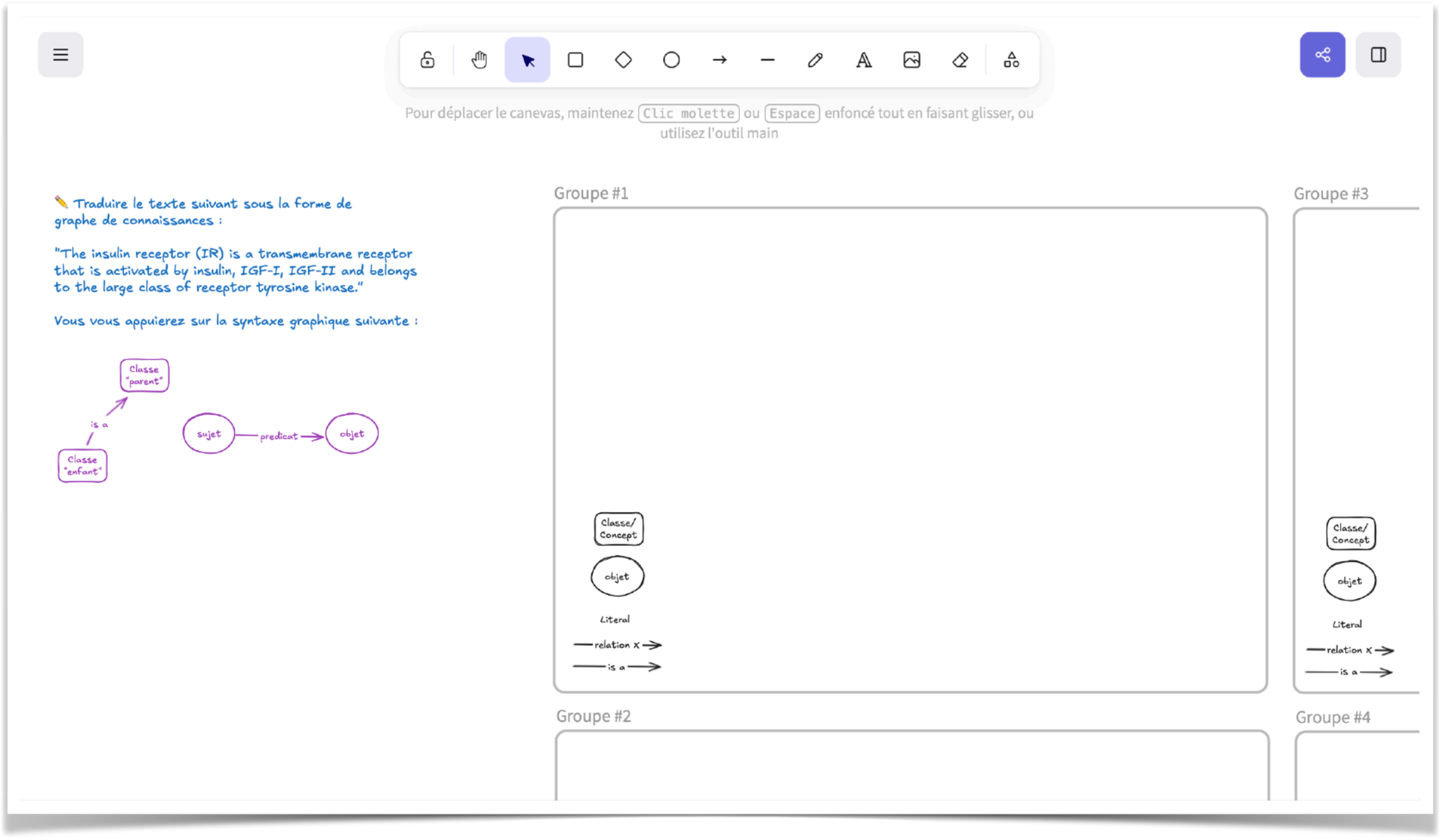
Question #1

From wikipedia : *"The insulin receptor (IR) is a transmembrane receptor that is activated by insulin, IGF-I, IGF-II and belongs to the large class of receptor tyrosine kinase."*

Draft a **graphical representation** of the associated knowledge graph.

- ✓ Identify verbs $\rightarrow$ RDF predicates
- ✓ Identify linked entities,
  - who is a subject of a relation ?
  - who is the object of a relation ?

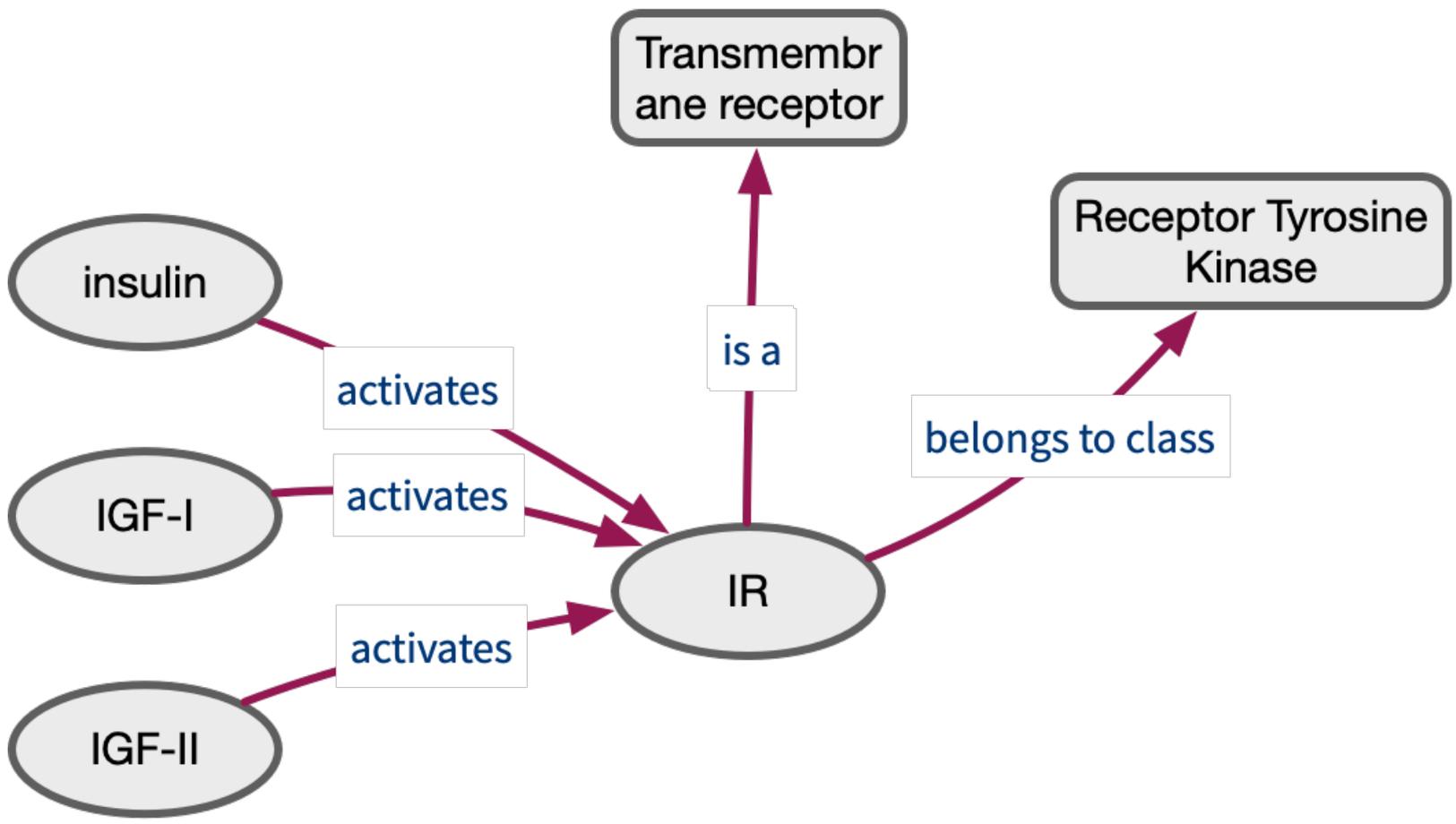
"Pen & paper" team work



https://tinyurl.com/4xf83nxd

One of the many solutions

"The insulin receptor (IR) is a transmembrane receptor that is activated by insulin, IGF-I, IGF-II and belongs to the large class of receptor tyrosine kinase."



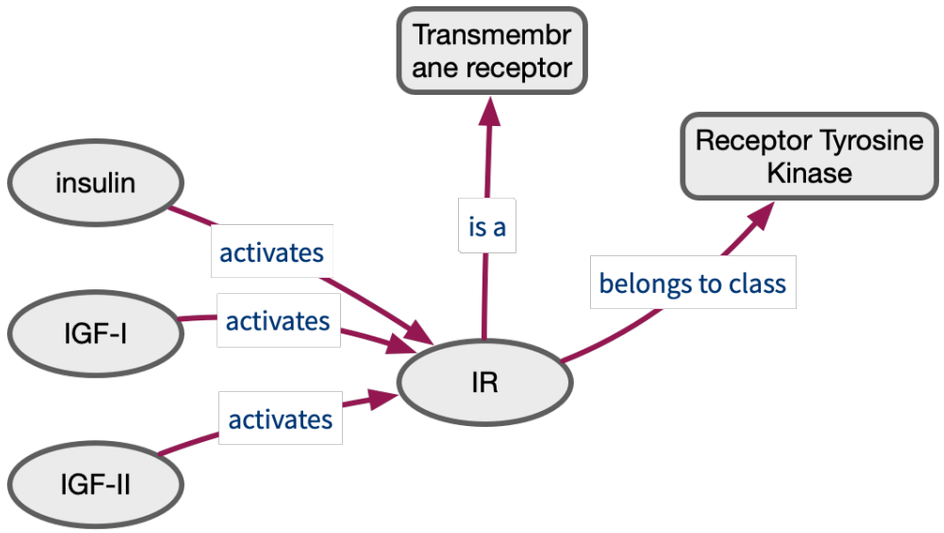
Hands-on session: from text to KG ✏️

Question #2

From wikipedia : *"The insulin receptor (IR) is a transmembrane receptor that is activated by insulin, IGF-I, IGF-II and belongs to the large class of receptor tyrosine kinase."*

Translate your KG into **RDF triples.**

In practice …




```
@prefix ns: <http://my/namespace/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

ns:insulin   ns:activates    ns:IR .
ns:IGF_I     ns:activates    ns:IR .
ns:IGF_II    ns:activates    ns:IR .

ns:IR   rdf:type  ns:TransmembraneReceptor ;
        ns:belongs_to_class ns:ReceptorTyrosineKinase .
```

Hands-on session: from text to KG ✏️

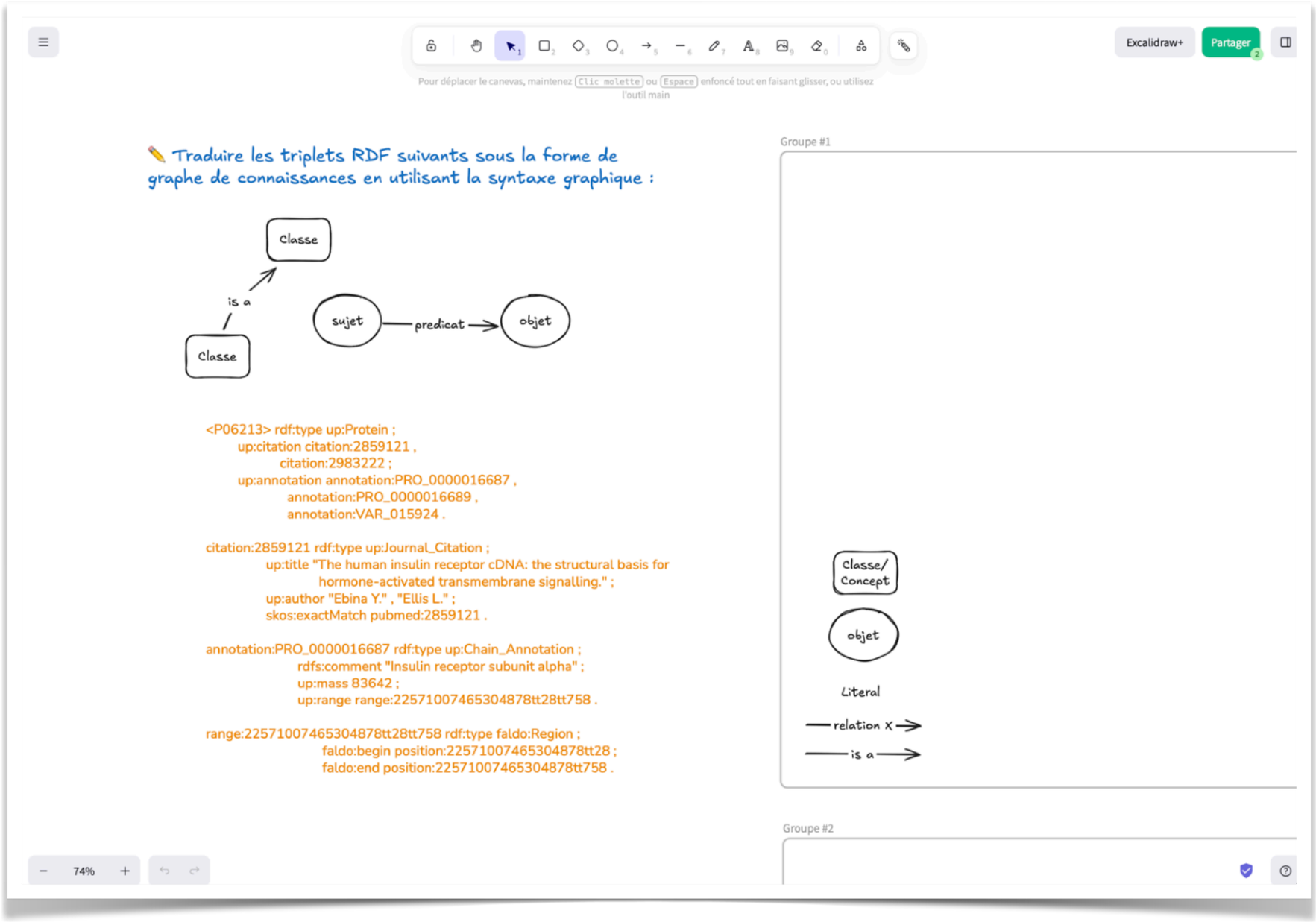
https://rest.uniprot.org/uniprotkb/P06213.ttl

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix up: <http://purl.uniprot.org/core/> .
@prefix annotation: <http://purl.uniprot.org/annotation/> .
@prefix citation: <http://purl.uniprot.org/citations/> .
@prefix range: <http://purl.uniprot.org/range/> .
@prefix faldo: <http://biohackathon.org/resource/faldo#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix pubmed: <http://purl.uniprot.org/pubmed/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix position: <http://purl.uniprot.org/position/> .

<P06213> rdf:type up:Protein ;
         up:citation citation:2859121 ,
                     citation:2983222 ;
         up:annotation annotation:PRO_0000016687 ,
                       annotation:PRO_0000016689 ,
                       annotation:VAR_015924 .

citation:2859121 rdf:type up:Journal_Citation ;
                 up:title "The human insulin receptor cDNA: the structural basis for
hormone-activated transmembrane signalling." ;
                 up:author "Ebina Y." , "Ellis L." ;
                 skos:exactMatch pubmed:2859121 .

annotation:PRO_0000016687 rdf:type up:Chain_Annotation ;
                          rdfs:comment "Insulin receptor subunit alpha" ;
                          up:mass 83642 ;
                          up:range range:22571007465304878tt28tt758 .

range:22571007465304878tt28tt758 rdf:type faldo:Region ;
                                 faldo:begin position:22571007465304878tt28 ;
                                 faldo:end position:22571007465304878tt758 .
```

Question #3

Draft the knowledge graph associated to some of the RDF triples representing the P06213 Uniprot entity.

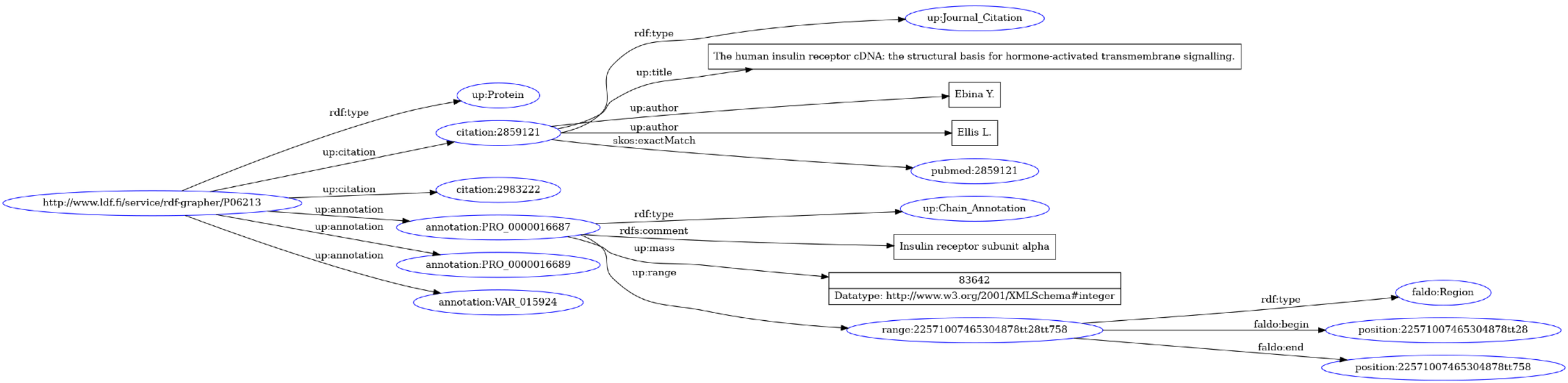
"Pen & paper" team work



https://tinyurl.com/4xf83nxd

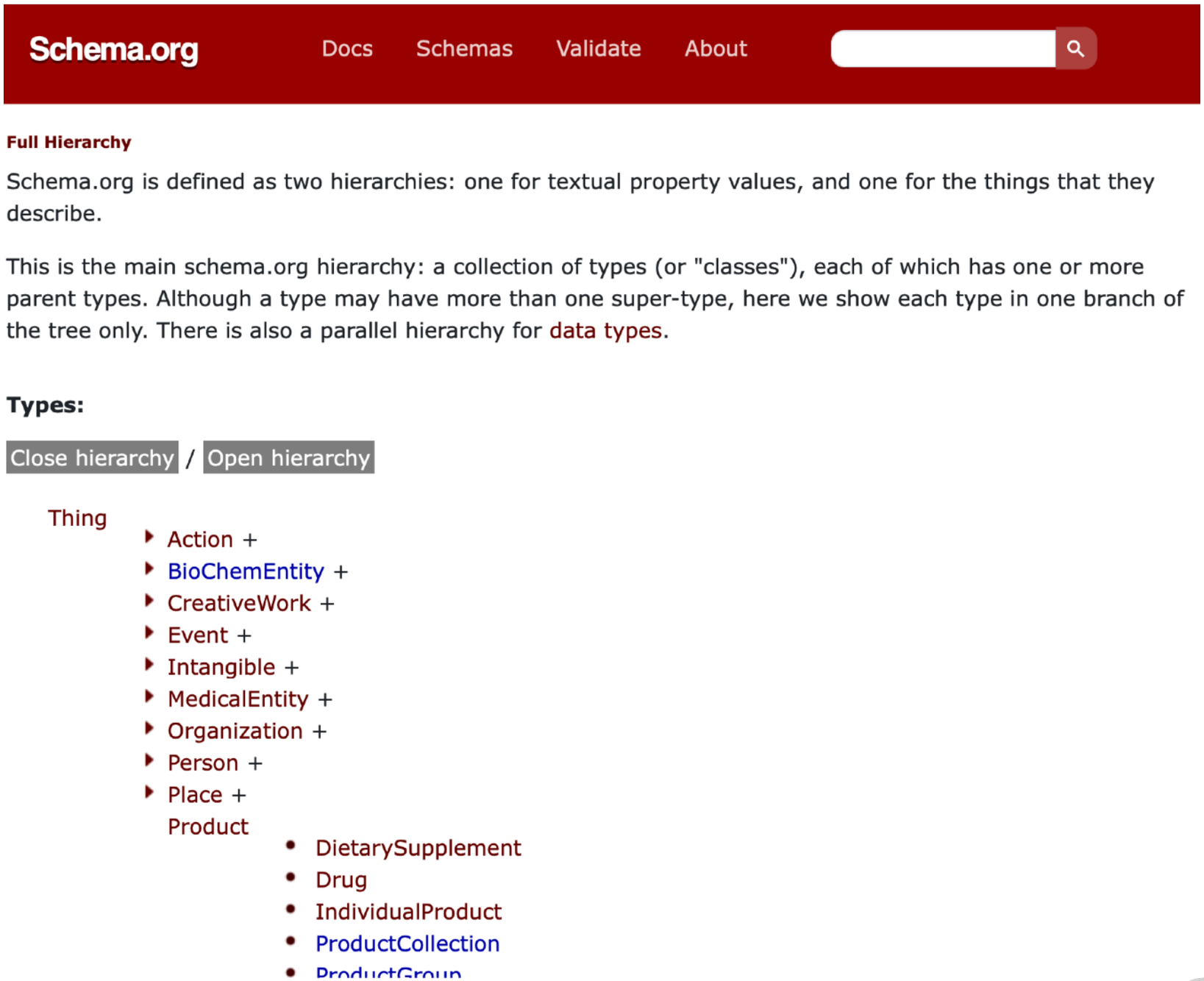
Practice … from KG to text

https://www.ldf.fi/service/rdf-grapher



Ontologies for FAIR bioinformatics tools and workflows

schema.org



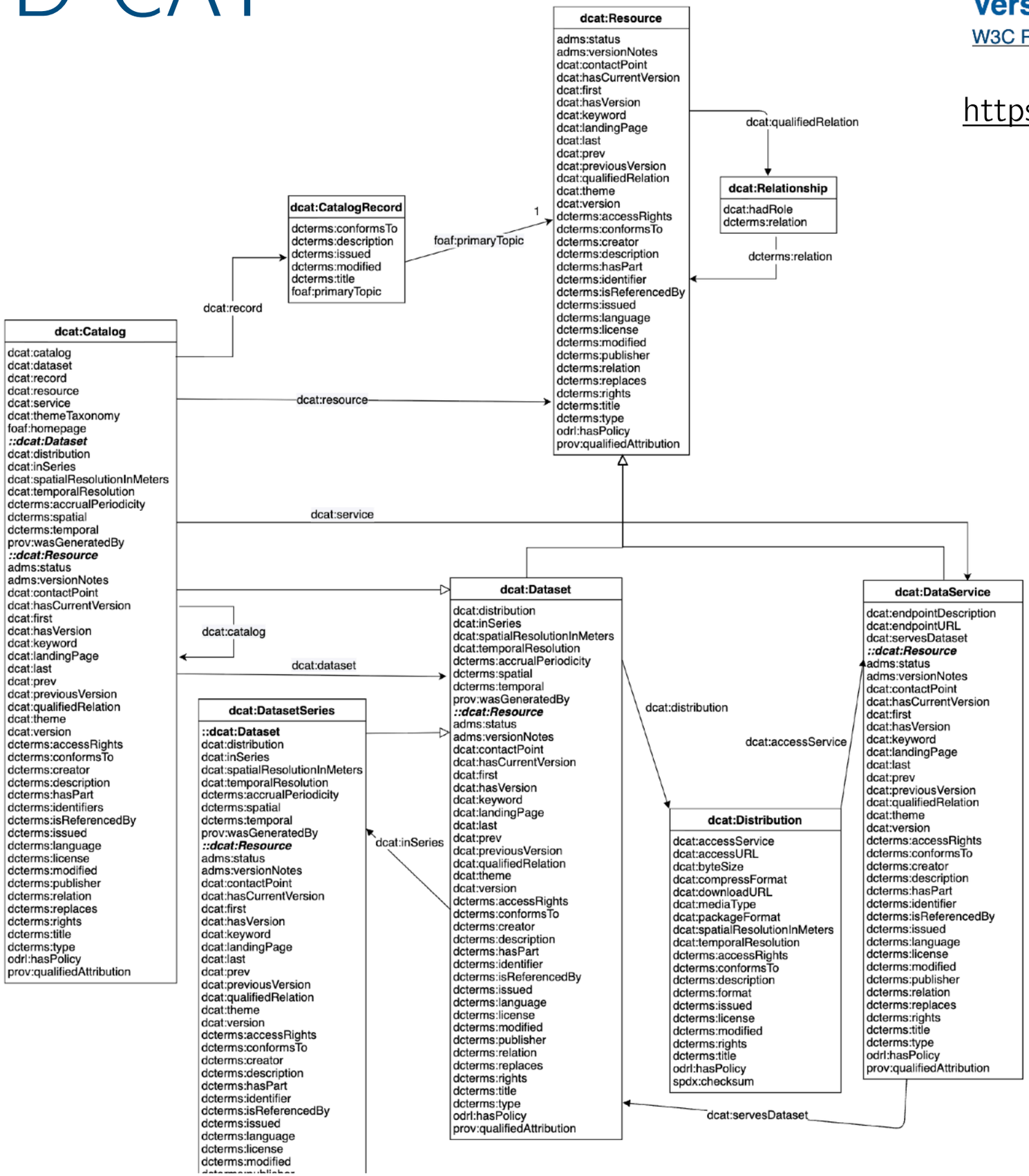
- General purpose **lightweight** ontology
- Aimed at **annotating web pages**
- Targetting **FINDABILITY**
- Originating from major search engines

Dublin Core (aka DC-Terms)


```
dcterms:Box
    dcterms:issued "2000-07-11"^^<http://www.w3.org/2001/XMLSchema#date> ;
    a rdfs:Datatype ;
    rdfs:comment "The set of regions in space defined by their geographic coordinates according to the DCMI Box Encoding Scheme."@en ;
    rdfs:isDefinedBy <http://purl.org/dc/terms/> ;
    rdfs:label "DCMI Box"@en ;
    rdfs:seeAlso <https://www.dublincore.org/specifications/dublin-core/dcmi-box/> .

dcterms:DCMIType
    dcterms:issued "2000-07-11"^^<http://www.w3.org/2001/XMLSchema#date> ;
    a dcam:VocabularyEncodingScheme ;
    rdfs:comment "The set of classes specified by the DCMI Type Vocabulary, used to categorize the nature or genre of the resource."@en ;
    rdfs:isDefinedBy <http://purl.org/dc/terms/> ;
    rdfs:label "DCMI Type Vocabulary"@en ;
    rdfs:seeAlso <http://purl.org/dc/dcmitype/> .

dcterms:DDC
    dcterms:issued "2000-07-11"^^<http://www.w3.org/2001/XMLSchema#date> ;
    a dcam:VocabularyEncodingScheme ;
    rdfs:comment "The set of conceptual resources specified by the Dewey Decimal Classification."@en ;
    rdfs:isDefinedBy <http://purl.org/dc/terms/> ;
    rdfs:label "DDC"@en ;
    rdfs:seeAlso <http://www.oclc.org/dewey/> .

dcterms:FileFormat
    dcterms:issued "2008-01-14"^^<http://www.w3.org/2001/XMLSchema#date> ;
    a rdfs:Class ;
    rdfs:comment "A digital resource format."@en ;
    rdfs:isDefinedBy <http://purl.org/dc/terms/> ;
    rdfs:label "File Format"@en ;
    rdfs:subClassOf dcterms:MediaType .

dcterms:Frequency
    dcterms:issued "2008-01-14"^^<http://www.w3.org/2001/XMLSchema#date> ;
    a rdfs:Class ;
    rdfs:comment "A rate at which something recurs."@en ;
    rdfs:isDefinedBy <http://purl.org/dc/terms/> ;
    rdfs:label "Frequency"@en .

dcterms:IMT
    dcterms:issued "2000-07-11"^^<http://www.w3.org/2001/XMLSchema#date> ;
    a dcam:VocabularyEncodingScheme ;
    rdfs:comment "The set of media types specified by the Internet Assigned Numbers Authority."@en ;
    rdfs:isDefinedBy <http://purl.org/dc/terms/> ;
    rdfs:label "IMT"@en ;
    rdfs:seeAlso <http://www.iana.org/assignments/media-types/> .

dcterms:ISO3166
    dcterms:issued "2000-07-11"^^<http://www.w3.org/2001/XMLSchema#date> ;
    a rdfs:Datatype ;
    rdfs:comment "The set of codes listed in ISO 3166-1 for the representation of names of countries."@en ;
    rdfs:isDefinedBy <http://purl.org/dc/terms/> ;
    rdfs:label "ISO 3166"@en ;
    rdfs:seeAlso <https://www.iso.org/obp/ui/#search> .
```

- **lightweight** ontology
- Generic metadata
- 22 classes
- 55 properties

D-CAT



Data Catalog Vocabulary (DCAT) - Version 3

W3C Recommendation 22 August 2024

https://www.w3.org/TR/vocab-dcat-3/

European Union

EU Vocabularies

Search

Publications Office | EU law | European data | EU tenders | EU research results | EU Whoiswh

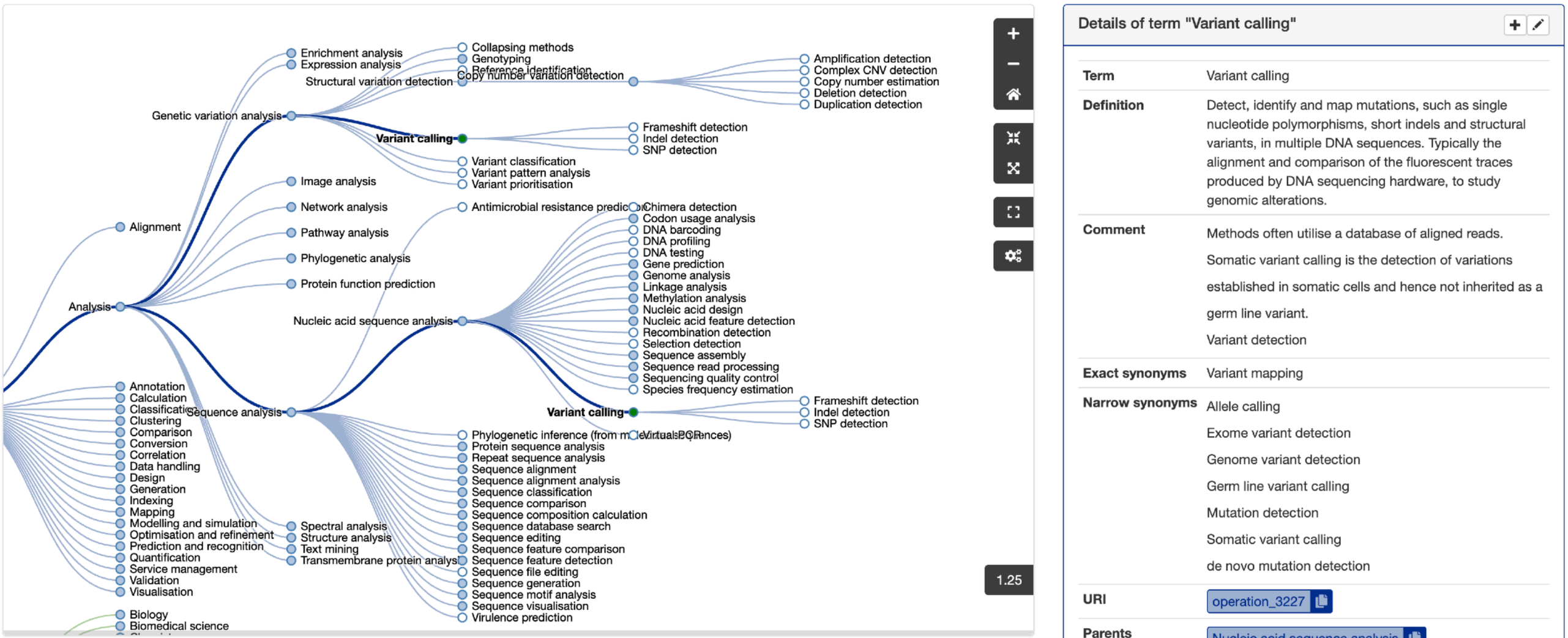
Publications Office › EU Vocabularies › Business collections › DCAT-AP for data portals in Europe

Home | Controlled vocabularies | Models | Business collections | Online tools

DCAT Application profile for data portals in Europe (DCAT-AP)

FAIR Data Point

Working Draft, 20 October 2023

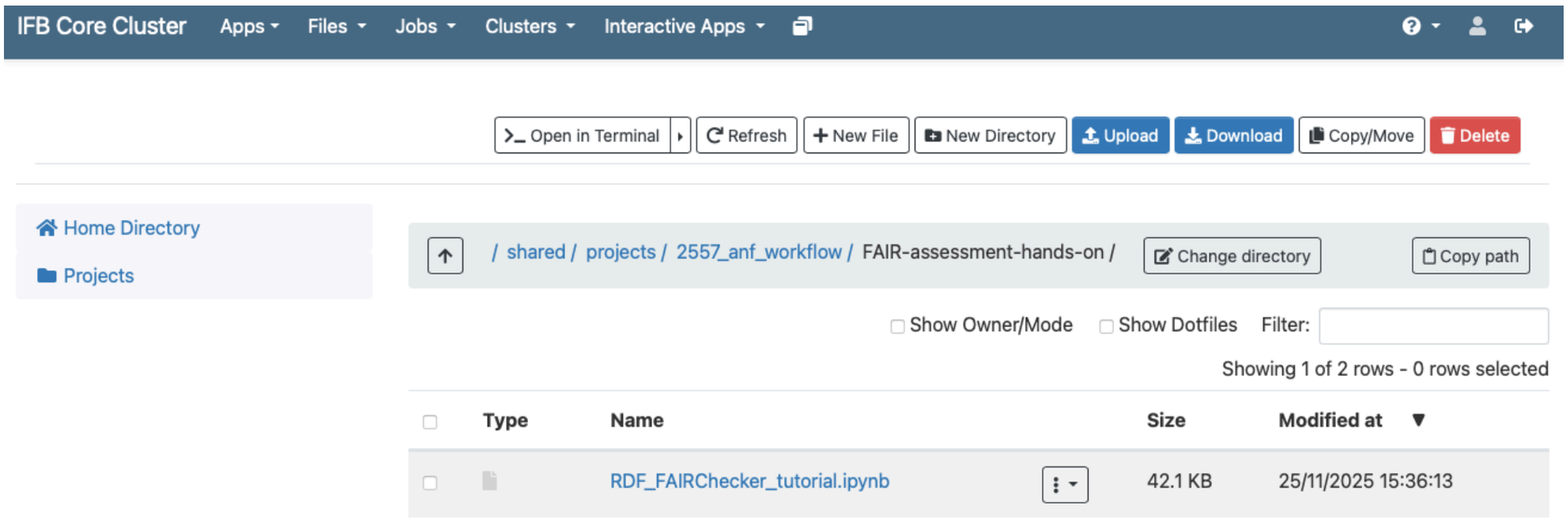


- A domain-specific ontology for Bioinformatics for annotating
  - **usage context** of bioinformatics tools (*EDAM Topics*)
  - **what does** the tool (*EDAM Operations*)
  - **input and output data** (*EDAM Data*, *EDAM Formats*)

Hands-on session: RDF graphs in Python ✏️

Question #4

Run step-by-step the PART1 section of the Jupyter notebook available on the IFB cluster :



Recap!

A practical overview on FAIR assessment, in 2 hours ...

- 📚 **Quick intro on FAIR principles**
- **Knowledge graphs and semantic metadata**
  - 📚 What are Knowledge Graphs, Computational Ontologies ?
  - ✏️ How to create Knowledge Graphs ? (team work, "pen & paper")
  - ✏️ How to write machine-readable semantic metadata ? (demo, python code)
  - ✏️ How to query knowledge graphs with SPARQL (demo, python code)
- **Tools and resources for increased FAIRness**
  - 📚 FAIR-Checker
  - 📚 Tools and workflow registries
  - ✏️ Using FAIR-Checker and interpreting the FAIR assement results
  - ✏️ Using the FAIR-Checker API and processing multiple resources

A practical overview on FAIR assessment, in 2 hours ...

- 📚 **Quick intro on FAIR principles**
- **Knowledge graphs and semantic metadata**
  - 📚 What are Knowledge Graphs, Computational Ontologies ?
  - ✏️ How to create Knowledge Graphs ? (team work, "pen & paper")
  - ✏️ How to write machine-readable semantic metadata ? (demo, python code)
  - ✏️ How to query knowledge graphs with SPARQL (demo, python code)
- **Tools and resources for increased FAIRness**
  - 📚 FAIR-Checker
  - 📚 Tools and workflow registries
  - ✏️ Using FAIR-Checker and interpreting the FAIR assement results
  - ✏️ Using the FAIR-Checker API and processing multiple resources

Coffee break