Webinar – 15th- 16th December 2020

Using ontologies to improve animal science research

Organizing center: INRAE (France)

Target participants: Scientists, RI technicians, Bachelor, Master or PhD students

Objectives

- Understand the usefulness of the ontologies to annotate experimental dataset
- Use these ontologies and reference them in publications
- Use the ontologies adapted to SmartCow

Pre requisite skills: Basic knowledge in biology

Program

Tuesday 15 December 2020, 9:45 – 15:30

9:45-10:00 Introduction of the course by R. Baumont and C. Hurtaud, INRAE (France)

10:00 -12:00 What is an ontology? by Olivier Dameron, University of Rennes 1 (France)

1. Definition
2. History
3. Examples of use of ontologies (application of ontologies)
4. Exercises:
    i. How to build a small ontology?
    ii. How to use ontologies to annotate data with the open-source editor Protégé?
    iii. How to use ontologies to retrieve data with the open-source editor Protégé?

12:00 – 13:30 lunchbreak

13:30-15:30 The ontology engineering lifecycle and its best practices by Monika Solanki, Agrimetrics (UK)

Wednesday 16 December, 10:00 -12:00

1. Presentation of French ontologies from INRAE
    a. ATOL ontology by Catherine Hurtaud and Linh Chi Nguyen, INRAE (France) (25 min)
    b. EOL ontology by Jérôme Bugeon, INRAE (France) (15 min)
    c. AHOL ontology by Marie-Christine Meunier-Salaün, INRAE (France) (15 min)
2. Example of application/use of an ontology AHOL for professional use, by Marie-Christine Meunier-Salaün, INRAE (France) (20 min)

3. Text based annotation with ontologies, Claire Nédellec INRAE (France) (45 min)

SmartCow

an integrated infrastructure for increased research capability and innovation in the European cattle sector

Webinar

Using ontologies to improve animal science research

Dr. René Baumont (INRAE, Project Coordinator)

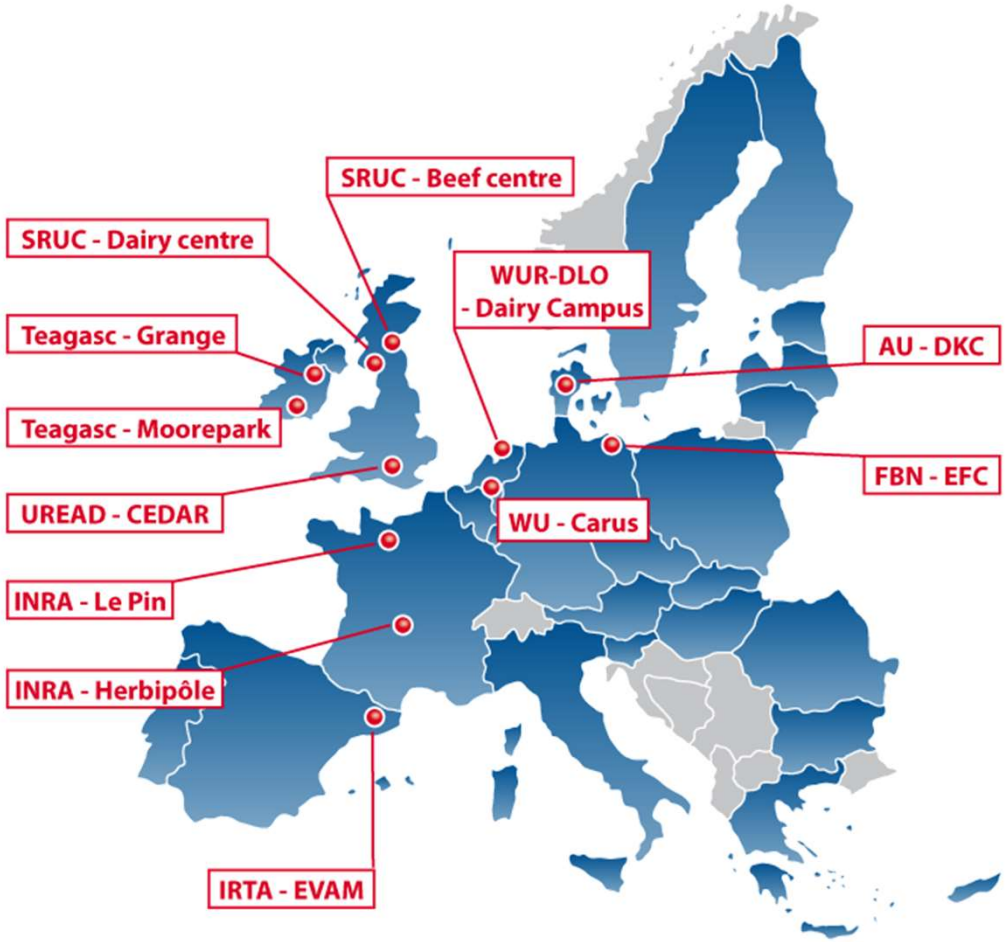
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n°730924

Concept and objective of SmartCow

- Better coordination of research infrastructures (RIs) in the cattle sector is necessary to develop more efficient approaches to address the various challenges in cattle breeding and research
- SmartCow is a first step towards the integration of RIs for the European cattle sector, developing:
  - ***Networking activities***: mapping RIs, adopting common language, the best standardized techniques and data sharing
  - ***Joint research activities***: improving GSM and 'proxies' to phenotype new and more complex animal traits
  - ***Transnational access:*** providing researchers an easy access to high quality services and resources

SmartCow at a glance

2018 - 2022



First-class Cattle Research Infrastructures (RIs) across Europe:

- 11 major RIs distributed in 7 EU countries
- 12 locations, which include 18 installations
- 2500 dairy and 1000 beef cows

- **Networking of RIs** to inventorize resources, harmonize procedures, and share data
- **Joint research activities** to improve experimental methods and phenotyping capability
- **Interaction with stakeholders** to stay in line with industry needs and improve dissemination

http://www.smartcow.eu/stakeholders/

TRAINING PROGRAM

For Scientists, Technicians, Stakeholders, PhD students

- Face-to-face training courses
- Free web-conferences
- One-day study tours in 4 different countries

http://www.smartcow.eu/resources/training/

TRANSNATIONAL ACCESS CALLS

Offers external users (academic and industry) free access to SmartCow RIs

- 30 projects during the 4 years of SmartCow
- Access to around 10,000 cow-weeks

http://www.smartcow.eu/calls/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n°730924.

SmartCow

A training and study tour program

It consists of:

- Face-to-face training courses (1-3 days)
- Web conference training courses, accessible with PC, Internet + phone (2 hours)
- One-day study tours

It aims to share the results of the project : methods, techniques, use of tools…

- *CH4 emissions, Feed efficiency, Ethics in animal experimentation, NIRS, Sensors, Ontologies…*

It includes lectures, exchanges with scientists, visit of RI facilities and farms…

…But Covid-19 crisis obliged turning into webinar all training activities since March 2020…

Research infrastructures in SmartCow

- **Large range of data are obtained on animals, feeds and environment in various conditions**









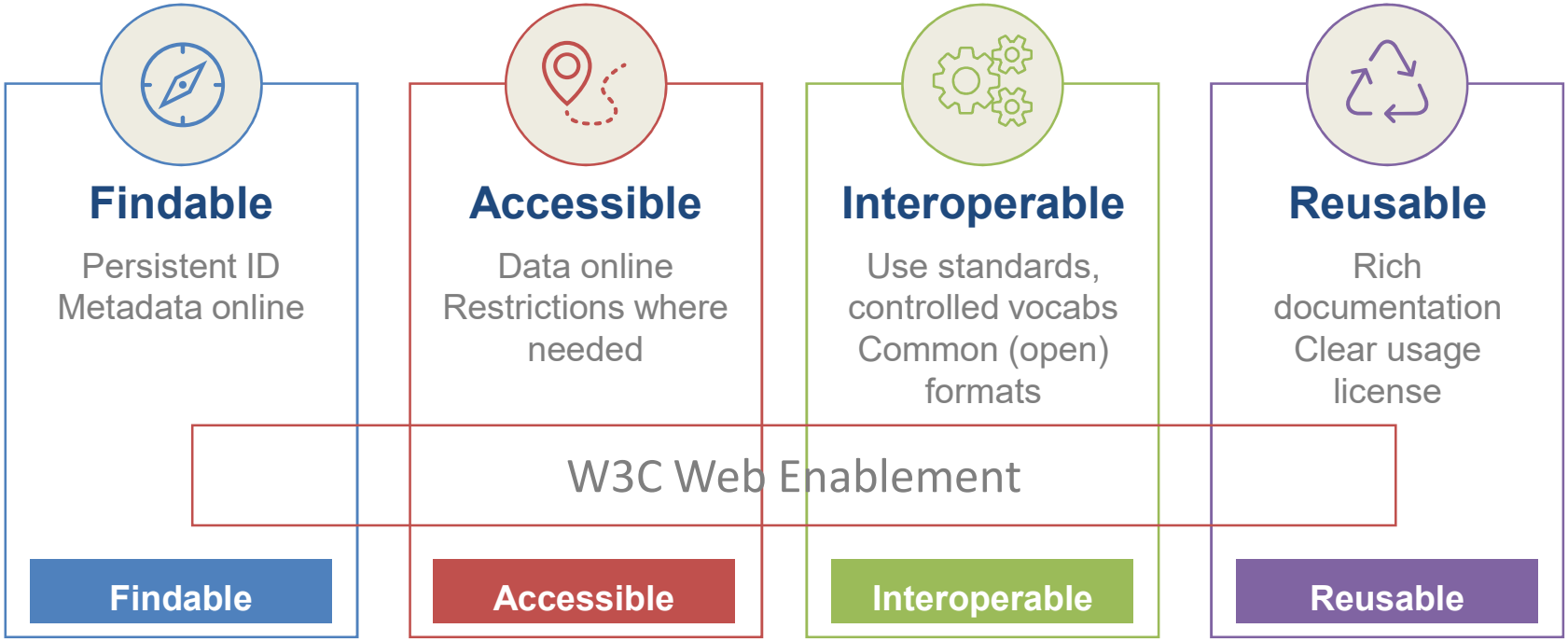




SmartCow

FAIR Principles in SmartCow

- Principles to ensure that data are shared in a way that enables & enhances reuse, by humans and machines



Source: https://doi.org/10.5281/zenodo.1065991

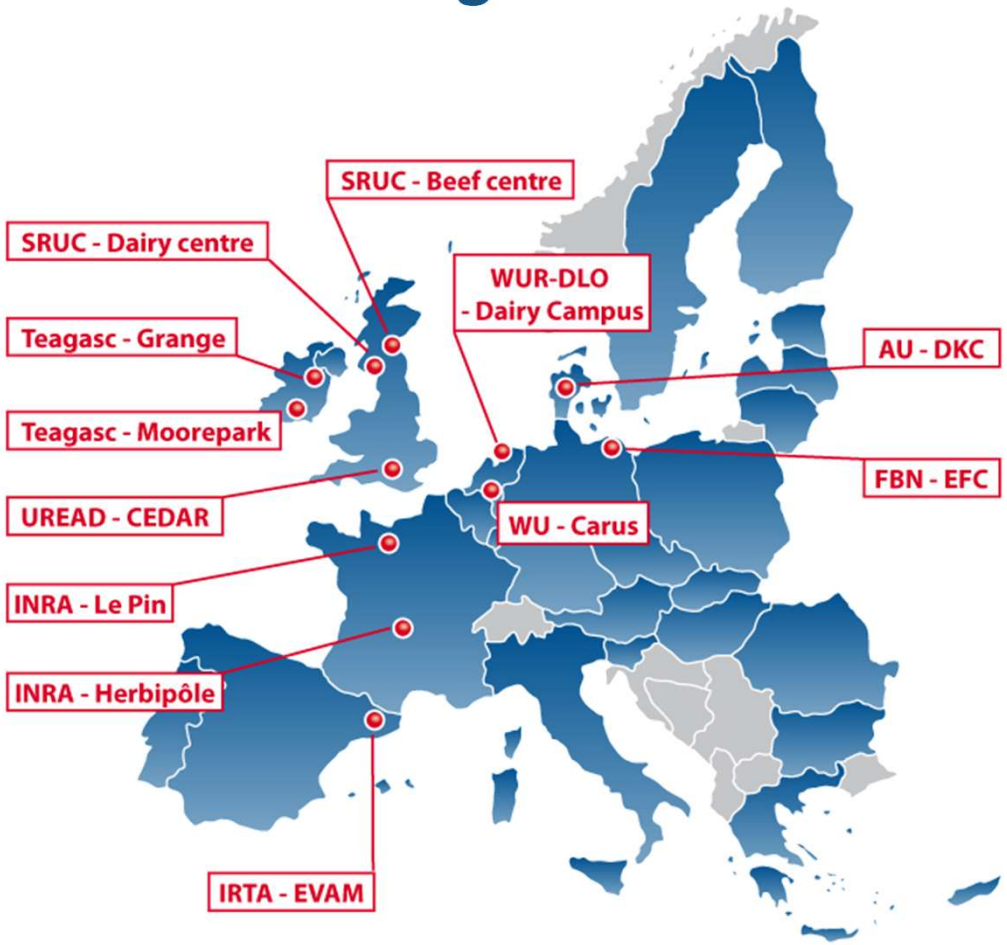
SmartCow

Ontologies in SmartCow

- The DMP outlines the use of a **common language to describe data** and the development of a tool that allow to exploit data in accordance with FAIR principles:
  - Use of open standards and open **data ontologies** to simplify access and inter-operability of data
    - ➔ *Improve and foster the use of ATOL, EOL and AHOL*
    - ➔ *92 new terms added since the beginning of SmartCow*
  - Development of a **could-based data platform** to upload, control sharing and link the data to get additional insights and store the data beyond the life of the project.
  - **Standardize the vocabulary** and make links between how observations are described and **livestock ontologies**

Thank you for your attention

SmartCow at a glance



First-class Cattle Research Infrastructures (RIs) across Europe:

- 11 major RIs distributed in 7 EU countries
- 12 locations, which include 18 installations
- 2500 dairy and 1000 beef cows

- **Networking of RIs** to inventorize resources, harmonize procedures, and share data
- **Joint research activities** to improve experimental methods and phenotyping capability
- **Interaction with stakeholders** to stay in line with industry needs and improve dissemination

http://www.smartcow.eu/stakeholders/

TRAINING PROGRAM

For Scientists, Technicians, Stakeholders, PhD students

- Face-to-face training courses
- Free web-conferences
- One-day study tours in 4 different countries

http://www.smartcow.eu/resources/training/

TRANSNATIONAL ACCESS CALLS

Offers external users (academic and industry) free access to SmartCow RIs

- 30 projects during the 4 years of SmartCow
- Access to around 10,000 cow-weeks

http://www.smartcow.eu/calls/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n°730924.

SmartCow

SmartCow

an integrated infrastructure for increased research capability and innovation in the European cattle sector

Webinar:

Using ontologies to improve animal science research

15-16 December 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n°730924

Program of the webinar

Tuesday 15 December morning 9:45-12:00

- 9:45-10:00 Introduction of the course by R. Baumont and C. Hurtaud, INRAE (France)
- 10:00 -12:00 What is an ontology? by Olivier Dameron, University of Rennes 1 (France)
    1. Definition
    2. History
    3. Examples of use of ontologies (application of ontologies)
    4. Exercises:
        i. How to build a small ontology?
        ii. How to use ontologies to annotate data with the open-source editor Protégé?
        iii. How to use ontologies to retrieve data with the open-source editor Protégé?

- 12:00 – 13:30 Lunchbreak

SmartCow

Program of the webinar

Tuesday 15 December afternoon

- 13:30-15:30 The ontology engineering lifecycle and its best practices by Monika Solanki, Agrimetrics (UK)

  Part 1 - 45 minutes presentation, Q&A and a 5 minutes break

  Part 2- 45 minutes presentation, hands-on & Q&A

Program of the webinar

Wednesday 16 December morning 10:00-12:00

- 10:00-10:55 Presentation of French ontologies from INRAE
  a. ATOL ontology by Catherine Hurtaud, INRAE (France) (25 min)
  b. EOL ontology by Jérôme Bugeon, INRAE (France) (15 min)
  c. AHOL ontology by Marie-Christine Meunier-Salaün, INRAE (France) (15 min)

- 10:55-11:15 Example of application/use of an ontology AHOL for professional use, by Marie-Christine Meunier-Salaün, INRAE (France) (20 min)

- 11:15-12:00 Text based annotation with ontologies, Claire Nédellec INRAE (France) (45 min)

SmartCow

During the webinar

- If you have a question, please use the chat.
- Linh Chi and I will be moderators for the questions. We will ask the questions for you during or after the presentations

Have a good webinar from :

Rennes, France

Paris, France

Reading, Great Britain

SmartCow

What is an ontology?

Olivier Dameron

Université de Rennes 1

15 December 2020

"Ontologies make people happier"
Ph. Rocca-Serra, 2011-10-18

Life science data

Big data and the deluge of life science data

Big data

Datasets so **large** or **complex** that traditional data processing is inadequate [Laney2001]

Life science: data deluge [Aldhous1993]

- computerized biomedical data
- genomics and bioinformatics

Science. 1993 Oct 22;262(5133):502-3.

Managing the genome data deluge.

Aldhous P.

PMID: 8211171 [PubMed - indexed for MEDLINE]

Science. 1995 Aug 4;269(5224):630.

Europe opens institute to deal with gene data deluge.

Williams N.

PMID: 7624788 [PubMed - indexed for MEDLINE]

Bottleneck

Too much data for current processing capabilities

- data production rates outpace CPU improvements
- current analysis methods do not scale up

The Widening Gulf between Genomics Data Generation and Consumption: A Practical Guide to Big Data Transfer Technology

Frank A. Feltus[1], Joseph R. Breen III[2], Juan Deng[3], Ryan S. Izard[3], Christopher A. Konger[4], Walter B. Ligon III[3], Don Preuss[5] and Kuang-Ching Wang[3]

BIOINFORMATICS AND BIOLOGY INSIGHTS 2015:9(S1)

What to expect for 2025?

Our estimation is that genomics is a "four-headed beast" – it is either on par with or the most demanding domain [...] in terms of

- data acquisition
- data storage
- data distribution
- data analysis

Big Data: Astronomical or Genomical?

Zachary D. Stephens[1], Skylar Y. Lee[1], Faraz Faghri[2], Roy H. Campbell[2], Chengxiang Zhai[3], Miles J. Efron[4], Ravishankar Iyer[1], Michael C. Schatz[5]*, Saurabh Sinha[3]*, Gene E. Robinson[6]*

PLOS Biology | DOI:10.1371/journal.pbio.1002195 July 7, 2015

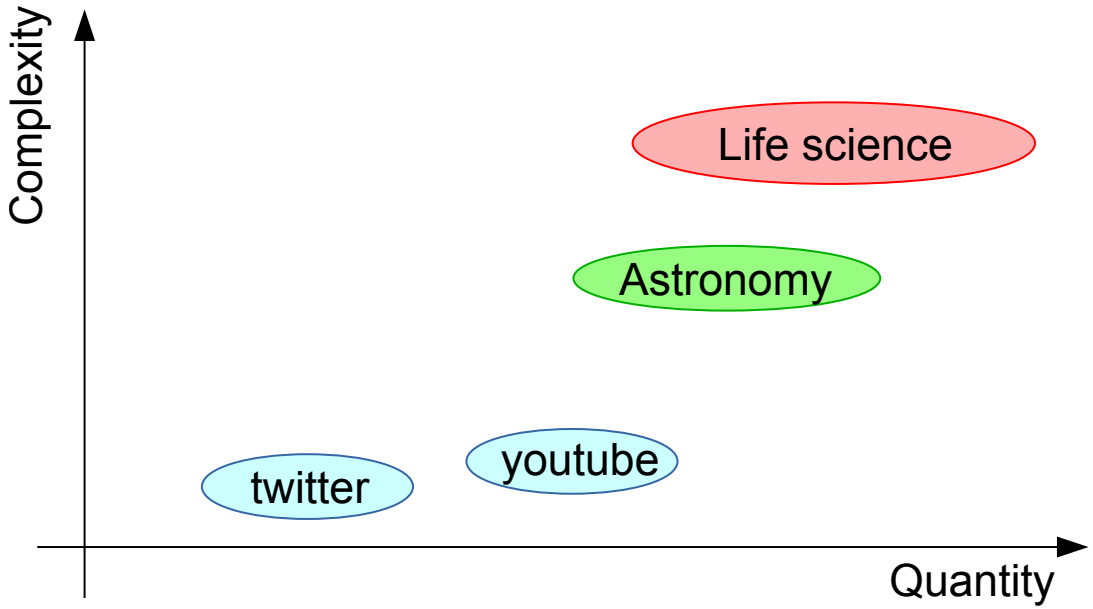
Table 1. Four domains of Big Data in 2025. In each of the four domains, the projected annual storage and computing needs are presented across the data lifecycle.

Data Phase	Astronomy	Twitter	YouTube	Genomics
Acquisition	25 zetta-bytes/year	0.5–15 billion tweets/year	500–900 million hours/year	1 zetta-bases/year
Storage	1 EB/year	1–17 PB/year	1–2 EB/year	2–40 EB/year
Analysis	In situ data reduction	Topic and sentiment mining	Limited requirements	Heterogeneous data and analysis
	Real-time processing	Metadata analysis		Variant calling, ~2 trillion central processing unit (CPU) hours
	Massive volumes			All-pairs genome alignments, ~10,000 trillion CPU hours
Distribution	Dedicated lines from antennae to server (600 TB/s)	Small units of distribution	Major component of modern user's bandwidth (10 MB/s)	Many small (10 MB/s) and fewer massive (10 TB/s) data movement

doi:10.1371/journal.pbio.1002195.t001

Degrees of data complexity

- multiple scales (heterogeneity)
- (highly) interdependent at each scale
- interdependent between scales
- variability
- incompleteness
- evolution
- distributed (and lack of interoperability)



Complexity of life science data

Challenge (computational)

How to handle this complexity?

- Experts are very good at doing it on their domain (hint)
  - on their domain
  - on their data
- The difficulty is to do it systematically
- Expertise = ability to use knowledge for interpreting data
- We should use their expertise, not try to outperform them

Annotations for describing complex and massive data

Capturing expertise with annotations

Annotation

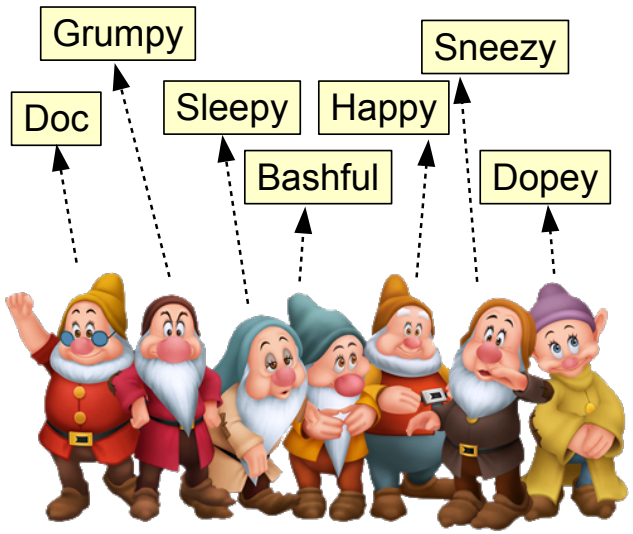
Annotation = result of some interpretation process

Capturing expertise with annotations

Annotation

Annotation = result of some interpretation process

- ideally by an expert (from big data to smart data)
- usually requires some background knowledge
- formalisation ranging from free text to controlled vocabularies to (shared) semantic framework [semantic spectrum]



The data annotation process

Annotate data =

- describe **explicitely** the relevant **elements** in your data and the **relations** between these elements
- so that users (incl. you and the non-experts) or programs do not have to go once again through the tedious process of interpreting them

- is important!
- seems easy...
- **exercice 1: describe a set of images**
- ... but turns out to be more difficult than expected
- describing and interpreting the data is for domain experts
- formalizing and representing the annotations and their dependencies is for data engineers

Using annotations for overcoming data complexity

Add annotations? But we have too much data already!

Benefits

- can be used as proxy to complex data
- simplifies by providing a compact abstraction
- overcomes variability
- enriches by making explicit the underlying meaning

Storing, sharing and reusing these annotations is the key to life science data systematic analysis

Ontologies for interpreting annotations

Ontologies

Knowledge underlying annotations remains to be represented

- "Much of biology works by applying prior knowledge [...] to an unknown entity" [Stevens2000]
- "The complex biological data stored in bioinformatics databases often require the addition of knowledge to specify and constrain the values held in that database" [Stevens2000]

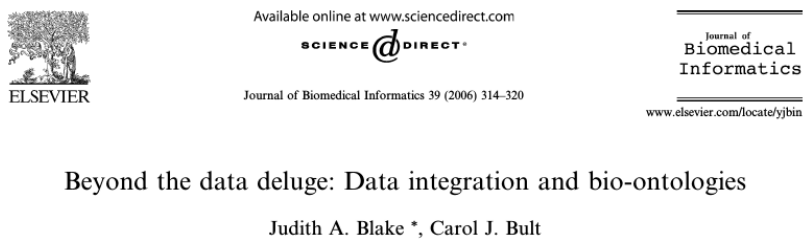
Ontology

Formal representation of knowledge in which the essential terms are combined with structuring rules that describe the relationships between them [Bard2004]


Brief Bioinform. 2000 Nov;1(4):398-414.

Ontology-based knowledge representation for bioinformatics.

Stevens R[1], Goble CA, Bechhofer S.


ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT

Journal of Biomedical Informatics 39 (2006) 314–320

Journal of Biomedical Informatics

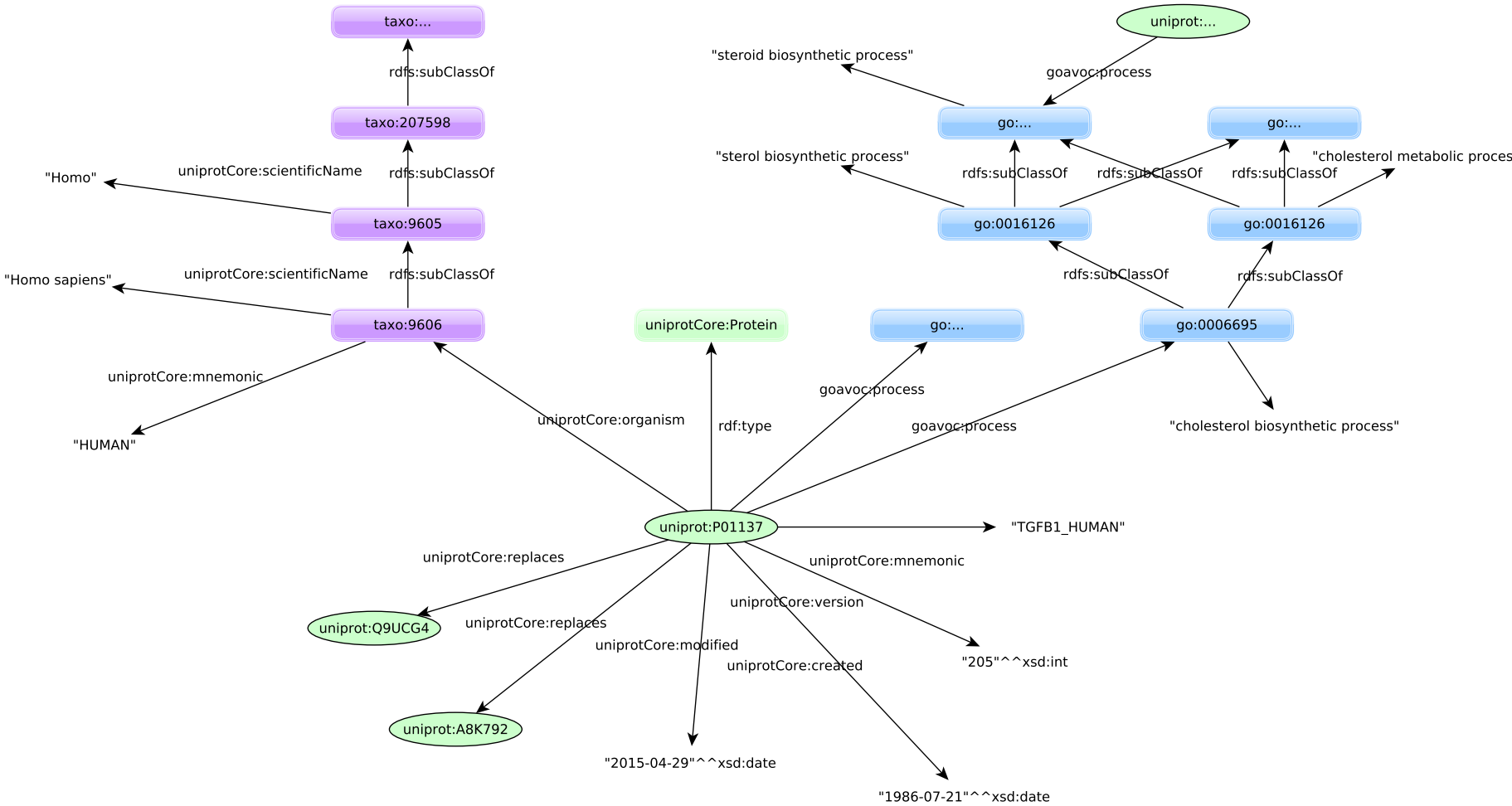
www.elsevier.com/locate/yjbin

Beyond the data deluge: Data integration and bio-ontologies

Judith A. Blake *, Carol J. Bult

Ontologies specify the meaning of annotations

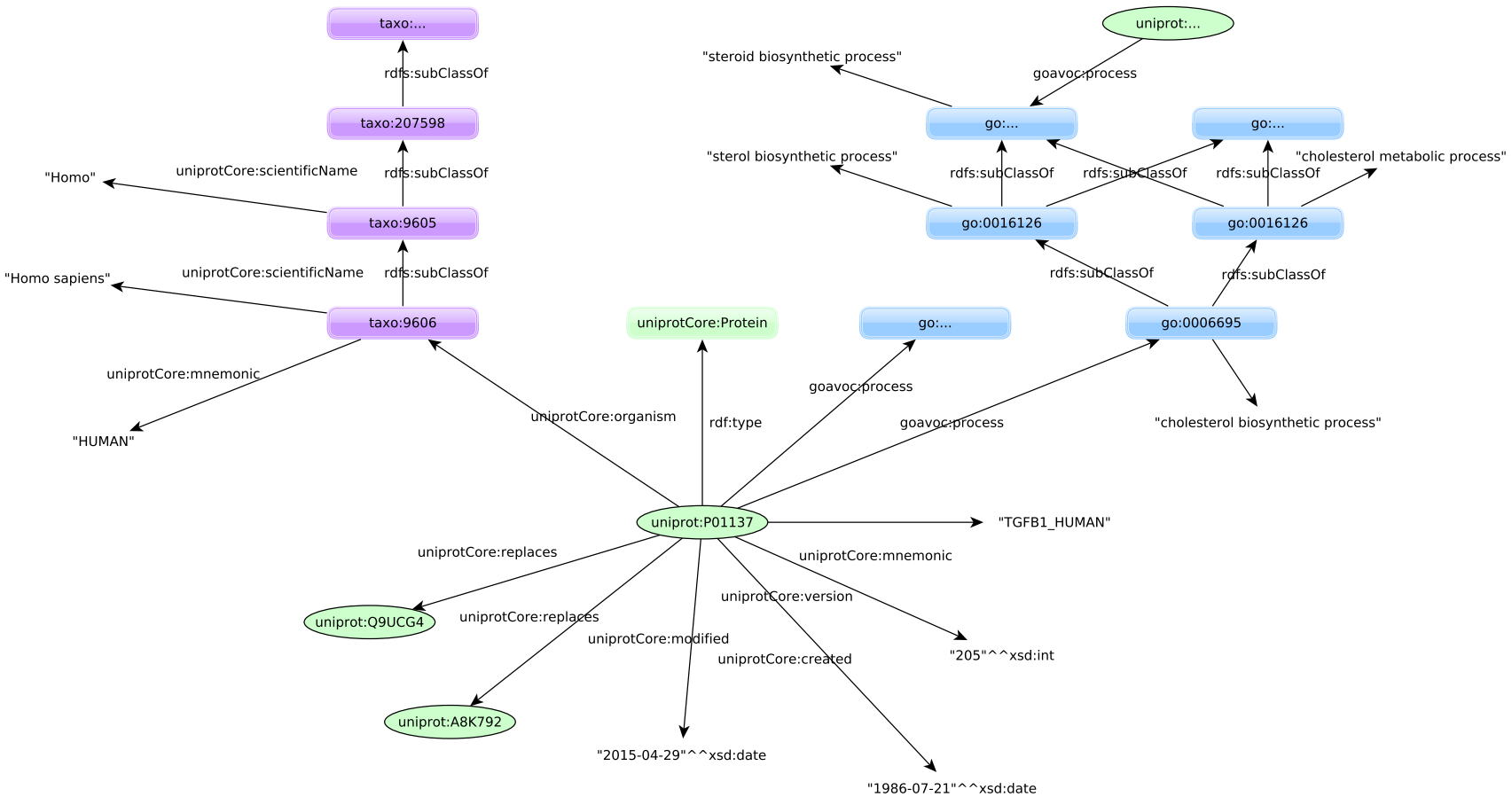
Knowledge is represented as relations between sets of entities



Ontologies support reasoning about annotations

Reasoning

Method for traversing or enriching the graph of data



Semantic Web and Linked (Open) Data

Semantic Web offers a unified framework to Linked Data

- **RDF** for representing and aggregating entities descriptions
- **RDFS+OWL** for representing domain knowledge (and combine it with data descriptions)
- **SPARQL** for querying everything (possibly from multiple repositories)

SPARQL endpoints offer unified query access to RDF repositories
ex: Fuseki, Virtuoso,...

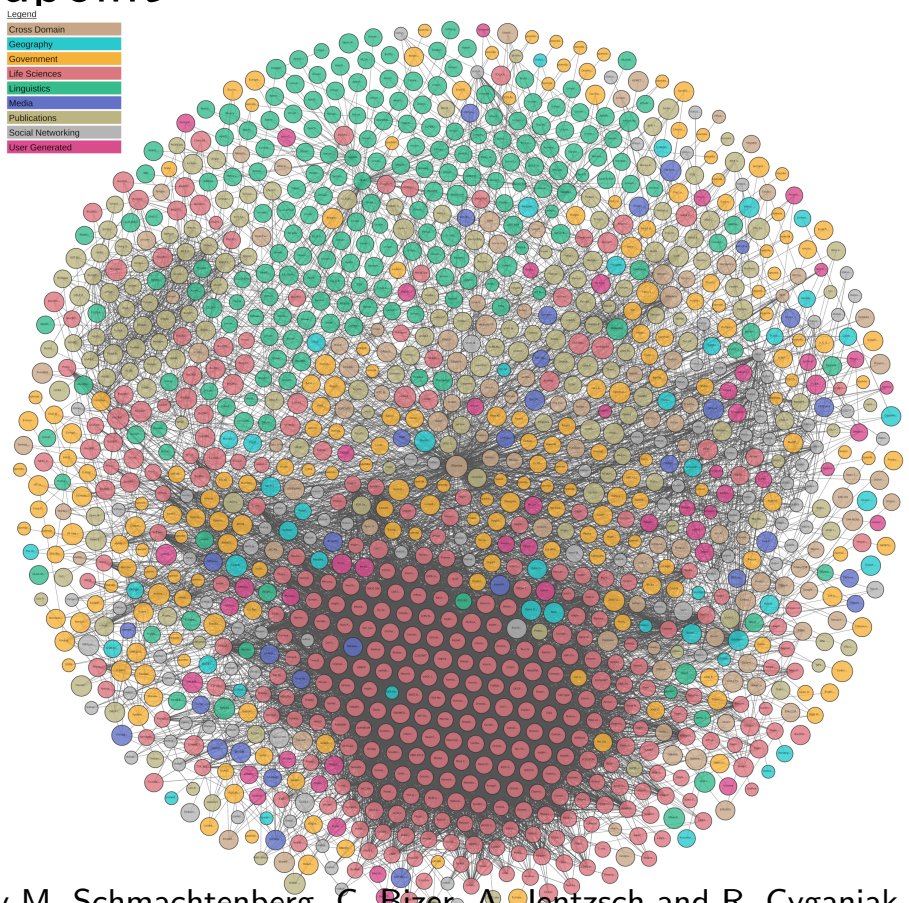
Linked Open Data: a federation of RDF repositories

LODStats (`http://lodstats.aksw.org/`) [Ermilov2016]

- 9960 datasets; 149.10^9 triples
- general scope; Life sciences = major field (size+density)

Linked open data (in 2019-03-29)

- RDF repositories can be queried in SPARQL via endpoints
- data from one endpoint can make references to data from another endpoint



Linked open data cloud, by M. Schmachtenberg, C. Bizer, A. Jentzsch and R. Cyganiak http://lod-cloud.net/

Ontologies: what they are

- Annotations' underlying structure
- (Formalized) representations of domain knowledge
- Relations that are **always** true, independently from anecdotic facts

Rule of thumb

what is implicit in a factual document is often background knowledge that belongs to an ontology

Ontologies: what they are not

- Ontology (the branch of philosophy)
- Controlled vocabularies, terminologies,... (although both are useful and often connected)
- Sets of annotated data (**genericity** is the key)

From controlled vocabularies to ontologies

- **Controlled vocabulary**: set of words used preferentially to ensure annotation uniformity
- **Terminology**: vocabulary with additional information (e.g. synonyms) and relations (e.g. antonyms, broader than, narrower than, related to,...). Ex: Wordnet
- **Taxonomy** (hopefully) formal organisation of concepts along the "is a" relation (subsumption). Ex: classification of species
  - language-independent
  - can be associated to controlled vocabularies and terminologies
- **Ontology**: taxonomy with additional relations, constraints, and (hopefully) formal definitions. Ex: Gene Ontology, ChEBI...

Ontology principles

Ontologies mostly deal with sets of entities (classes)

- Which properties are shared by these entities (necessary)
- How some of these properties define the set (necessary and sufficient)
- What are the possible values for some of the properties (fractures are located on bones, no "fracture of the eyelid")
- What are the relations with other sets of entities

- Classes are organized from general to particular (taxonomy)
- The subclass – superclass relation corresponds to set inclusion
- A class can have several parents
- The properties of the superclass also apply to all its descendents (inheritance)

Ontologies in use: use the MeSH to query pubmed

handle synonyms

In pubmed (`https://pubmed.ncbi.nlm.nih.gov/`)

- look for articles about vitamin c as a full text search
- examine one of the articles and look for the MeSH annotations
- look for the MeSH term vitamin c and the articles it annotates
- look for the MeSH term ascorbic acid and the articles it annotates

handle taxonomy

In MeSH (`https://www.ncbi.nlm.nih.gov/mesh/`)

- look for cardiovascular disease
- select the relevant MeSH term (`https://www.ncbi.nlm.nih.gov/mesh/68002318`)
- look at its synonyms and its descendents
- add it to the search builder
- search on pubmed

Ontologies in use: Gene Ontology

In the GO website (`http://geneontology.org/`)

- look for "glucose metabolic process"
- select "ontology" in the radio box below the search field, and search
- select the GO term (`http://amigo.geneontology.org/amigo/term/GO:0006006`)
- select the "Graph view" or the "Inferred tree view" tab and visualize its ancestors and its descendants
- for Homo sapiens, how many proteins are annotated by this GO term? How many miRNA?

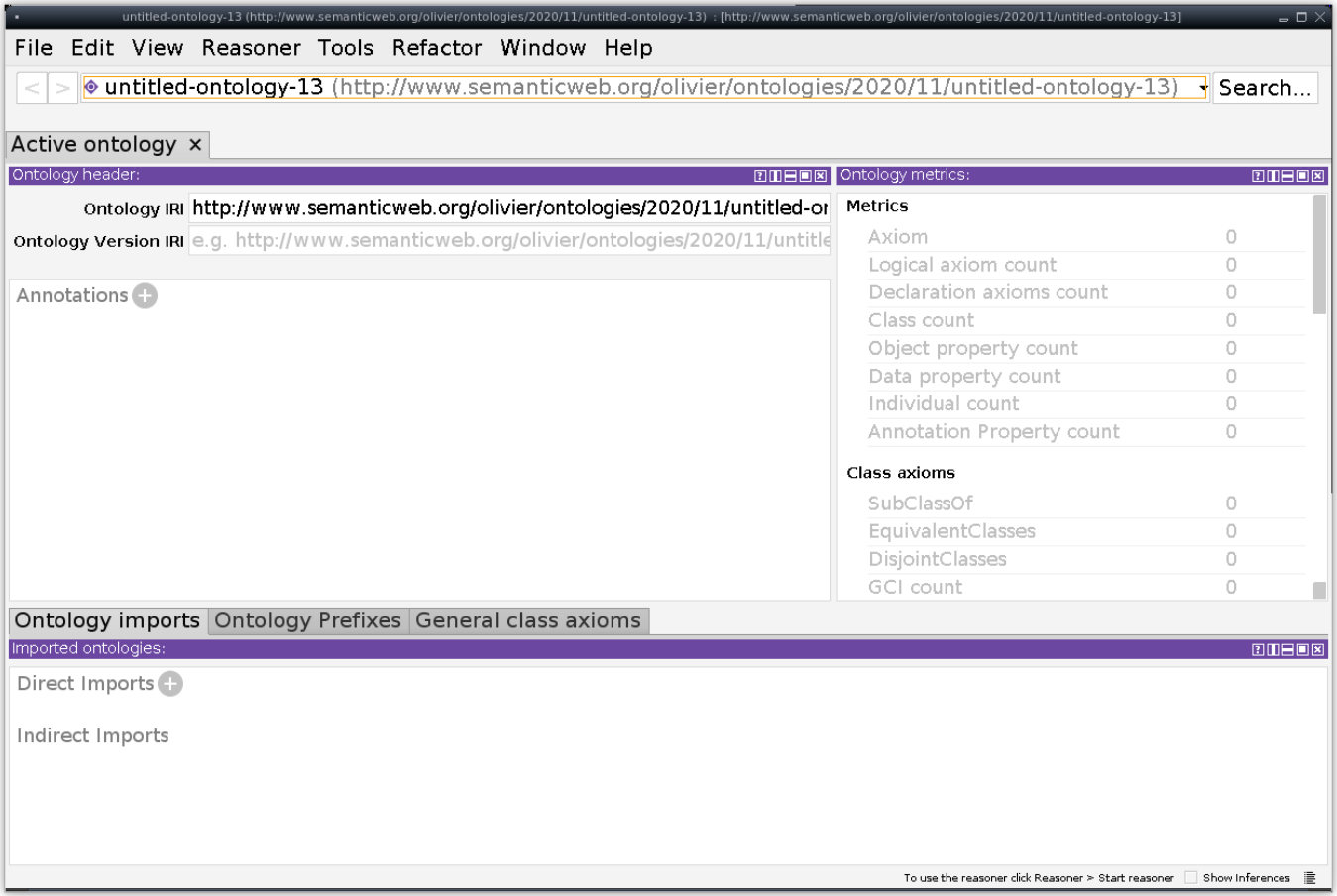
Synthesis

Synthesis

- Life science **data** are distributed and highly complex
- **Annotations** address complexity
- Ontologies formalize the **knowledge** underlying annotations
  - Reused across datasets
  - Shared for interoperability
  - Support reasoning
  - For programs more than humans
- the Semantic Web provides an infrastructure supporting integration and **reasoning**
- Life science annotations and ontologies have matured

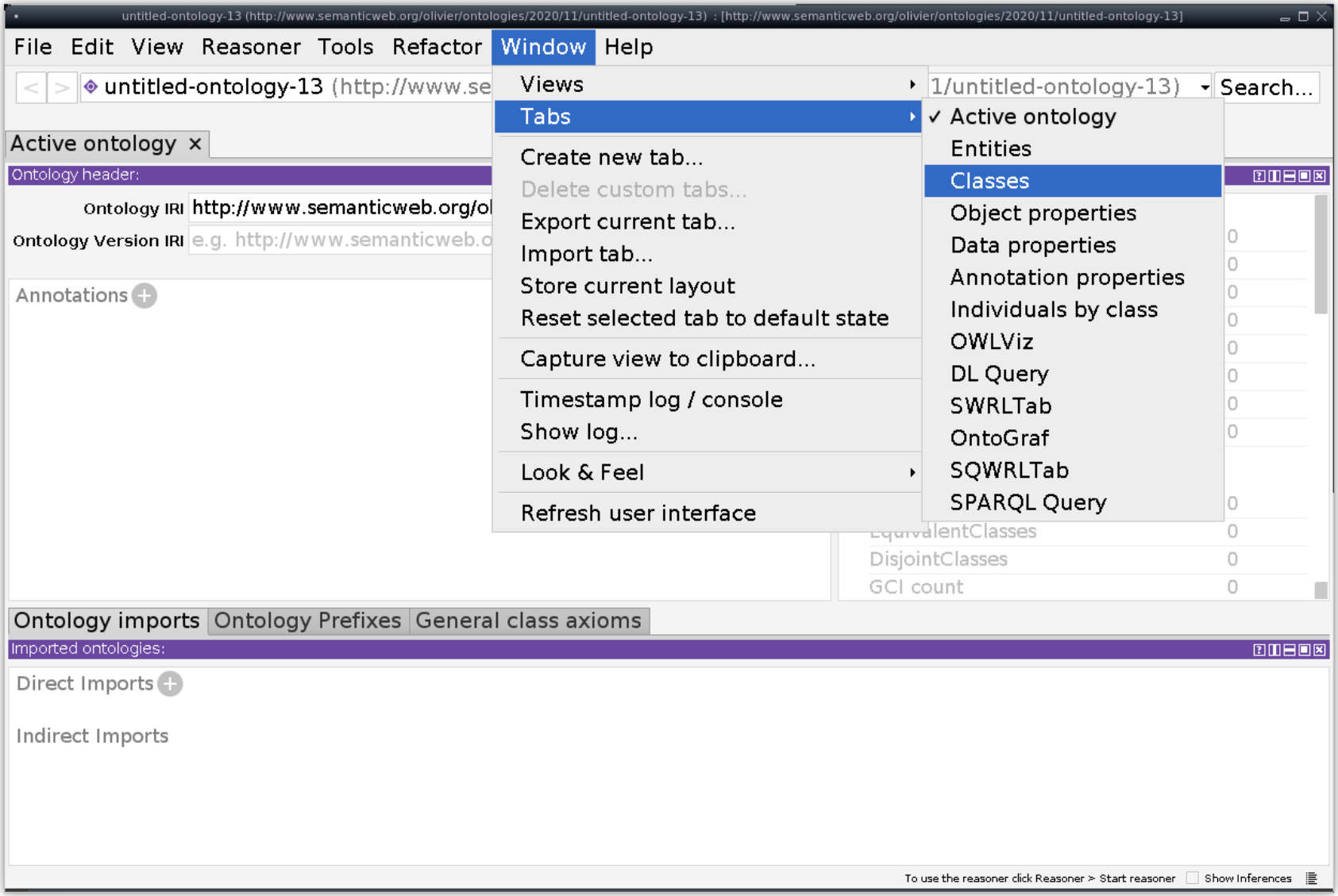
Hands-on session: build your own ontologies

Your first ontology with Protégé

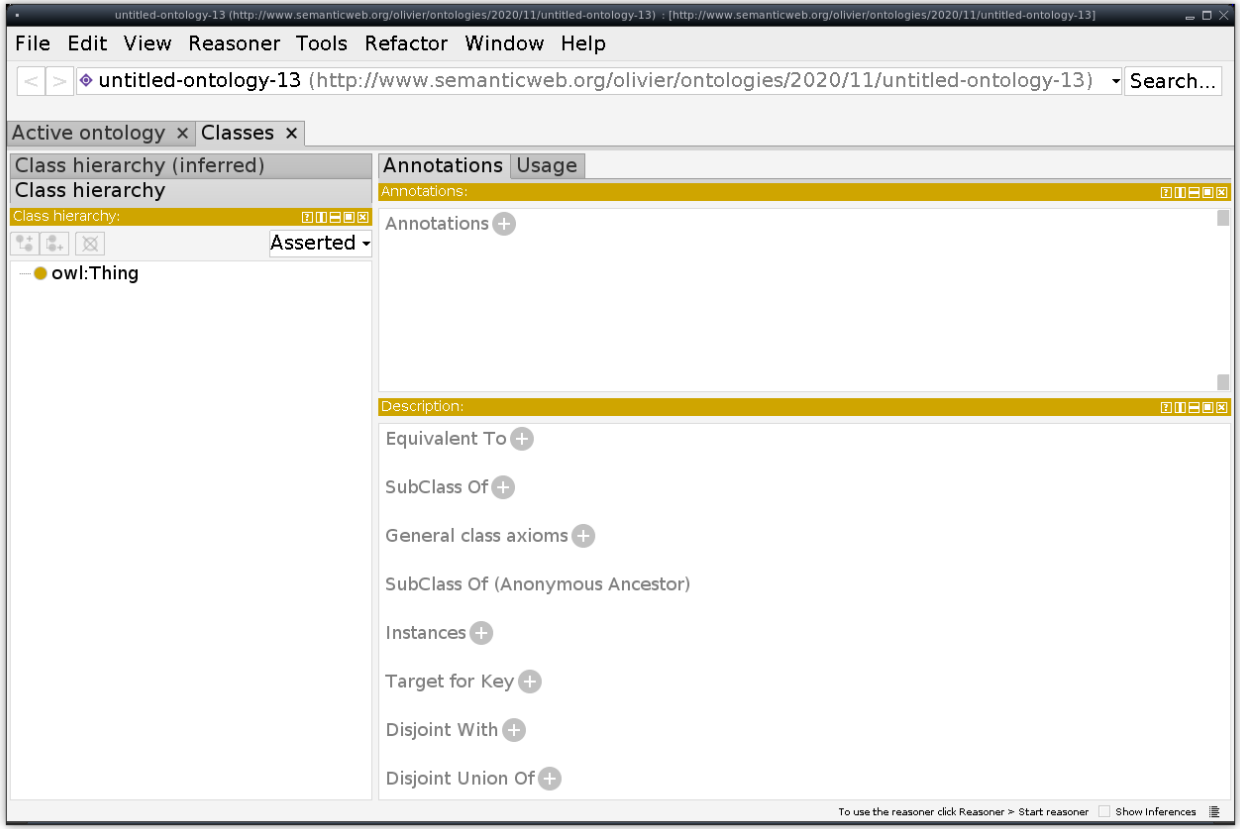


1. Retrieve Protégé from `https://protege.stanford.edu/`
2. select the "Active ontology" tab
3. adjust the ontology IRI (= template for the future identifiers)

Activate the “Classes” tab (if not already visible)

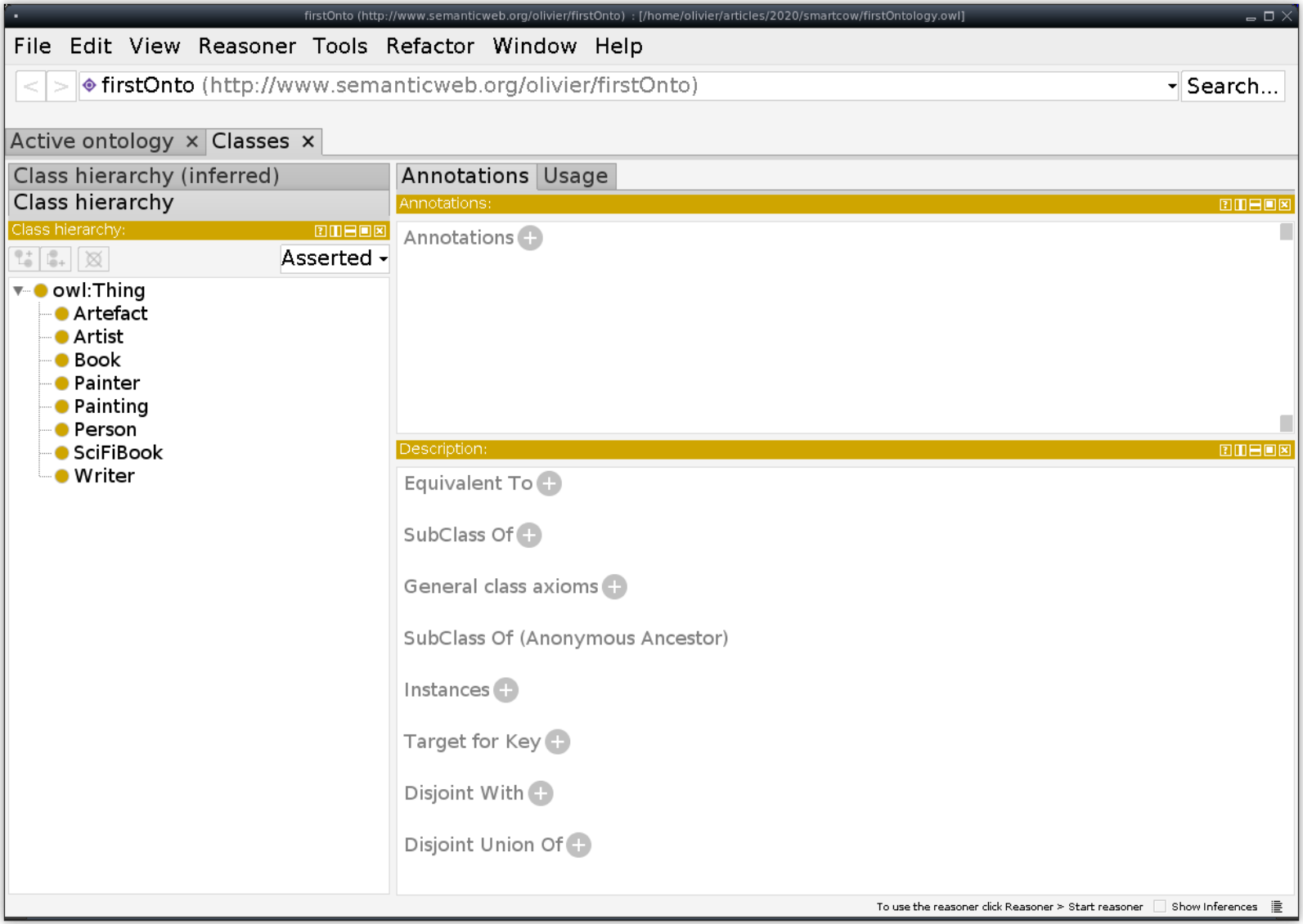


“Classes” tab organization

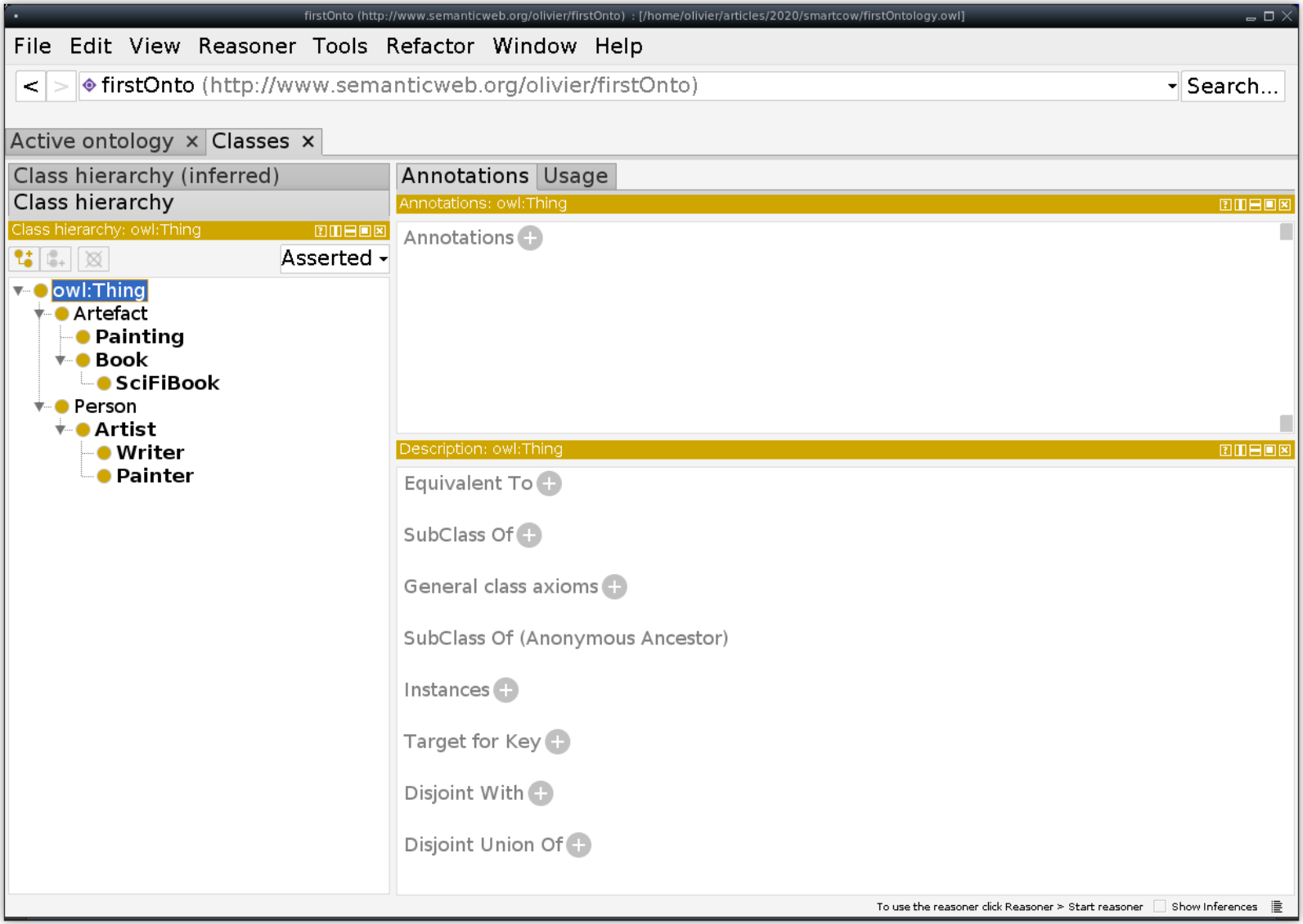


- left: hierarchy of classes (top = general; bottom = particular)
- top right: annotations for describing the hierarchy active class
- bottom right: properties formalizing the hierarchy active class

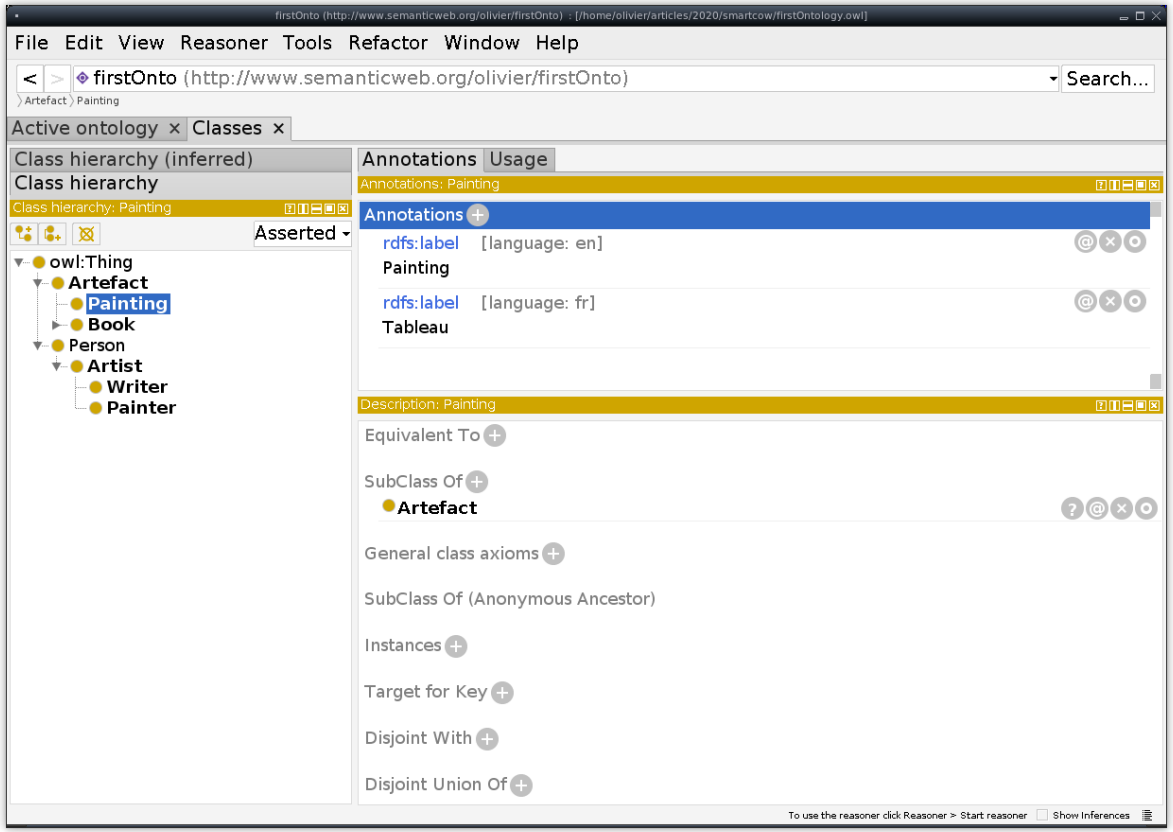
Create new classes



Organize classes into a hierarchy

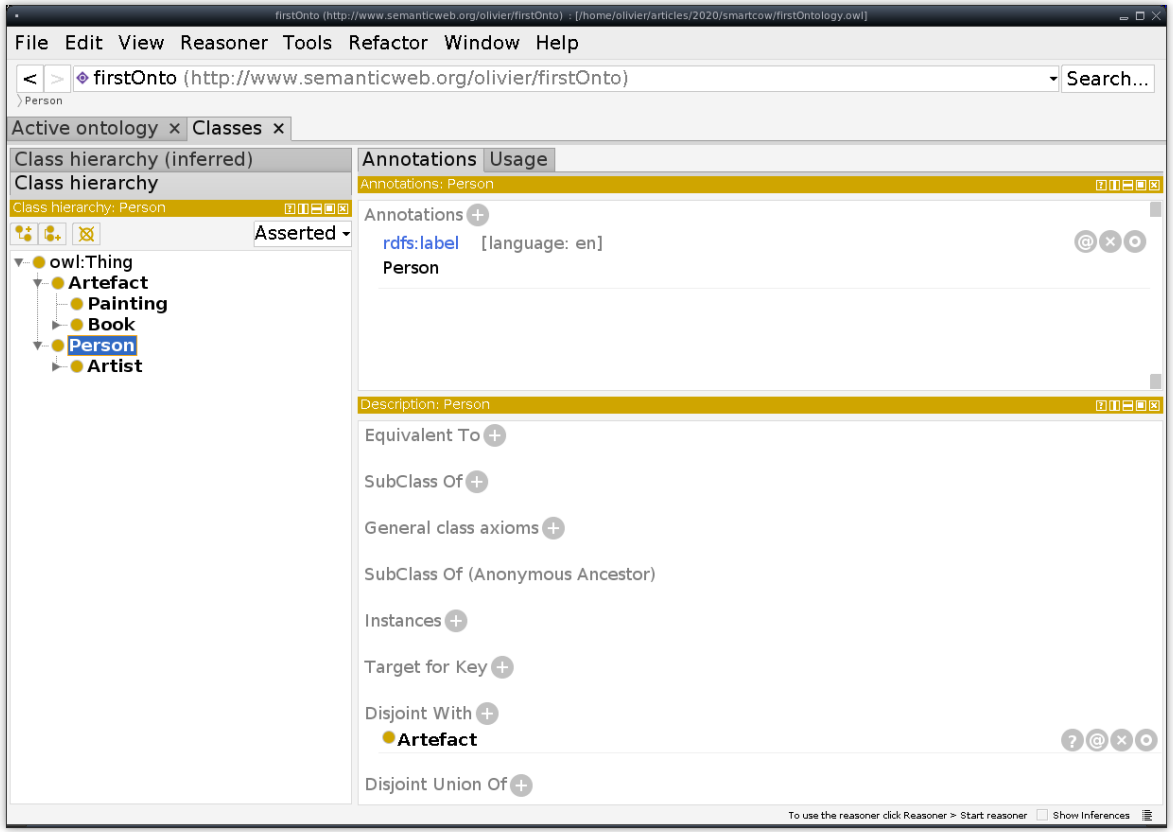


Annotate the classes



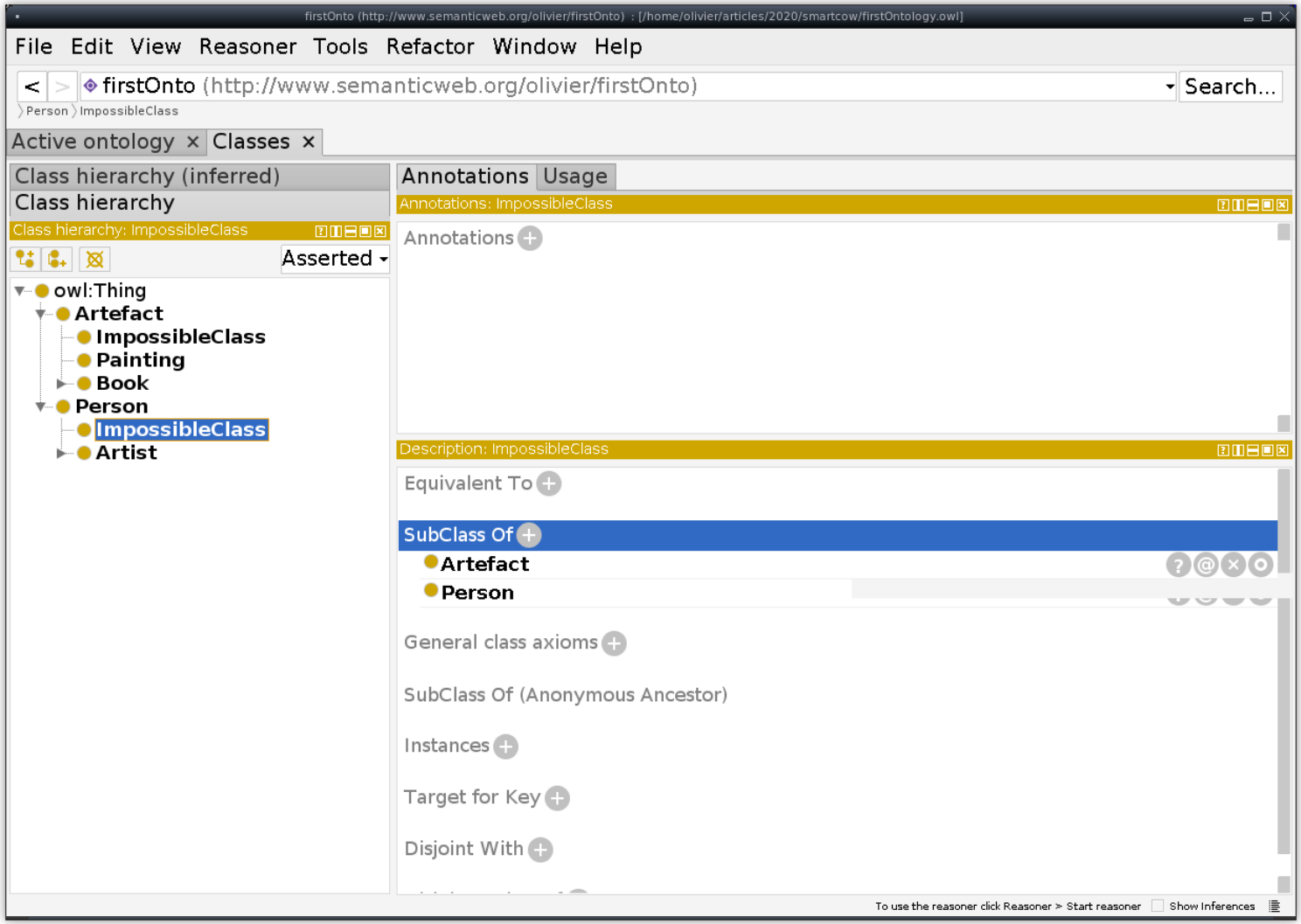
- `rdfs:label` = name(s) of the class
- `rdfs:comment` = description of textual definition

Disjoint classes cannot have common instances



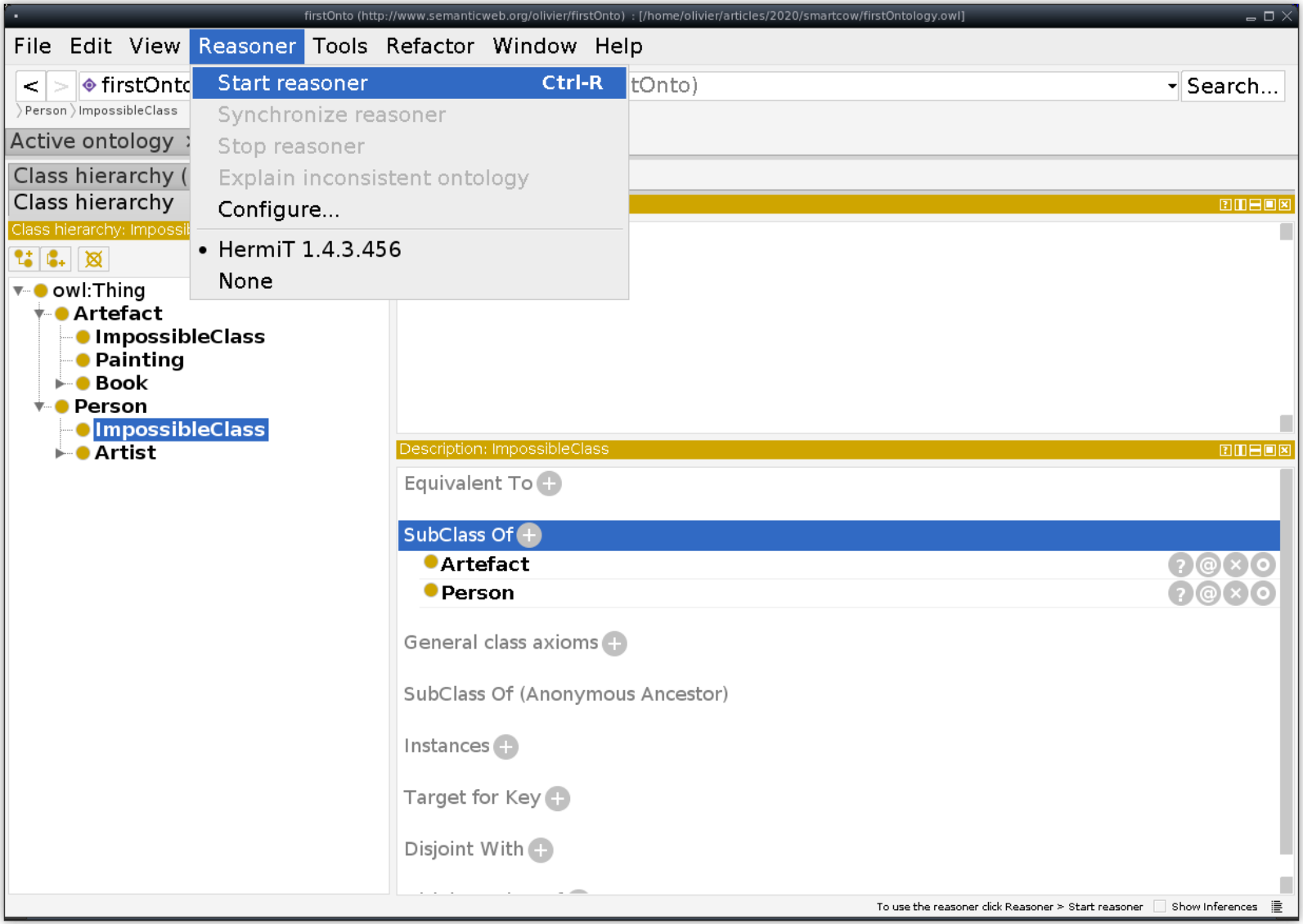
- `Person` and `Artefact` are disjoint
- `Painter` and `Writer` are not
- notice that by inheritance, `Painter` and `Artefact` are also disjoint

Classes can have multiple parents

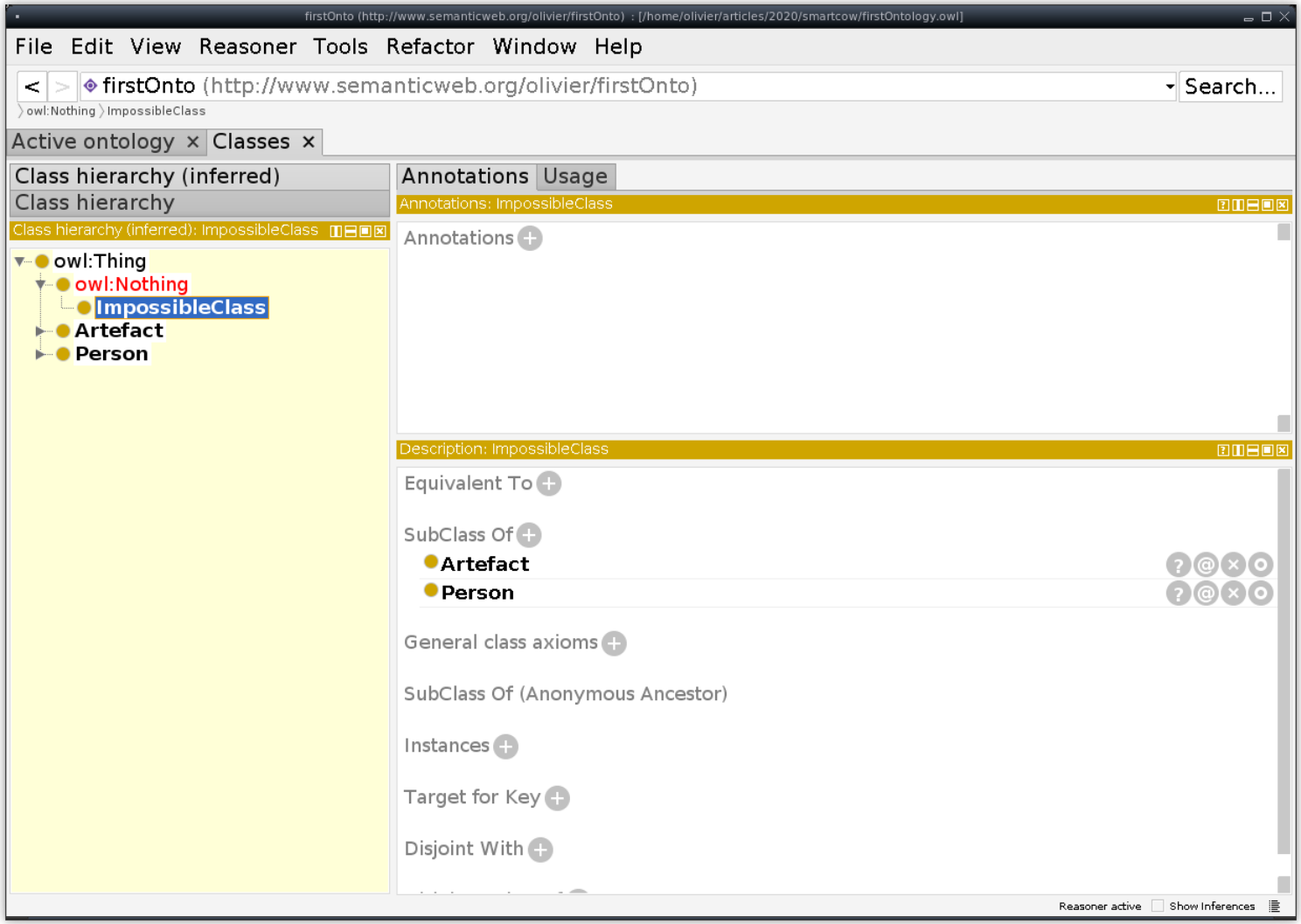


- notice that `ImpossibleClass` appears twice in the hierarchy

Start the reasoner

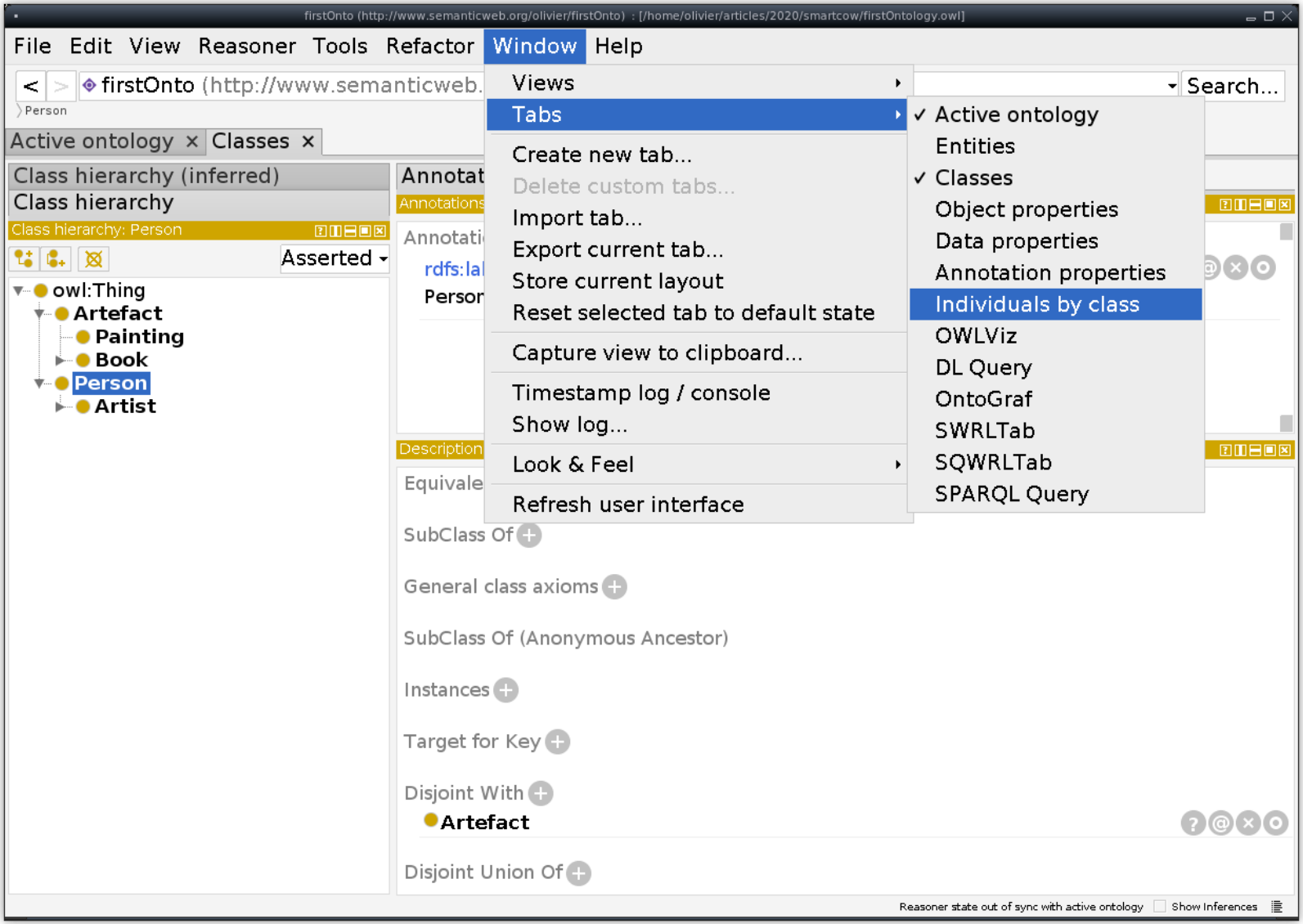


The reasoner detects inconsistent classes

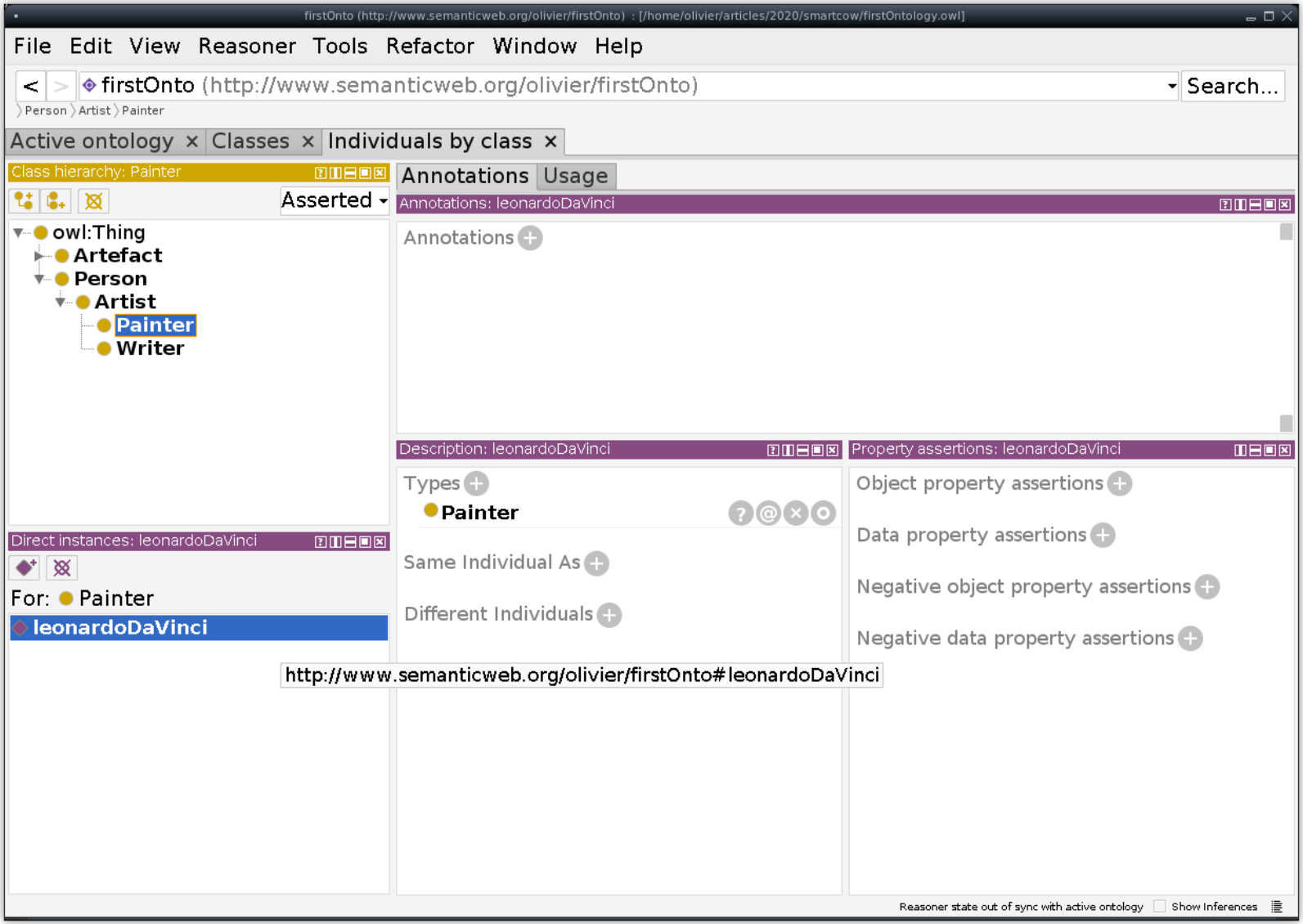


- notice that we switched to the inferred class hierarchy

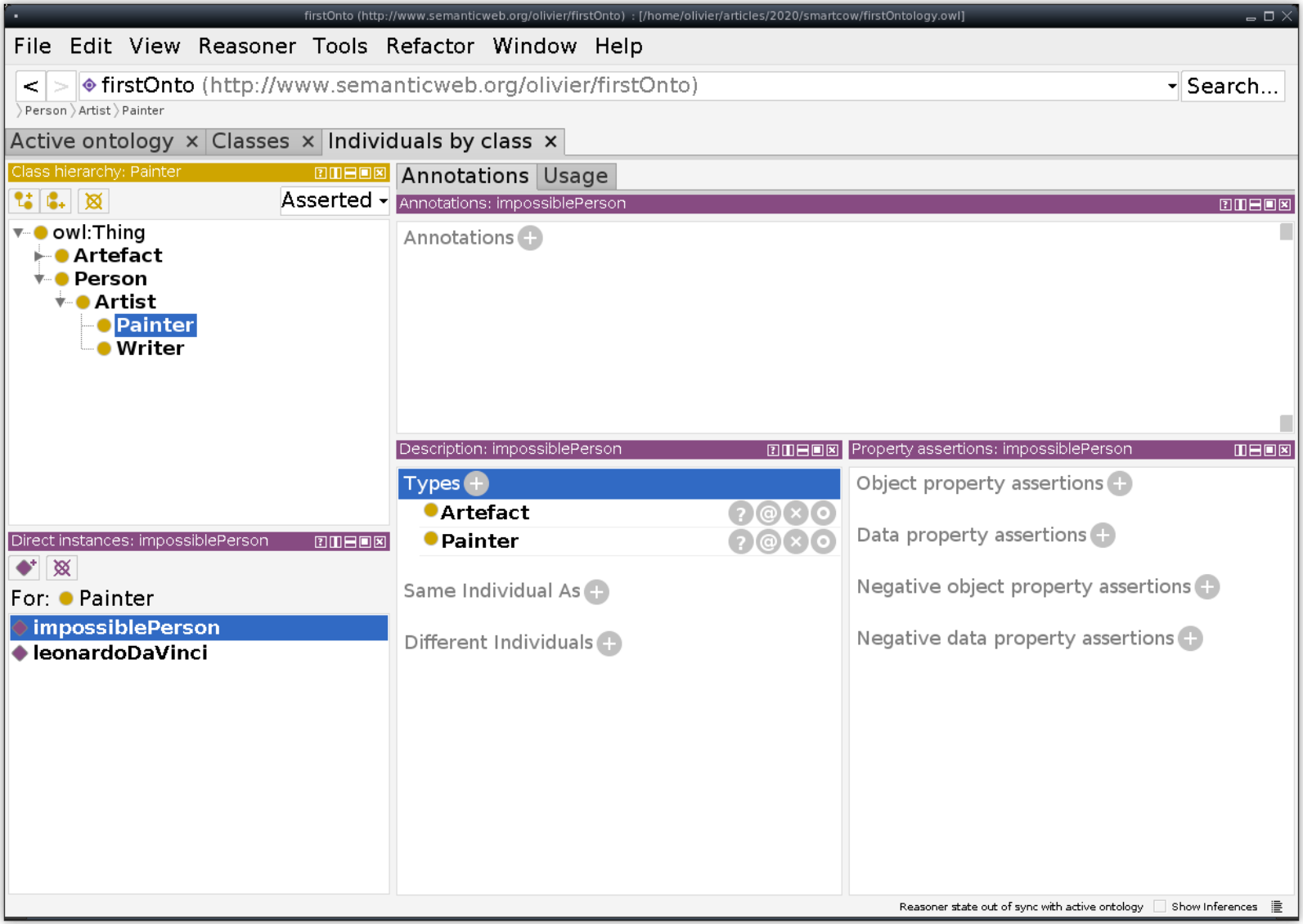
Activate the “Individuals” tab



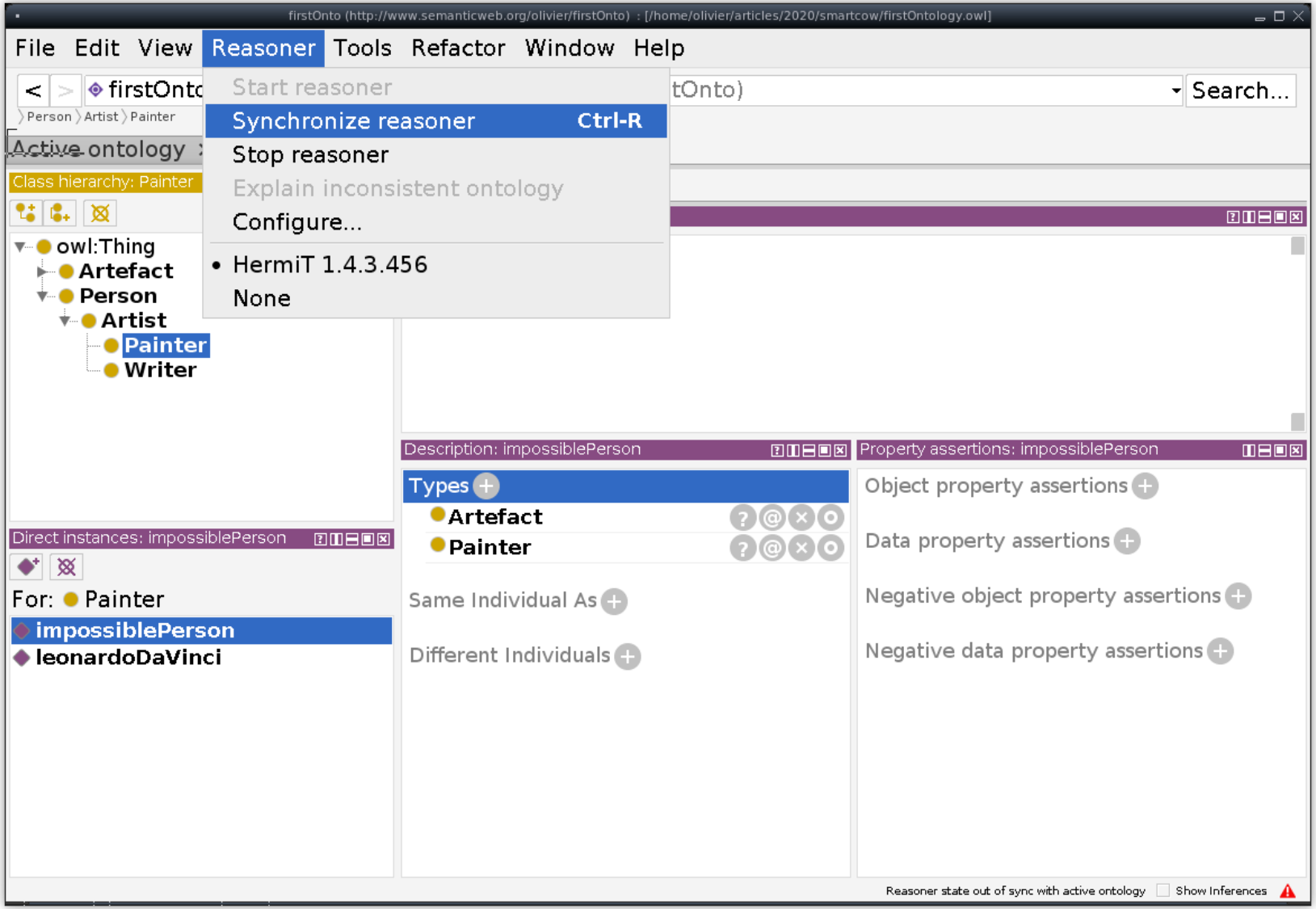
Create an instance of the `Painter` class



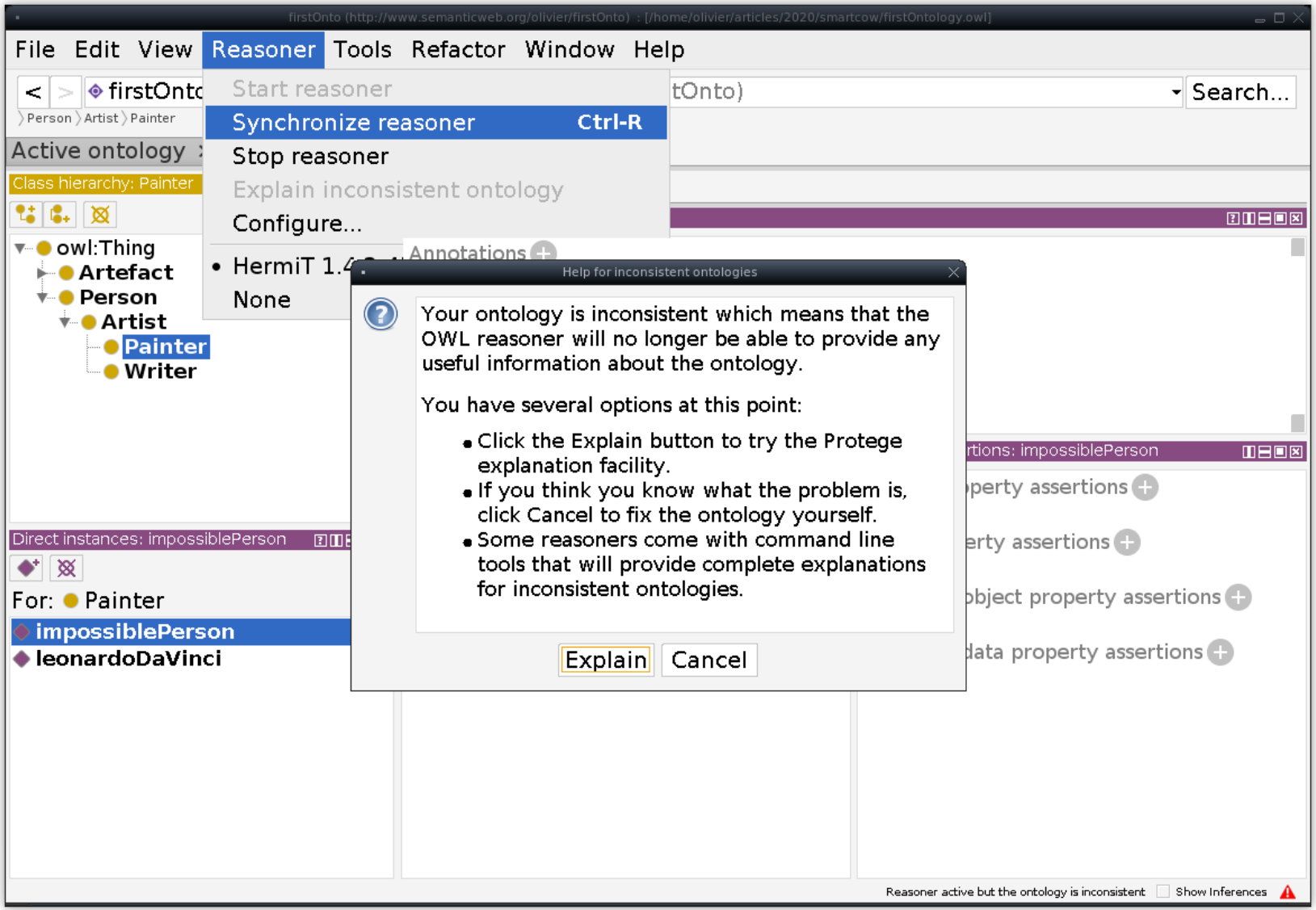
Create an instance of disjoint classes



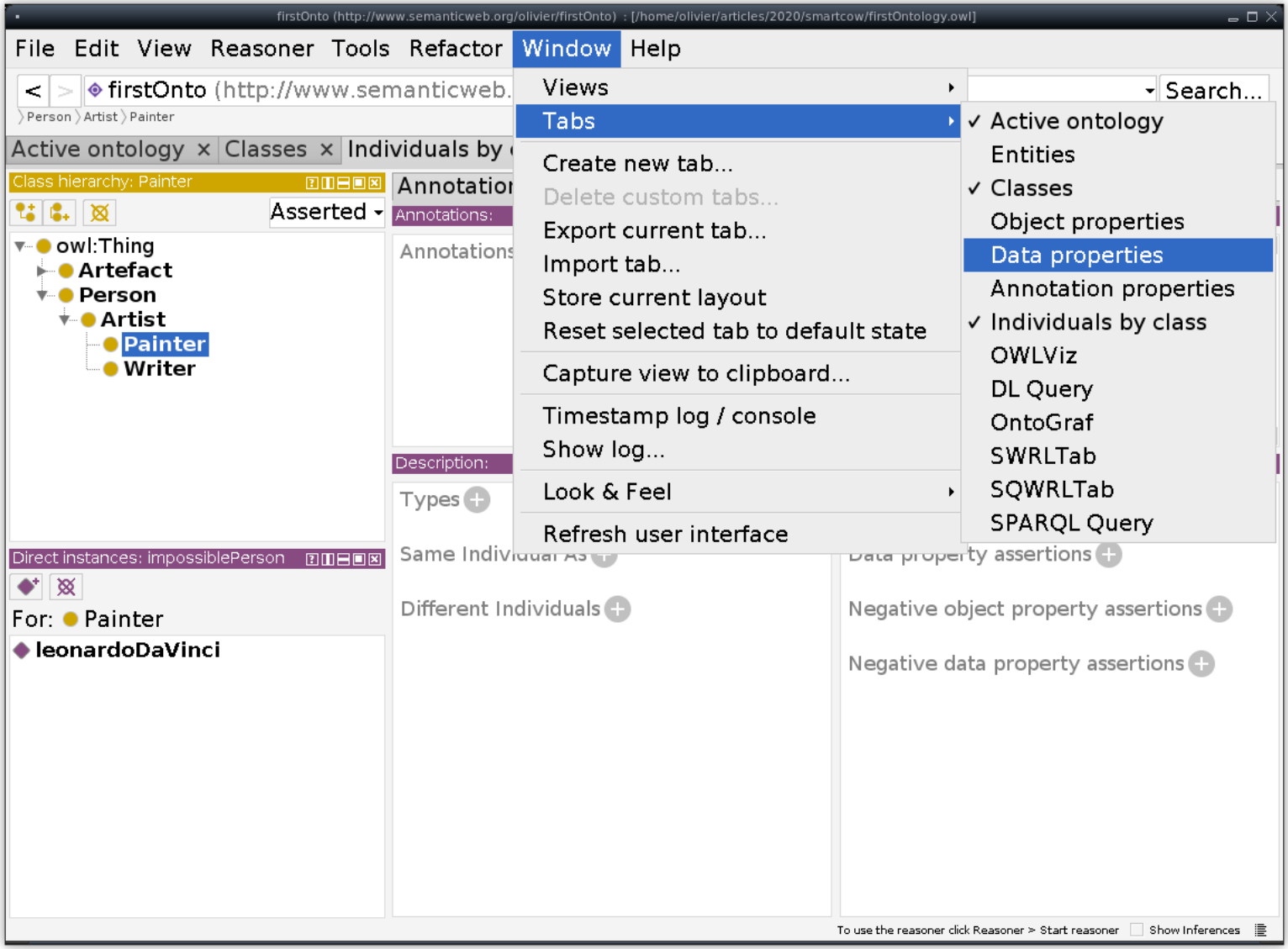
Synchronize the reasoner



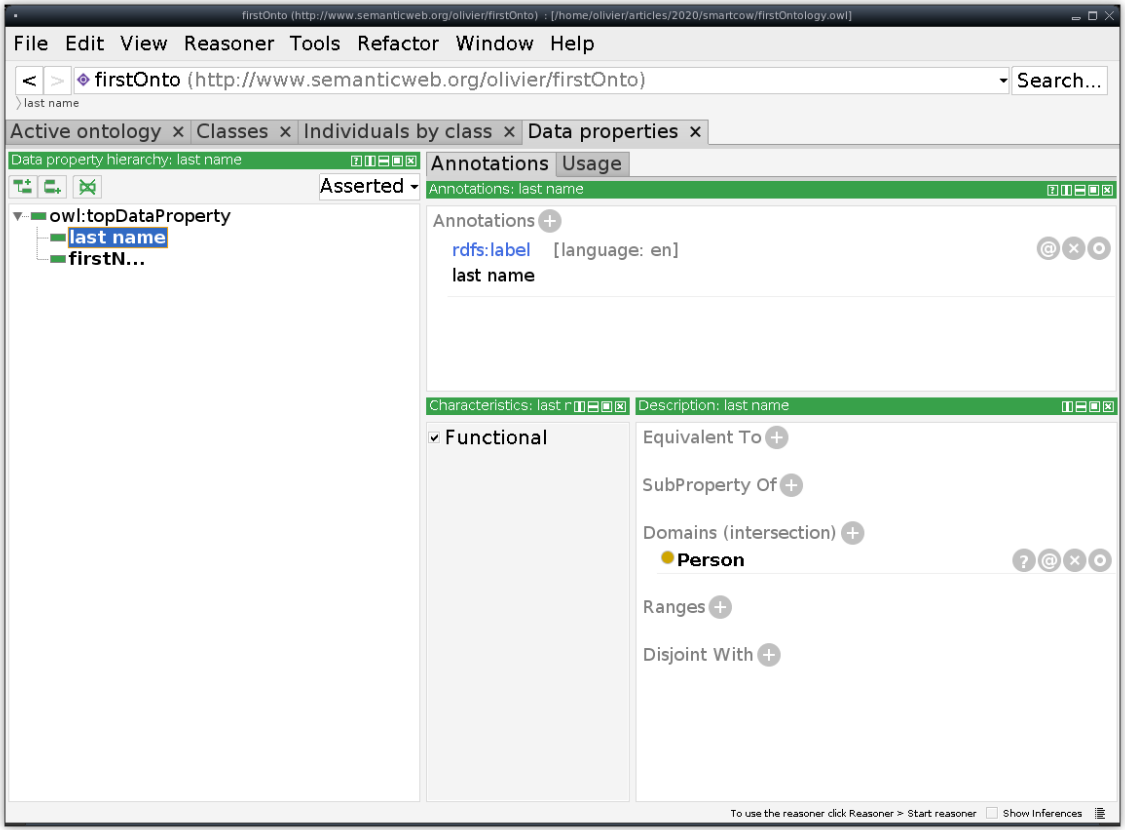
The reasoner detects inconsistent individuals



Activate the “Datatype properties” tab

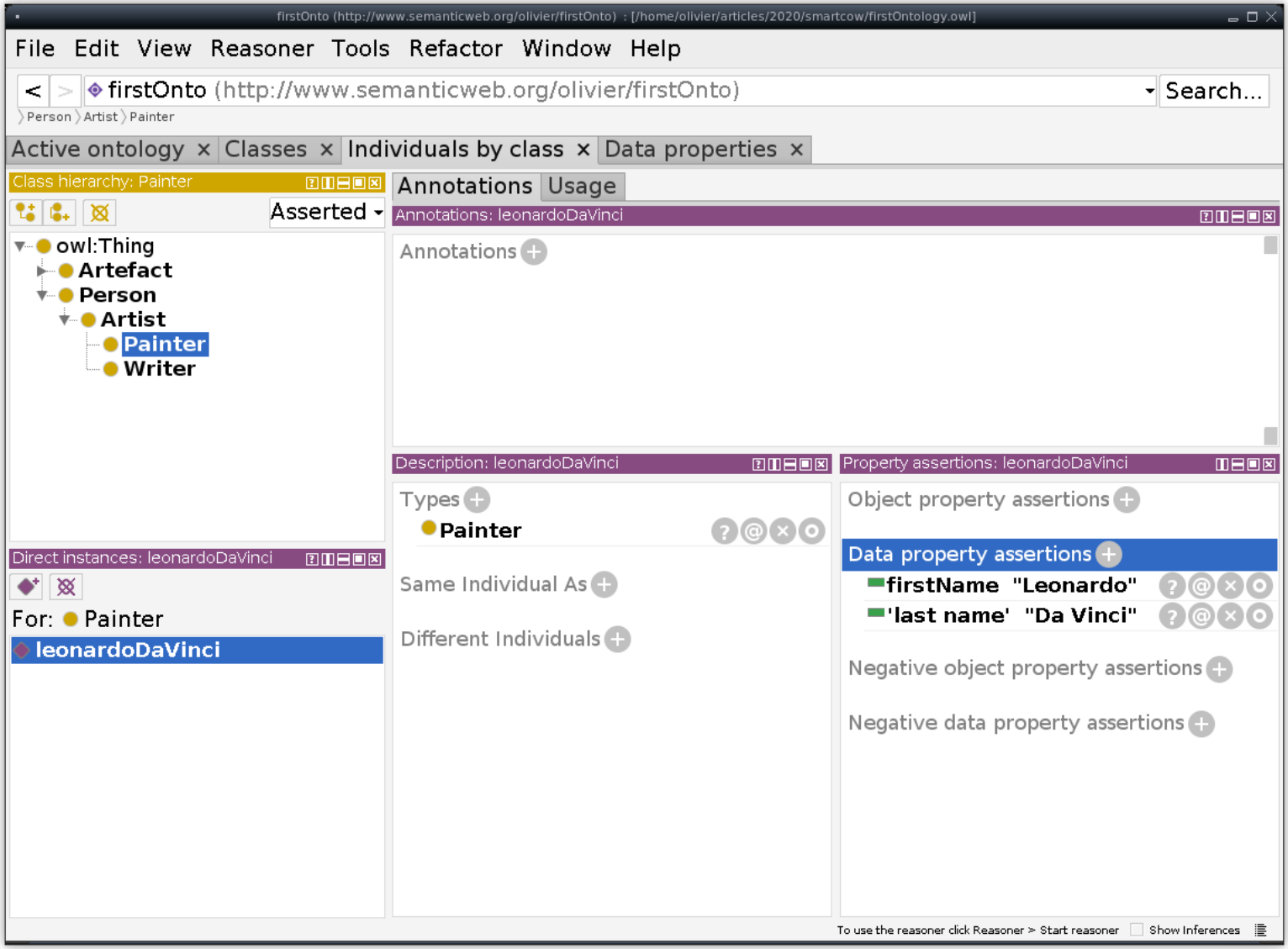


Create datatype properties

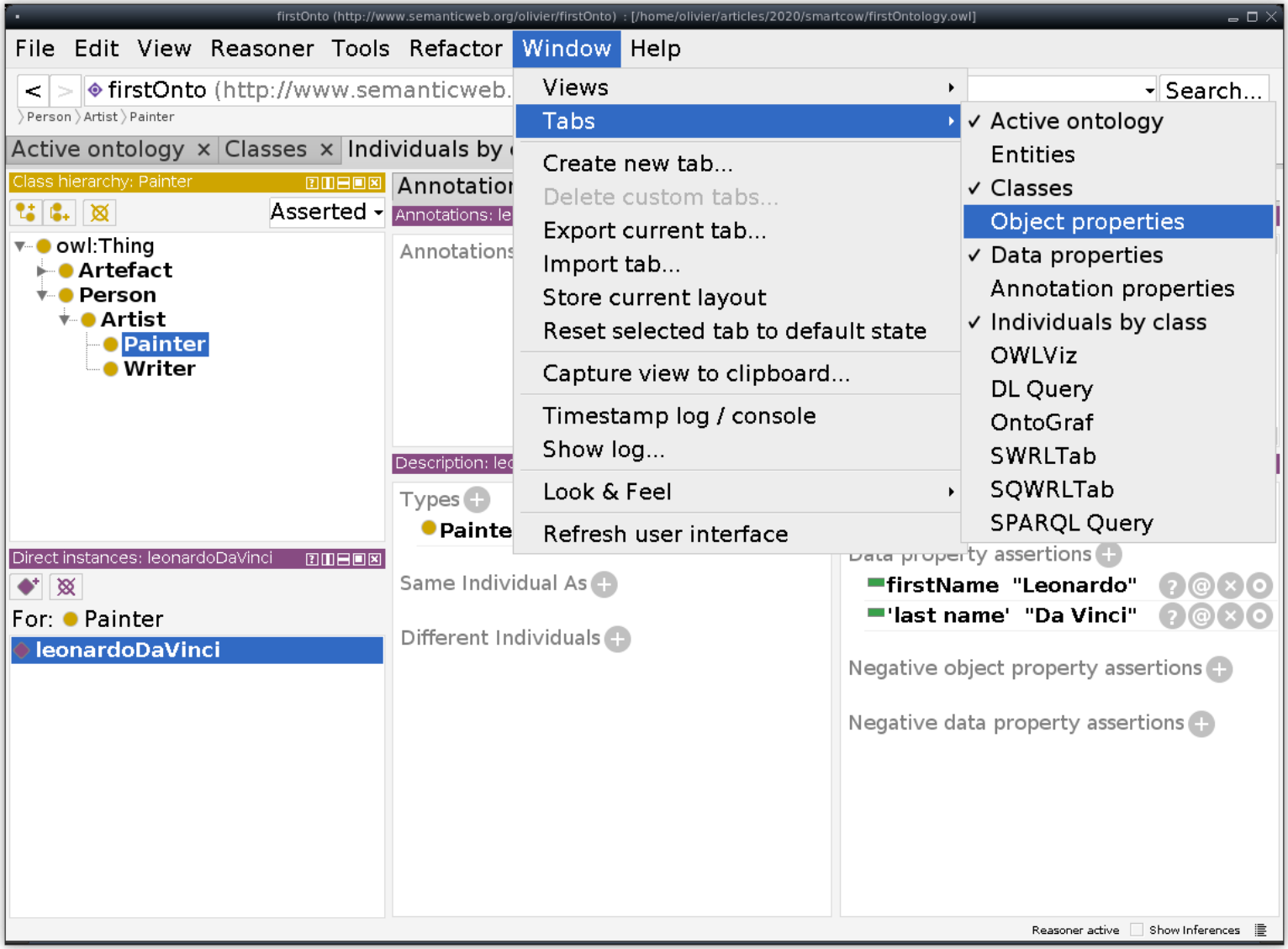


- create relations which values are strings, numbers, dates...
- the domain is the set of individuals for which this relation makes sense
- functional property (checkbox) can have at most one value

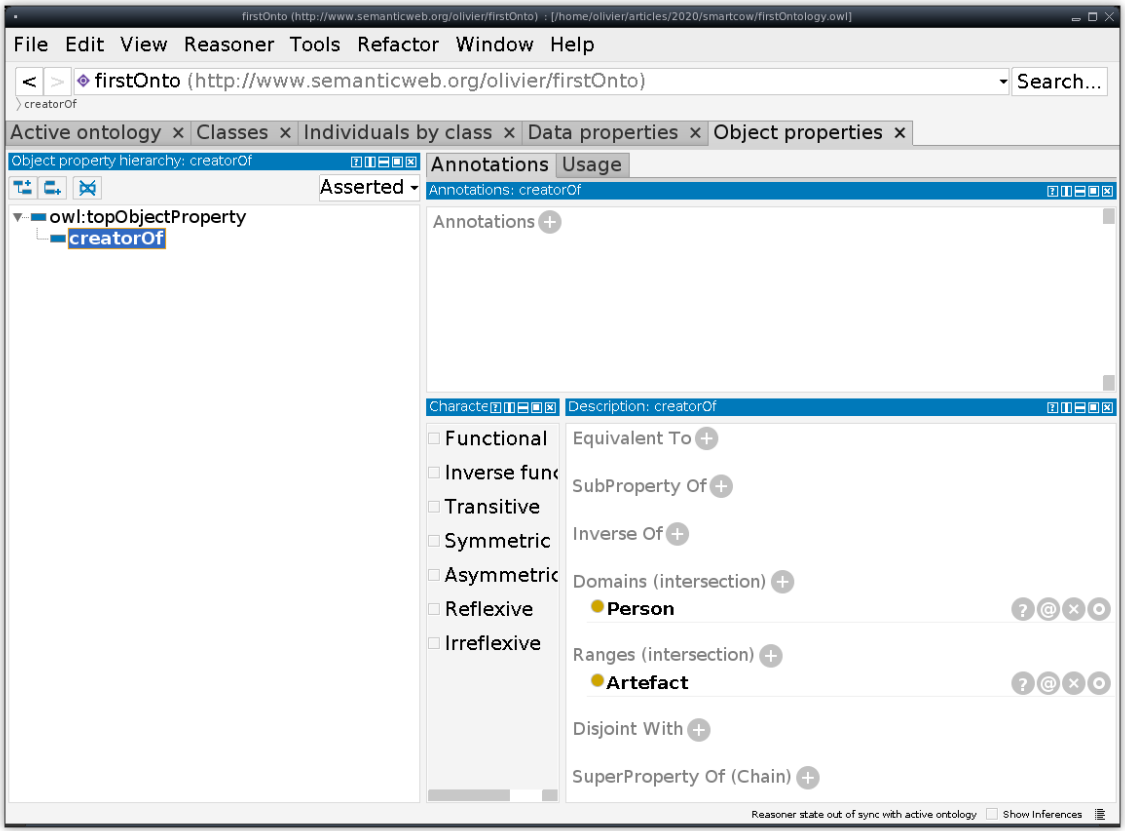
Use datatype properties to describe individuals



Activate the “Object properties” tab

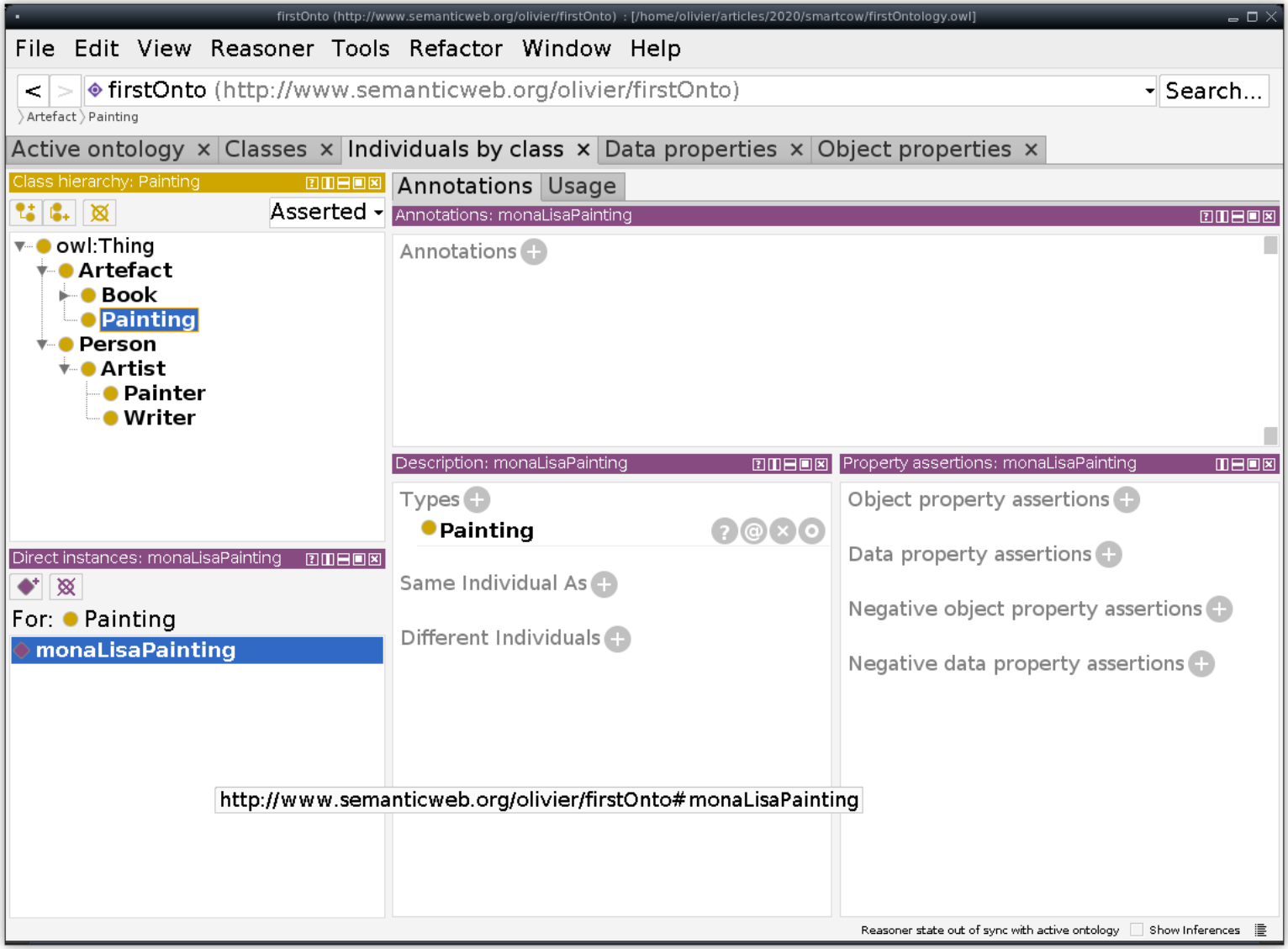


Create an object property

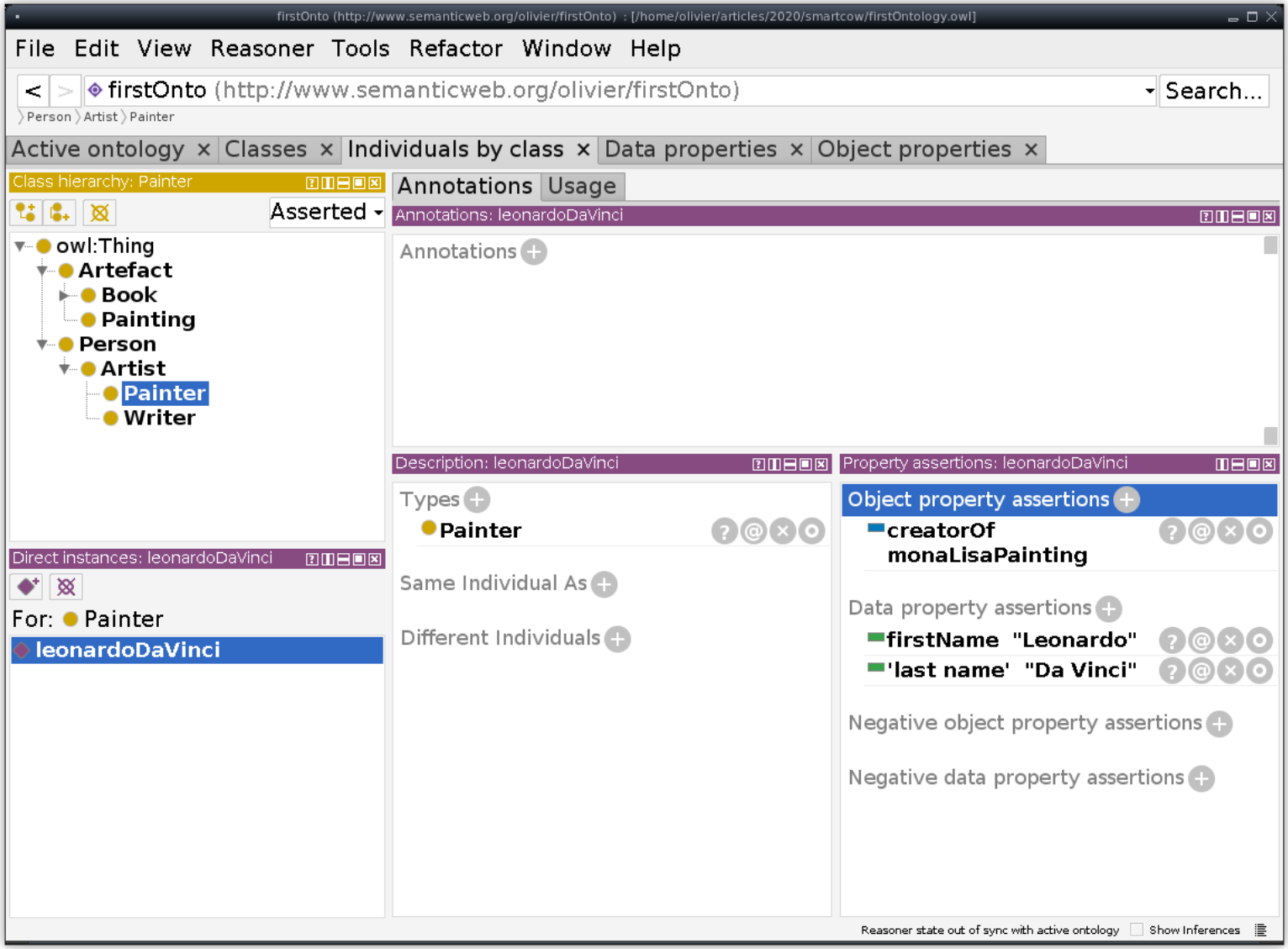


- create relations between individuals
- domain = set of individuals for which this relation makes sense
- range = set of possible values for the relation

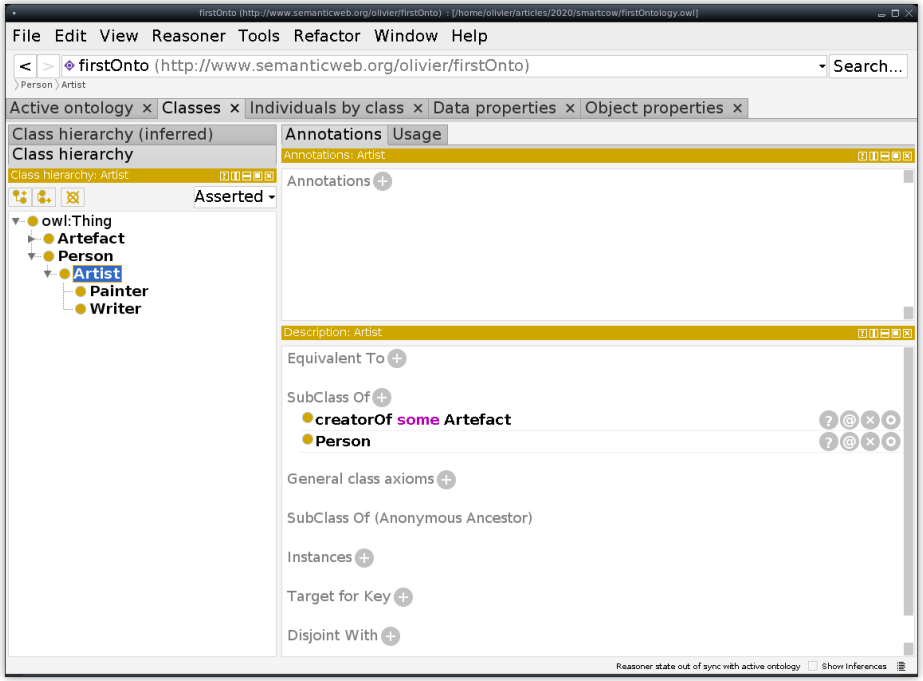
Create an instance of `Painting`



Use an object property to connect Da Vinci and Mona Lisa

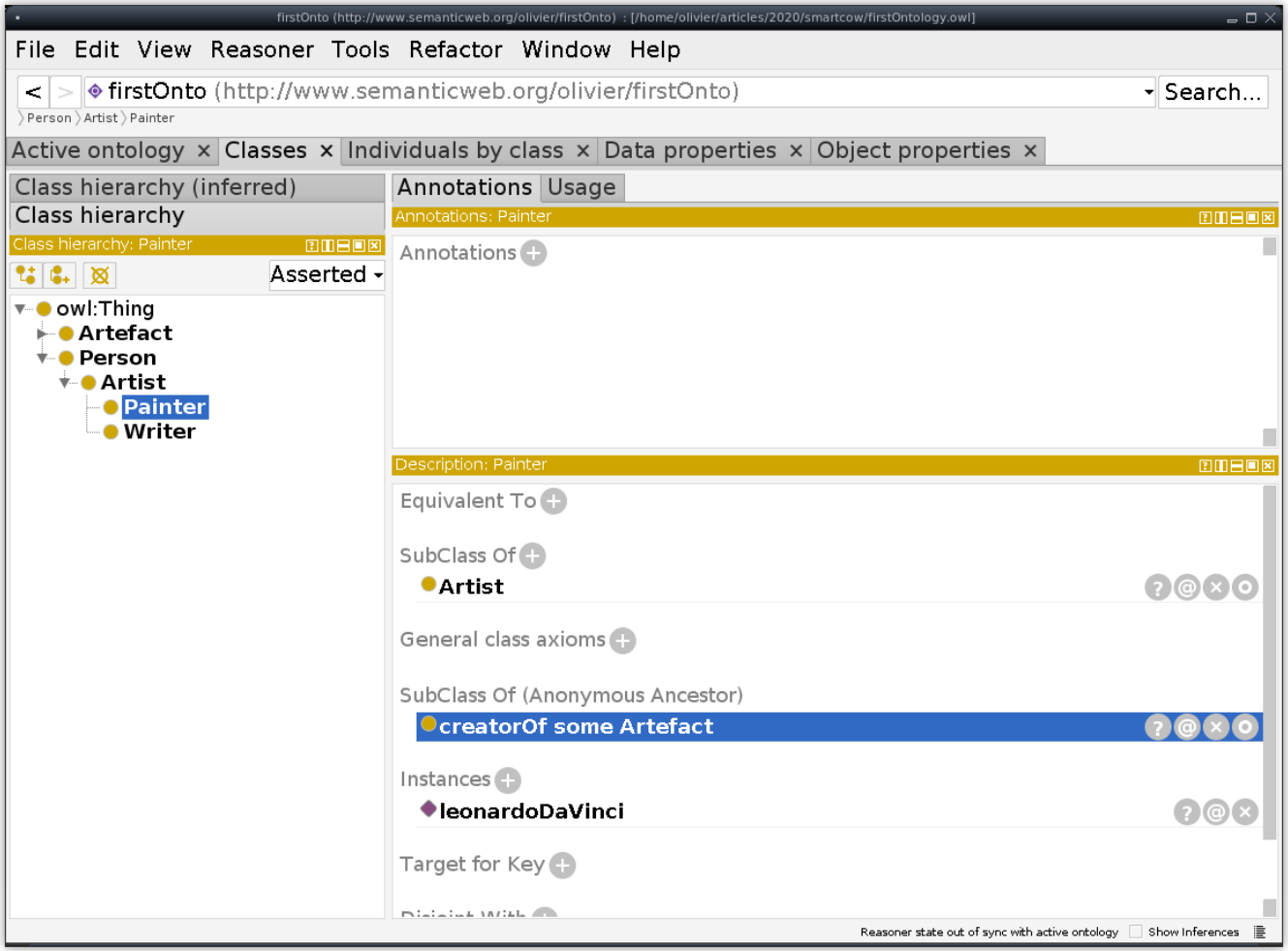


Use properties to add constraints to classes



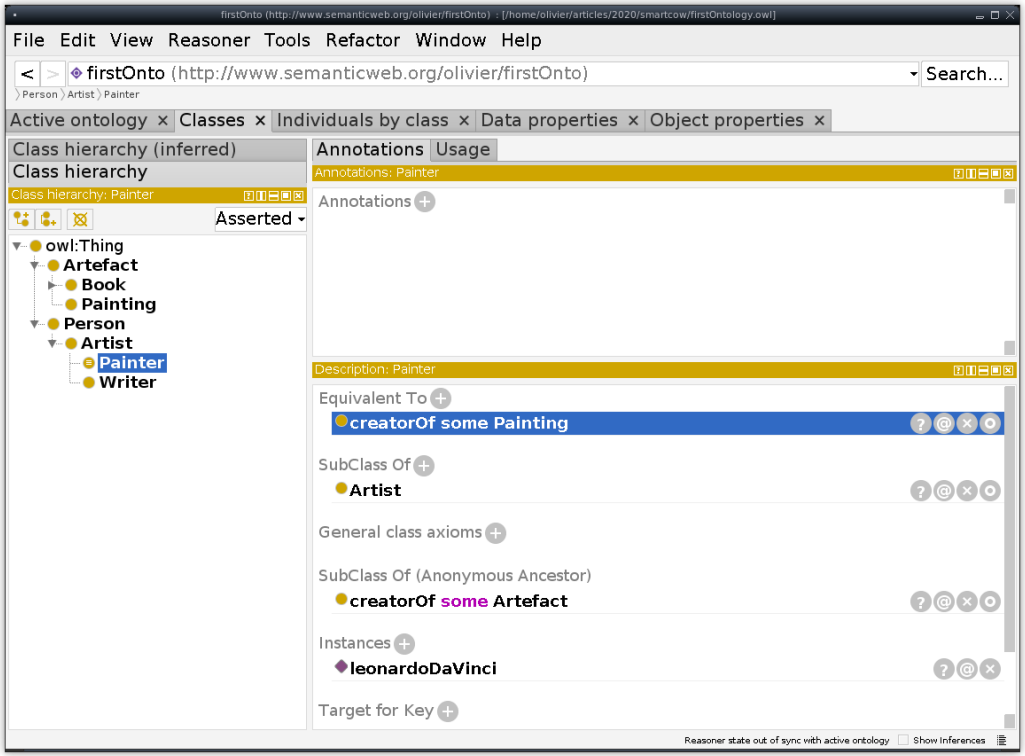
- if x is an instance of `Artist`, then
  - x is the `creatorOf` at least one y (notice the "some" keyword)
  - this y is an instance of `Artefact`
- NB1: y can be unspecified; the reasoner knows there is one
- NB2: being an artist $\Rightarrow$ having created an artefact
  but having created an artefact does not make you an artist

A class' constraints are inherited by its subclasses



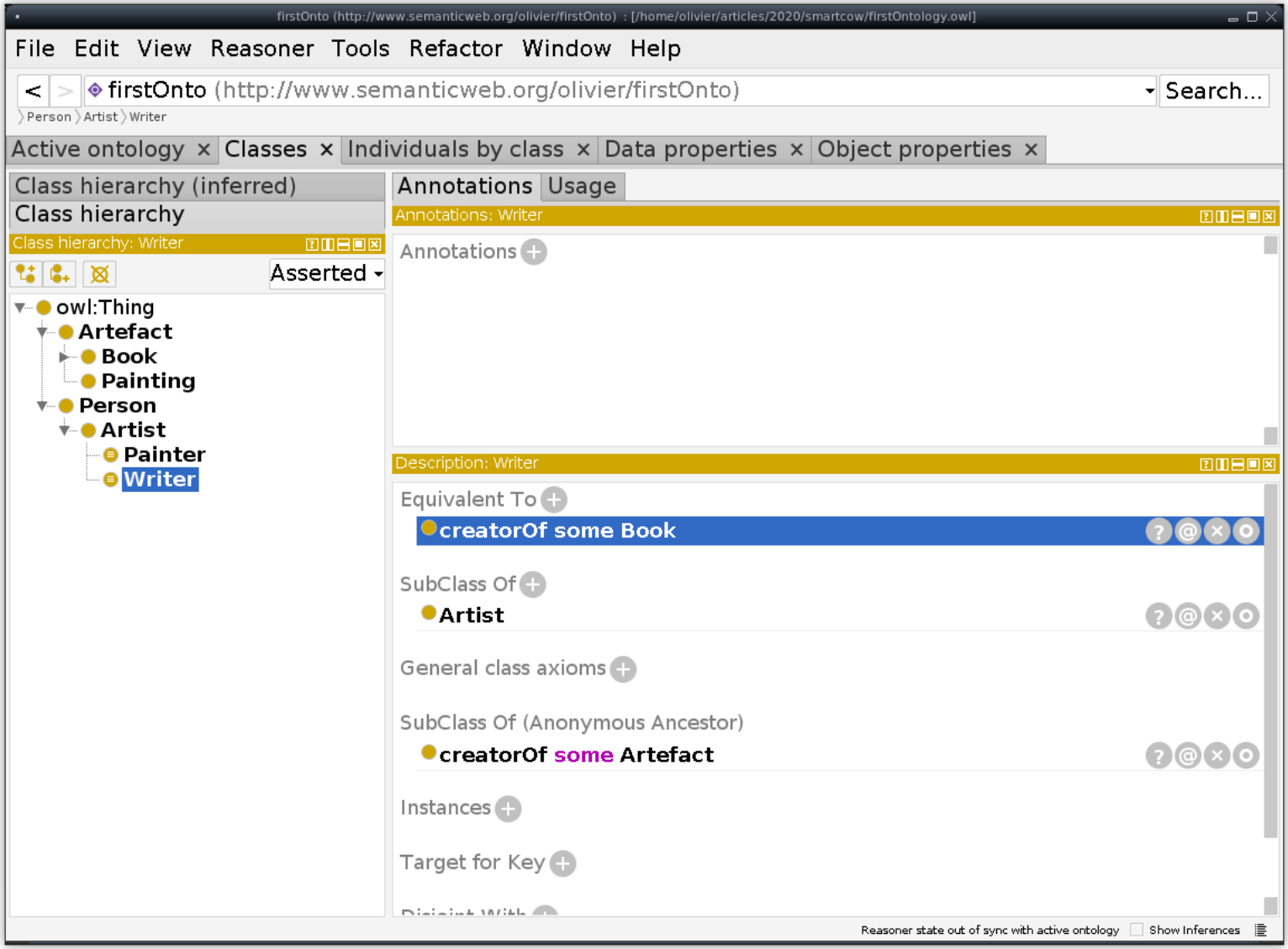
- because all `Painters` are `Artists`, they must have created at least one `Artefact`

Defined classes

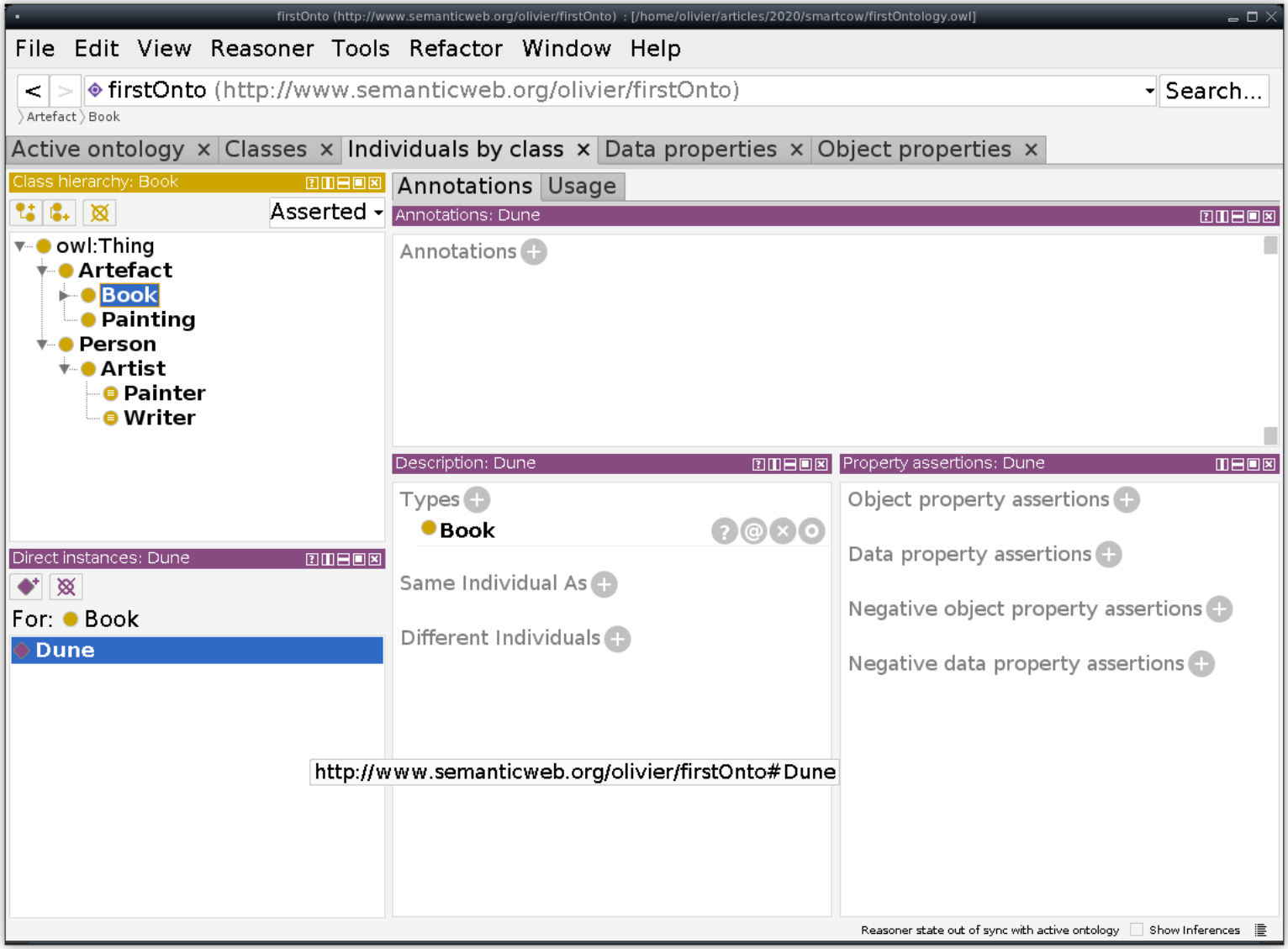


- we refine the (inherited) constraint for `Painter` so that the must have created at least one `Painting` (which is a subclass of `Artefact`)
- defined instead of subclass:
  having created a painting $\Leftrightarrow$ being a painter

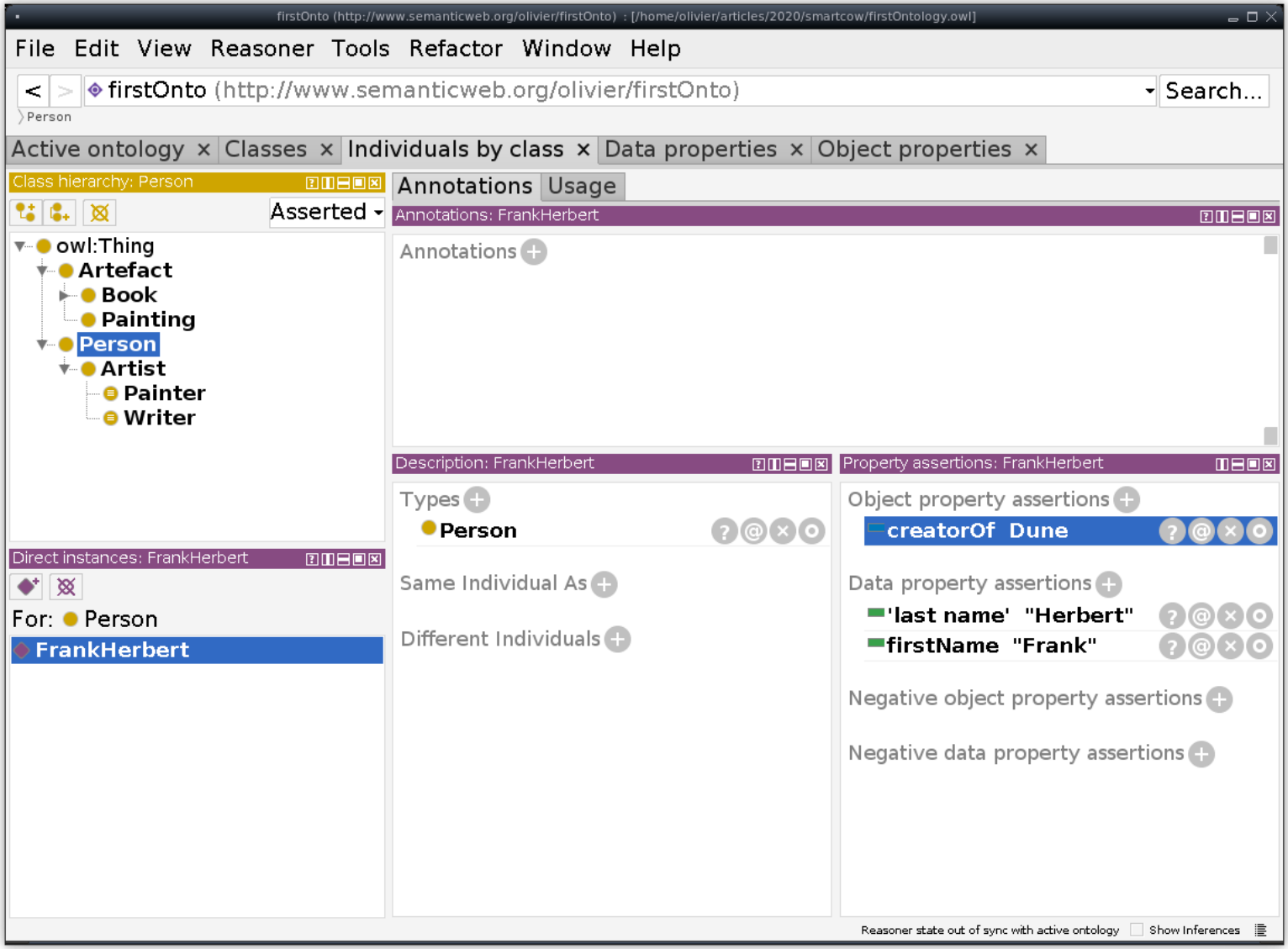
Defined classes



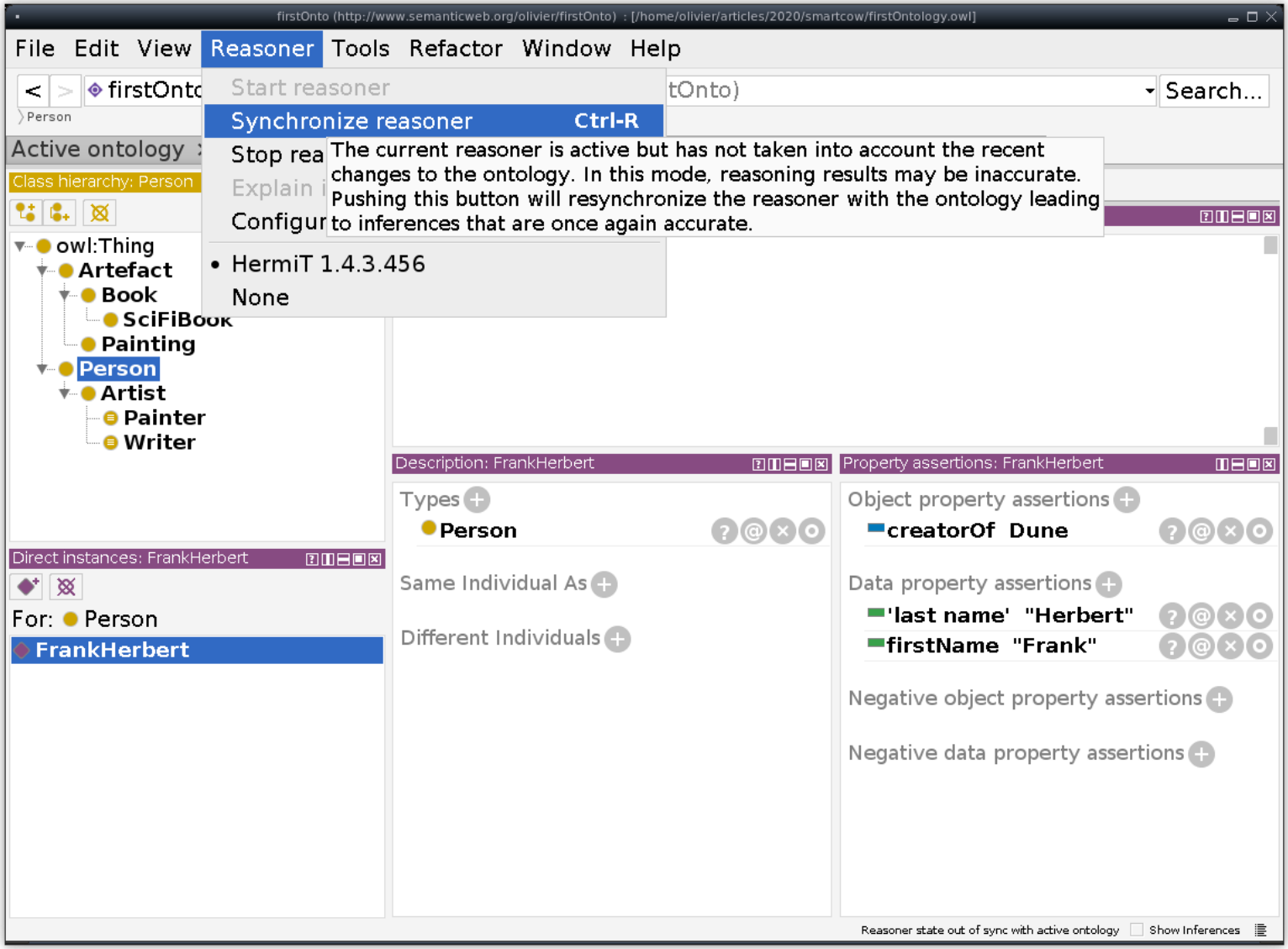
Dune is a book



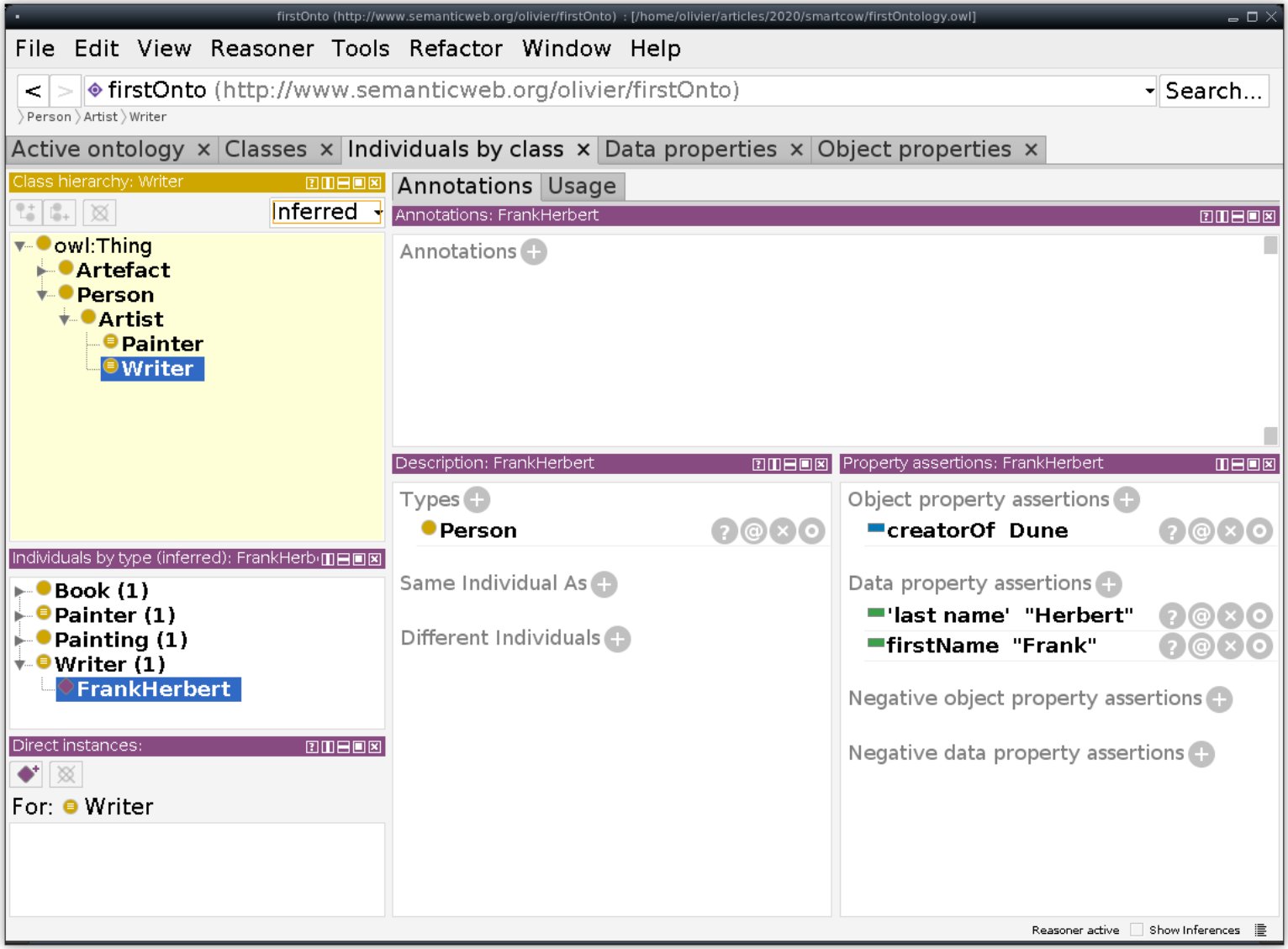
Frank Herbert is a Person and created Dune



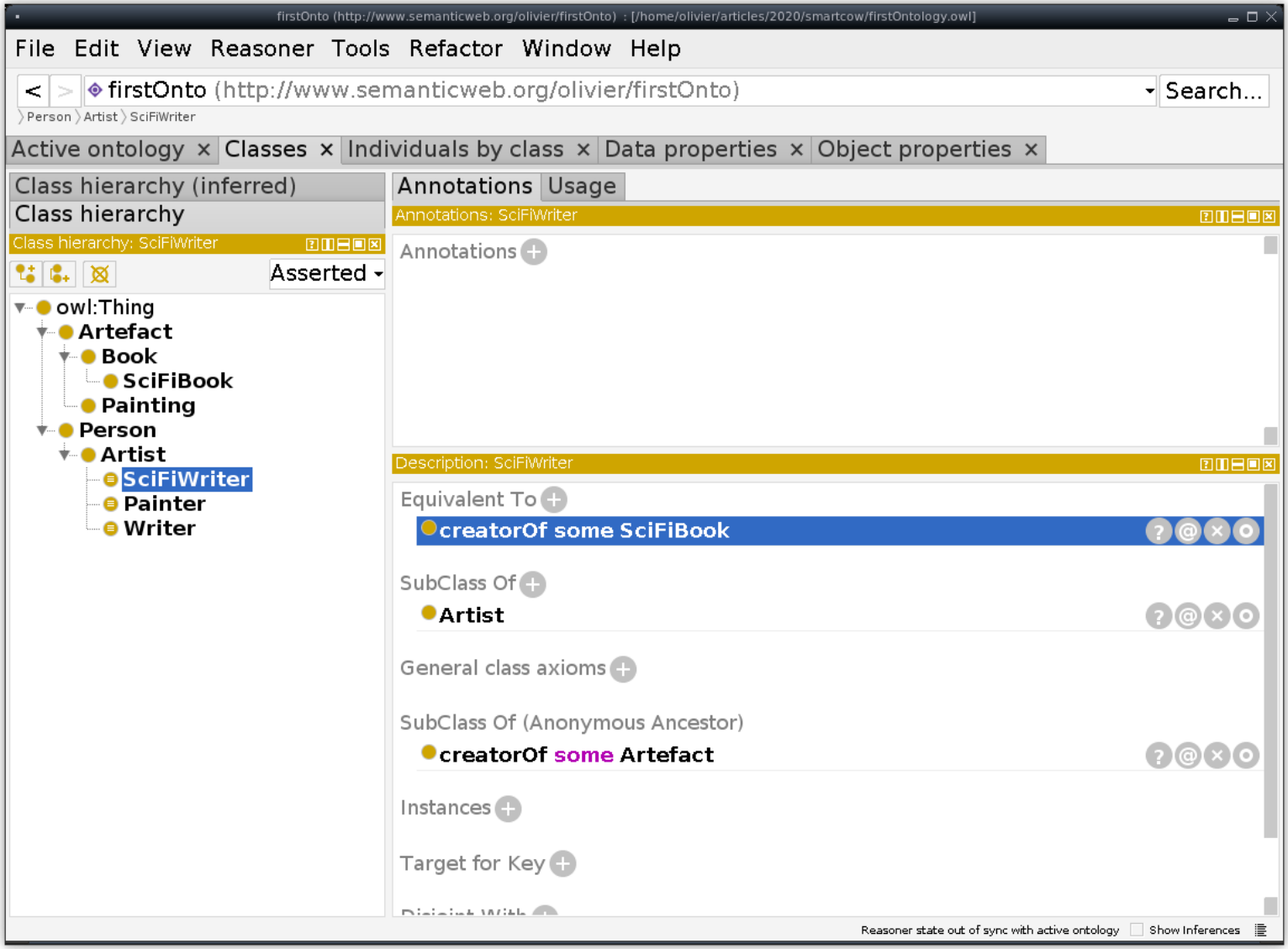
Synchronize the reasoner



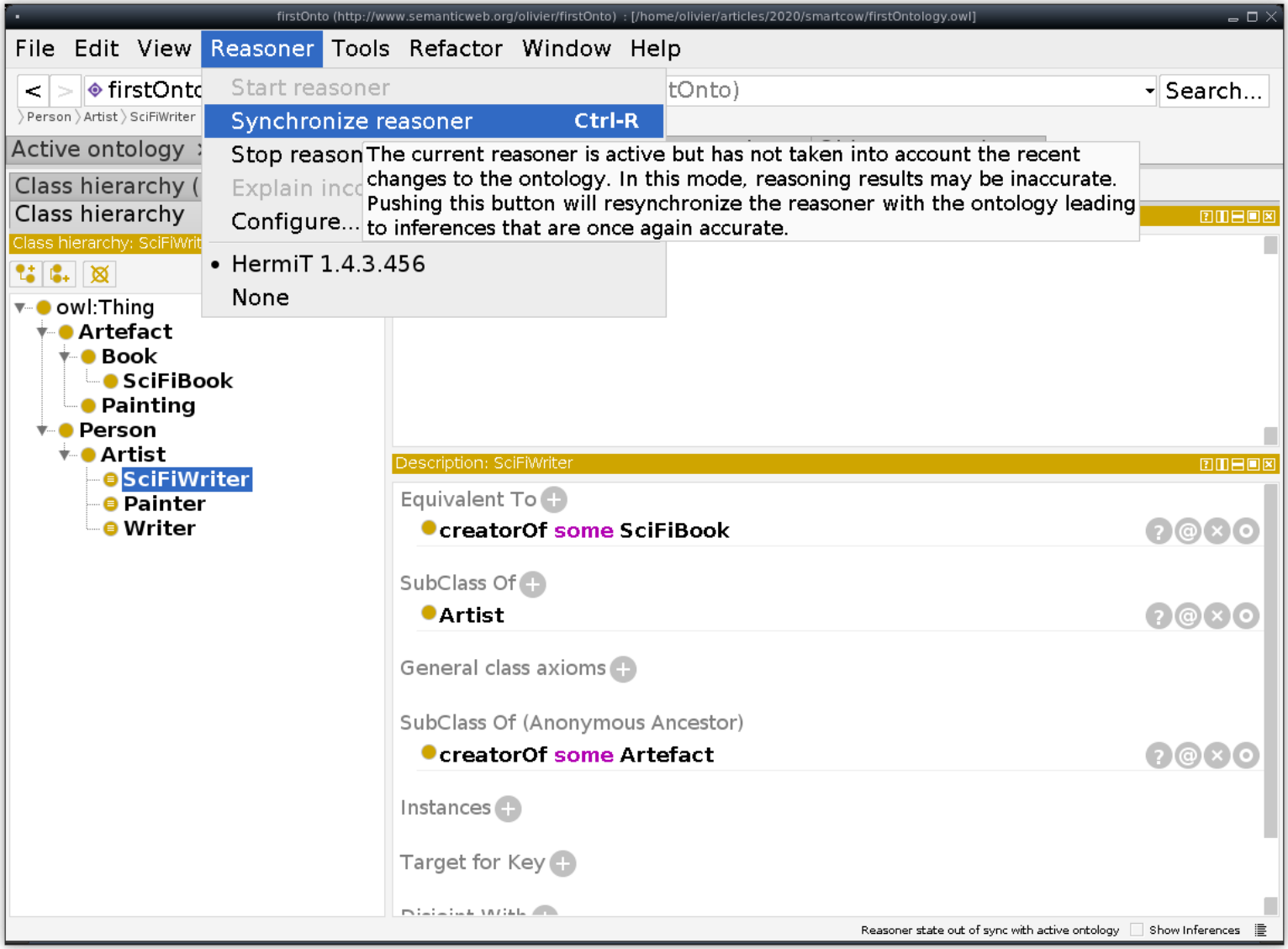
Class definition: Frank Herbert is recognized as a `Writer`



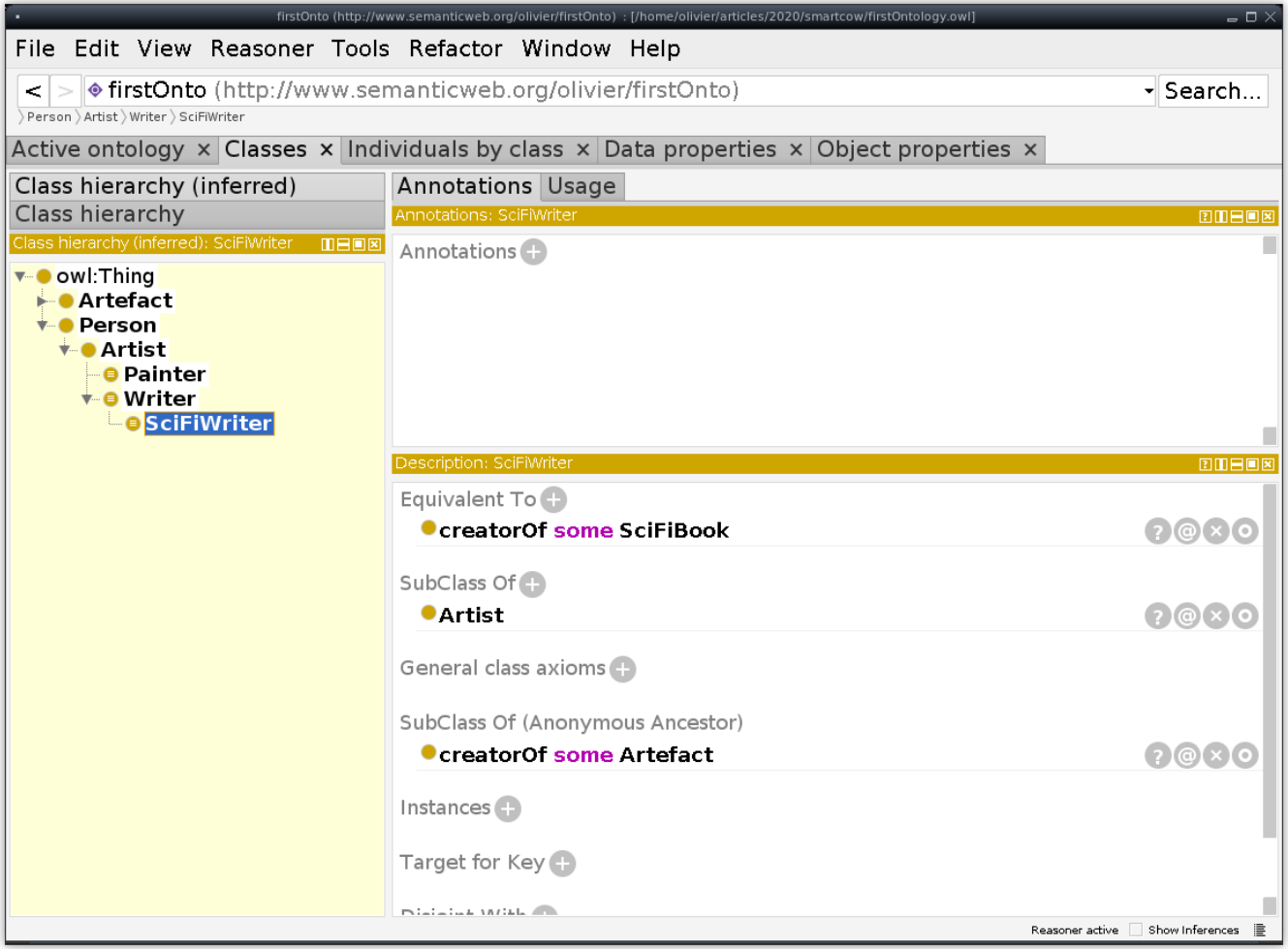
Create and define SciFiWriter as a subclass of Person



Synchronize the reasoner

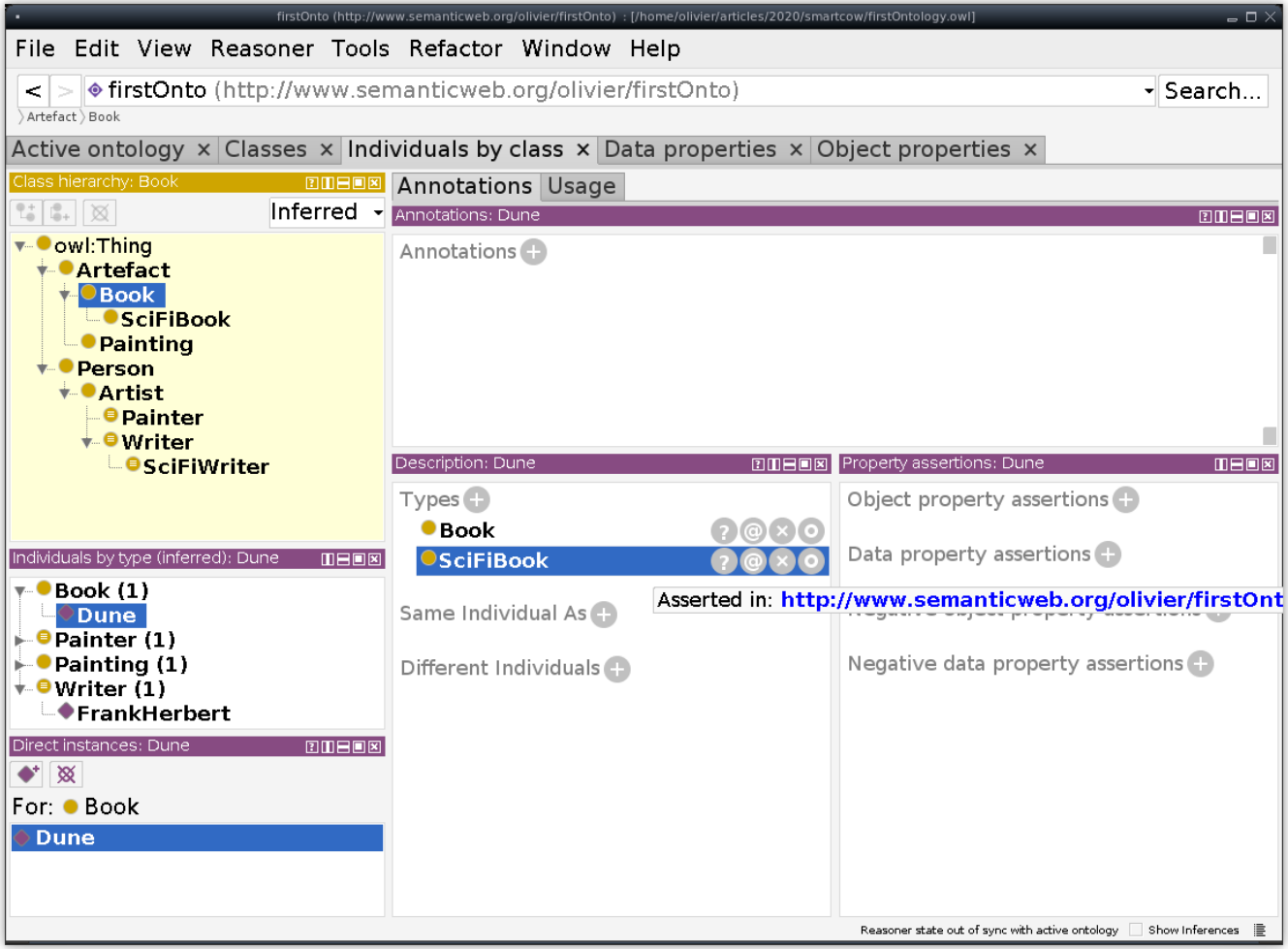


SciFiWriter is inferred to a subclass of Writer



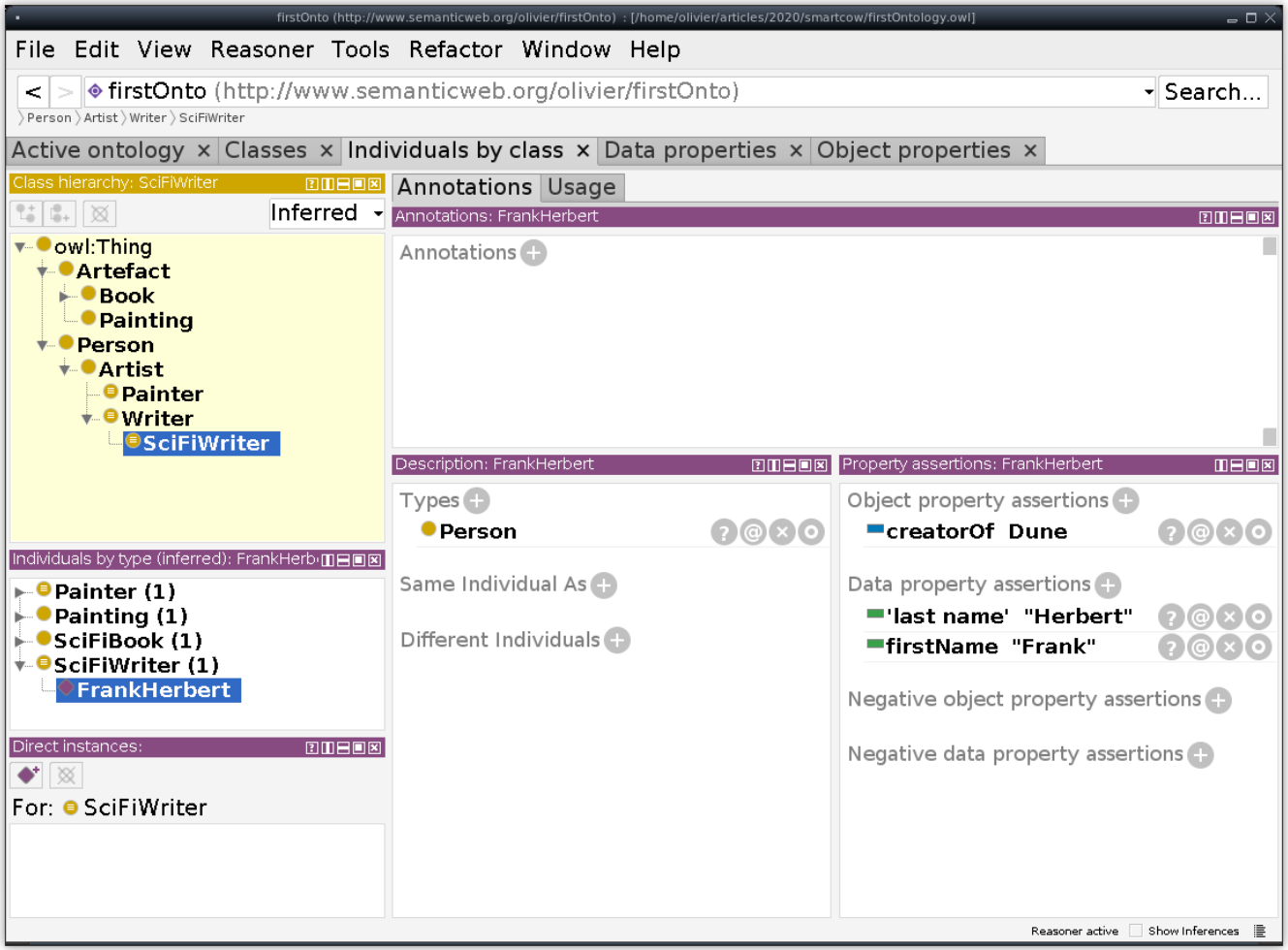
- let the reasoner organize the classes for you!

Make Dune an instance of SciFiBook...



- an individual can be an instance of multiple classes

... automatically makes Frank Herbert a SciFiWriter



Synthesis

Synthesis

- describe your data as precisely as possible
- formalize your ontology as finely as possible
  - disjoint classes help detect inconsistencies
    - in your ontology (cf. `ImpossibleClass`)
    - in your data (cf. `impossiblePerson`)
  - necessary constraints (cf. `Artist`)
  - definitions (cf. `Painter` and `Writer`)
- let the reasoner do the tedious part of ontology maintenance
- enjoin the new capability to query your data

SmartCow

an integrated infrastructure for increased research capability and innovation in the European cattle sector

Webinar:

Using ontologies to improve animal science research

15-16 December 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n°730924

Program of the webinar

Wednesday 16 December morning 10:00-12:00

- 10:00-10:55 Presentation of French ontologies from INRAE
  a. ATOL ontology by Catherine Hurtaud, INRAE (France) (25 min)
  b. EOL ontology by Jérôme Bugeon, INRAE (France) (15 min)
  c. AHOL ontology by Marie-Christine Meunier-Salaün, INRAE (France) (15 min)

- 10:55-11:15 Example of application/use of an ontology AHOL for professional use, by Marie-Christine Meunier-Salaün, INRAE (France) (20 min)

- 5 min break

- 11:20-12:00 Text based annotation with ontologies, Claire Nédellec INRAE (France) (40 min)

SmartCow

Animal Trait Ontology for Livestock: ATOL

ATOL project

INRAE motivations

- Promote a common culture for INRAE scientists using animal models and different approaches
- Develop a common language with geneticists
- Have an ontological base of the traits of interest for production animals and associated measurement techniques, in order to promote collaborative projects, referenced publications, data management and help teaching
- Actively participate in the implementation of the VT-USA database by introducing generic animal traits used in Europe

Motivations

J Anim Sci. 2008 June ; 86(6): 1485–1491. doi:10.2527/jas.2008-0930.

Animal trait ontology: The importance and usefulness of a unified trait vocabulary for animal species

L. M. Hughes*, J. Bao†, Z.-L. Hu*, V. Honavar†, and J. M. Reecy*

* *Department of Animal Science, Center for Integrated Animal Genomics, Iowa State University, 2255 Kildee Hall, Ames 50011*

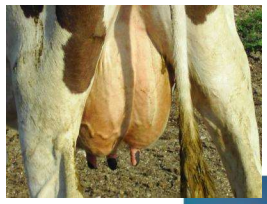
† *Department of Computer Science, Artificial Intelligence Research Laboratory, Iowa State University, 211 Atanasoff Hall, Ames 50011*

- Meeting with James Reecy (december 2008): presentation of his project VT sustained by **USDA-National Animal Genome Research = objective of Production**
- Program is aimed at the development of a **standardized trait ontology for farm animals and software tools** to assist the research community in collaborative creation, editing, maintenance, and use of such an ontology

SmartCow

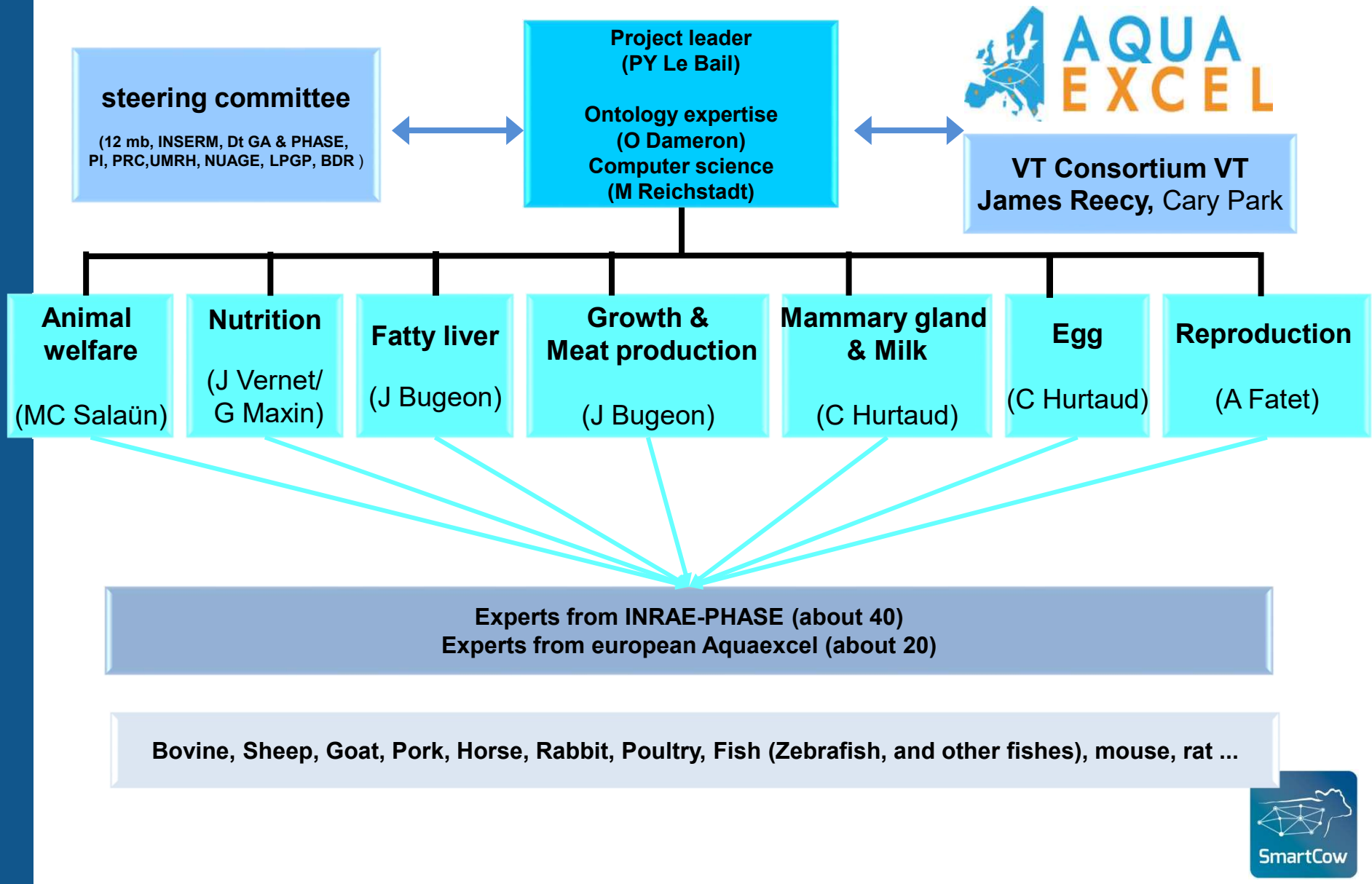
The ATOL project at the origin: the objectives

- Have a **reference ontology** for the phenotyping of farm animals and shared by the international scientific and teaching community
- Have a **language** usable by **computer programs** (database management, semantic analysis, modeling, etc.)
- Have the most **generic** traits possible for vertebrates
- Make the ontology as **operational** as possible and close to measurement techniques (collectors, users)
- Structure the base for **production** purposes



SmartCow

ATOL: Structure and network



ATOL: the complete hierarchy

ATOL: the complete hierarchy

- animal trait of livestock
  - animal welfare trait
  - egg trait
  - fatty liver trait
  - growth and meat production trait
  - mammary gland and milk production trait
  - nutrition trait
  - reproduction trait

- animal welfare trait
  - animal performance trait
    - animal production trait
    - body conformation trait
    - growth trait
  - health trait
    - disease susceptibility trait
    - health body trait
    - immunity system trait
  - psychoneurophysiological state trait
    - behaviour trait
    - biological rhythm trait
    - cognitive function trait
    - emotional and stress function tra
    - pain response trait
    - reflex response trait
    - sensory capacities trait

SmartCow

ATOL: the complete hierarchy

- animal trait of livestock
  - animal welfare trait
  - egg trait
  - fatty liver trait
  - growth and meat production trait
  - mammary gland and milk production
  - nutrition trait
  - reproduction trait

- egg trait
  - egg formation trait
    - hormonal mechanisms°°°
    - organ°°°
  - egg production trait
    - egg morphological production trai
    - egg quantitative production trait °
  - egg quality trait
    - egg biological activity trait
    - egg composition trait
    - egg mechanical properties trait
    - egg organoleptic trait
    - egg technological trait

SmartCow

ATOL: the complete hierarchy

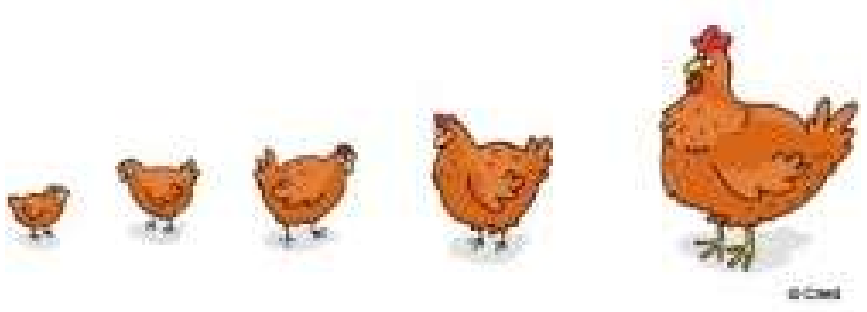
- animal trait of livestock
  - animal welfare trait
  - egg trait
  - fatty liver trait
  - growth and meat production trait
  - mammary gland and milk production tra
  - nutrition trait
  - reproduction trait

- fatty liver trait
  - fatty liver physiology
    - fatty liver enzyme activity
  - fatty liver quality
    - fatty liver nutritional quality
    - technological and organoleptic qu

SmartCow

ATOL: the complete hierarchy

- animal trait of livestock
  - animal welfare trait
  - egg trait
  - fatty liver trait
  - growth and meat production trait
  - mammary gland and milk production trait
  - nutrition trait
  - reproduction trait



- growth and meat production trait
  - adipose tissue trait
  - body slaughter trait
  - carcass quality trait
  - growth trait
  - meat quality trait
  - muscular system trait

SmartCow

ATOL: the complete hierarchy

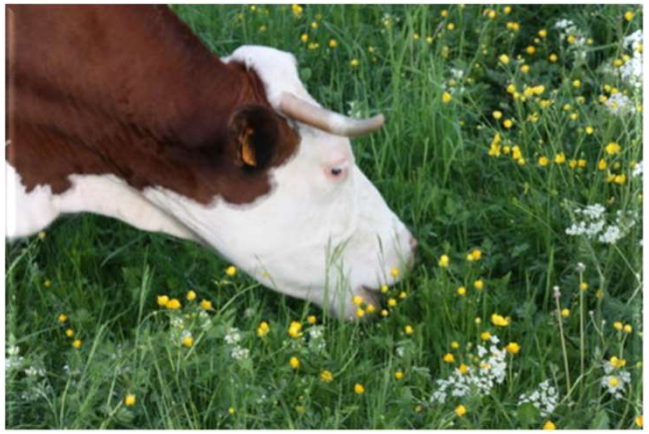
- animal trait of livestock
  - animal welfare trait
  - egg trait
  - fatty liver trait
  - growth and meat production trait
  - mammary gland and milk production trait
  - nutrition trait
  - reproduction trait

- mammary gland and milk production tr
  - mammary gland production trait
    - colostrum trait
    - milk ejection trait
    - milk trait
  - mammary gland trait
    - mammary gland microbiology trai
    - mammary gland morphology trait
    - mammary gland physiology trait

SmartCow

ATOL: the complete hierarchy

- animal trait of livestock
  - animal welfare trait
  - egg trait
  - fatty liver trait
  - growth and meat production trait
  - mammary gland and milk production trait
  - nutrition trait
  - reproduction trait



- nutrition trait
  - animal feed efficiency
  - animal requirement
  - digestive system morphology
  - digestive system physiology
  - ingestion
  - nutrient balance
  - supplementation

SmartCow

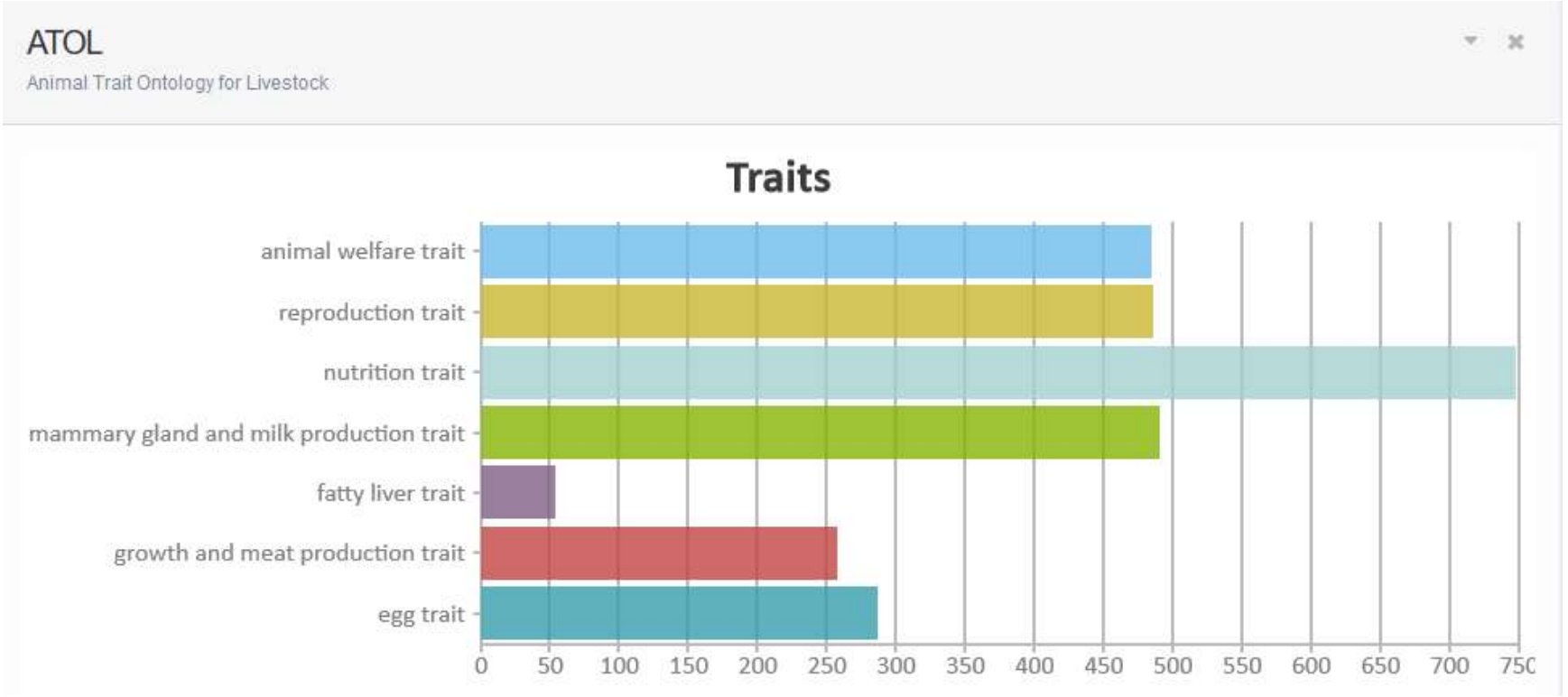
ATOL: the complete hierarchy

- animal trait of livestock
  - animal welfare trait
  - egg trait
  - fatty liver trait
  - growth and meat production trait
  - mammary gland and milk production trait
  - nutrition trait
  - reproduction trait

- reproduction trait
  - conception physiology
  - endocrine reproductive system
  - female reproductive feature
  - male reproductive feature
  - sex determination

SmartCow

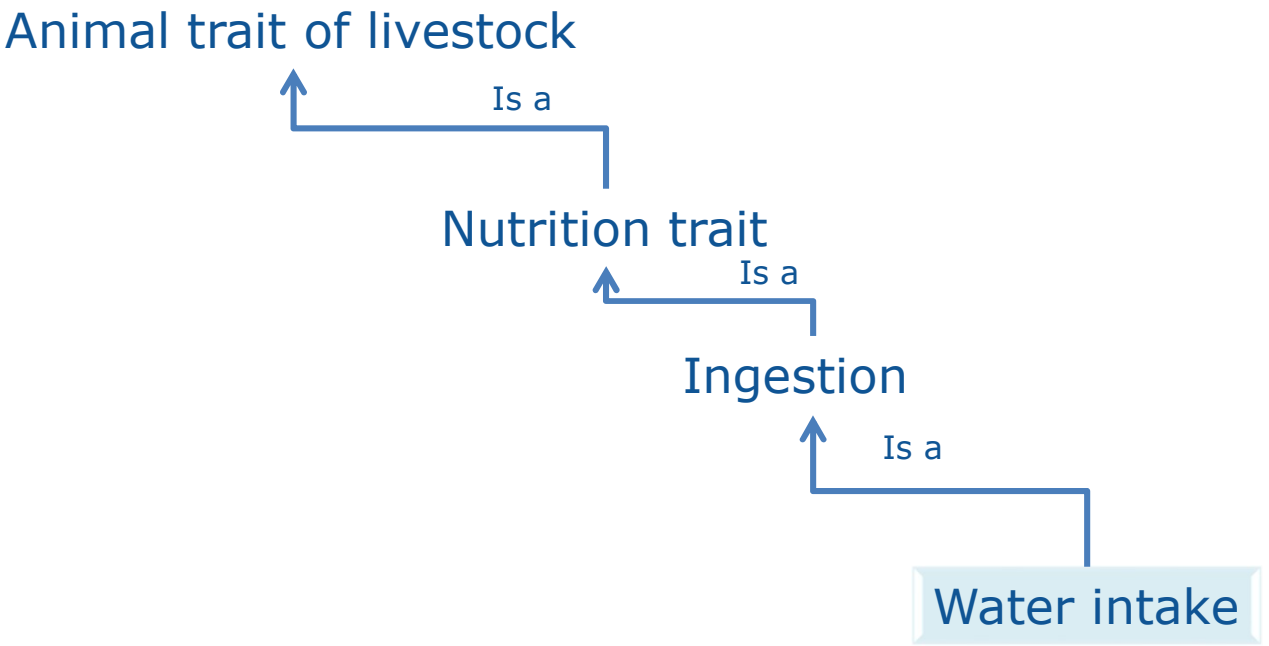
ATOL: number of traits



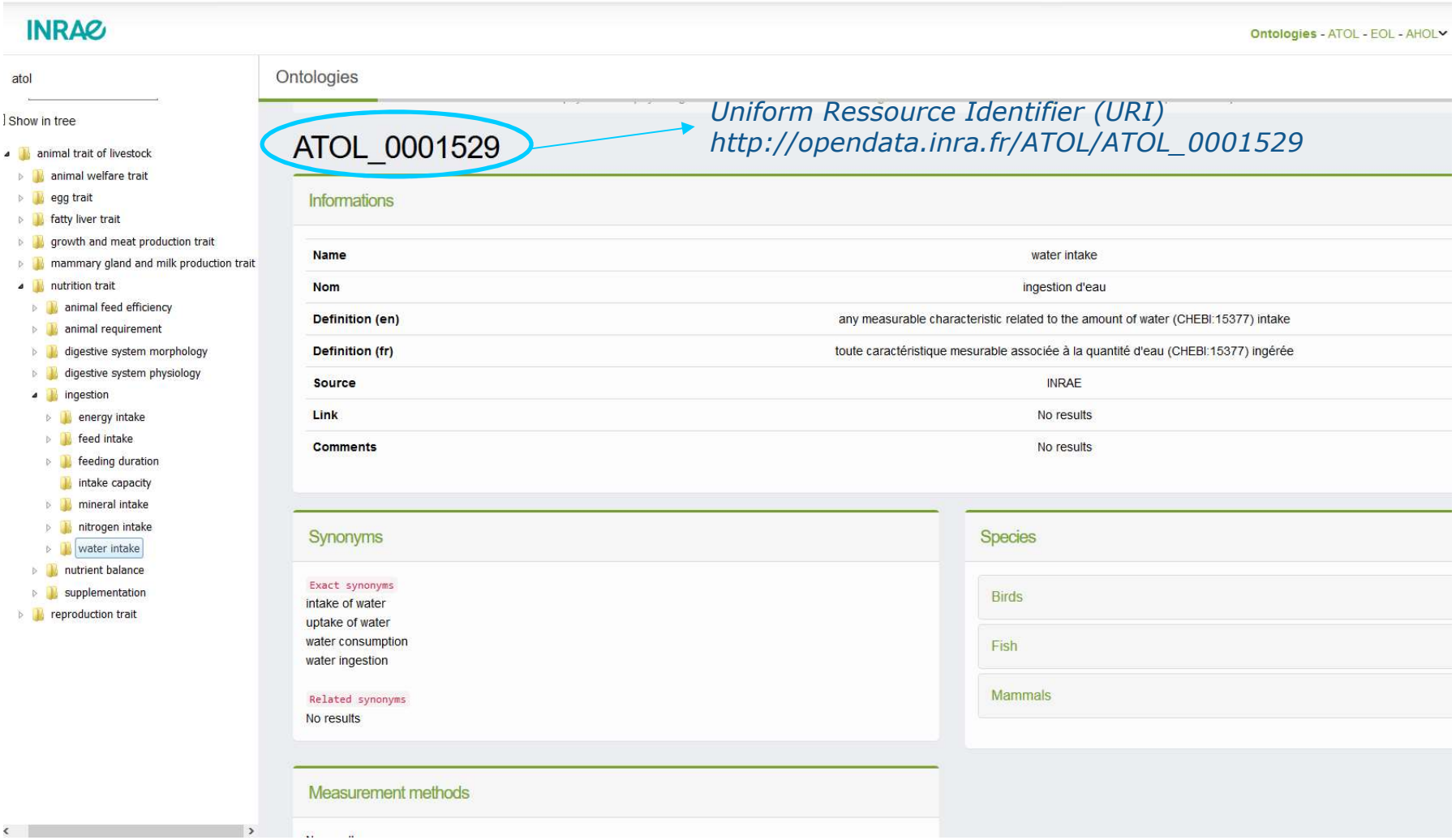
More than 3000 traits in September 2020

SmartCow

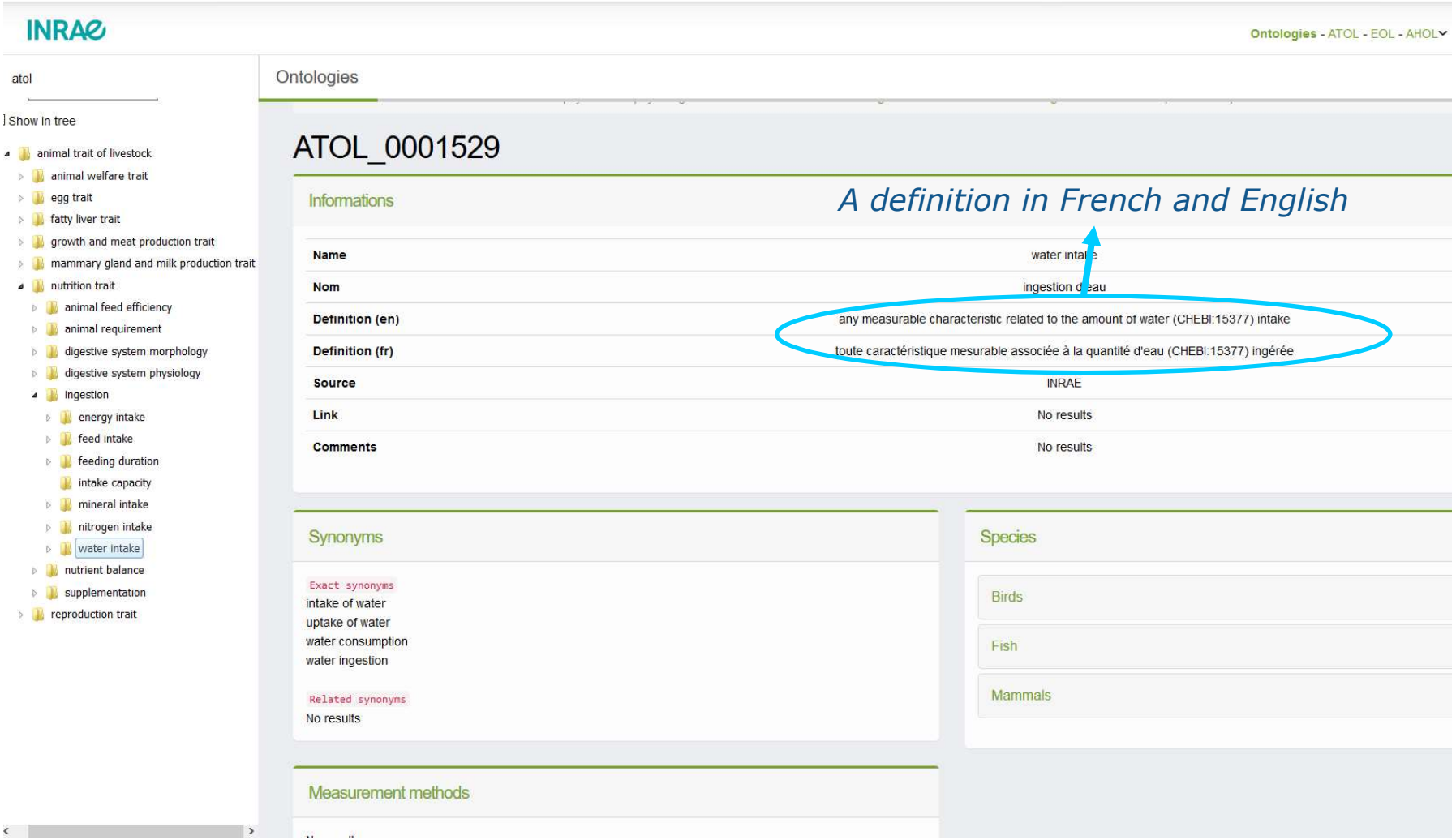
ATOL: structure of a trait « water intake »



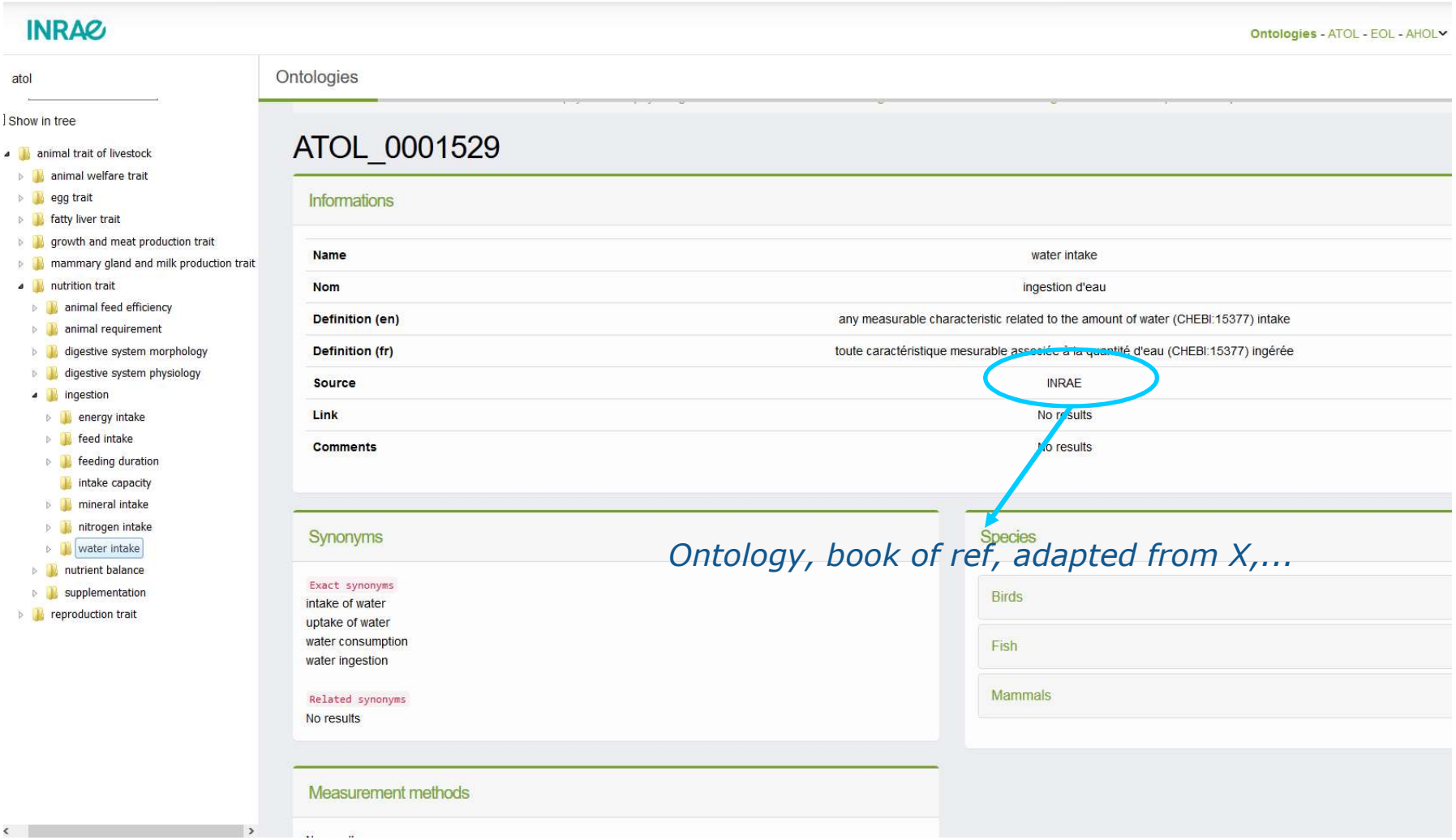
ATOL: structure of a trait « water intake »



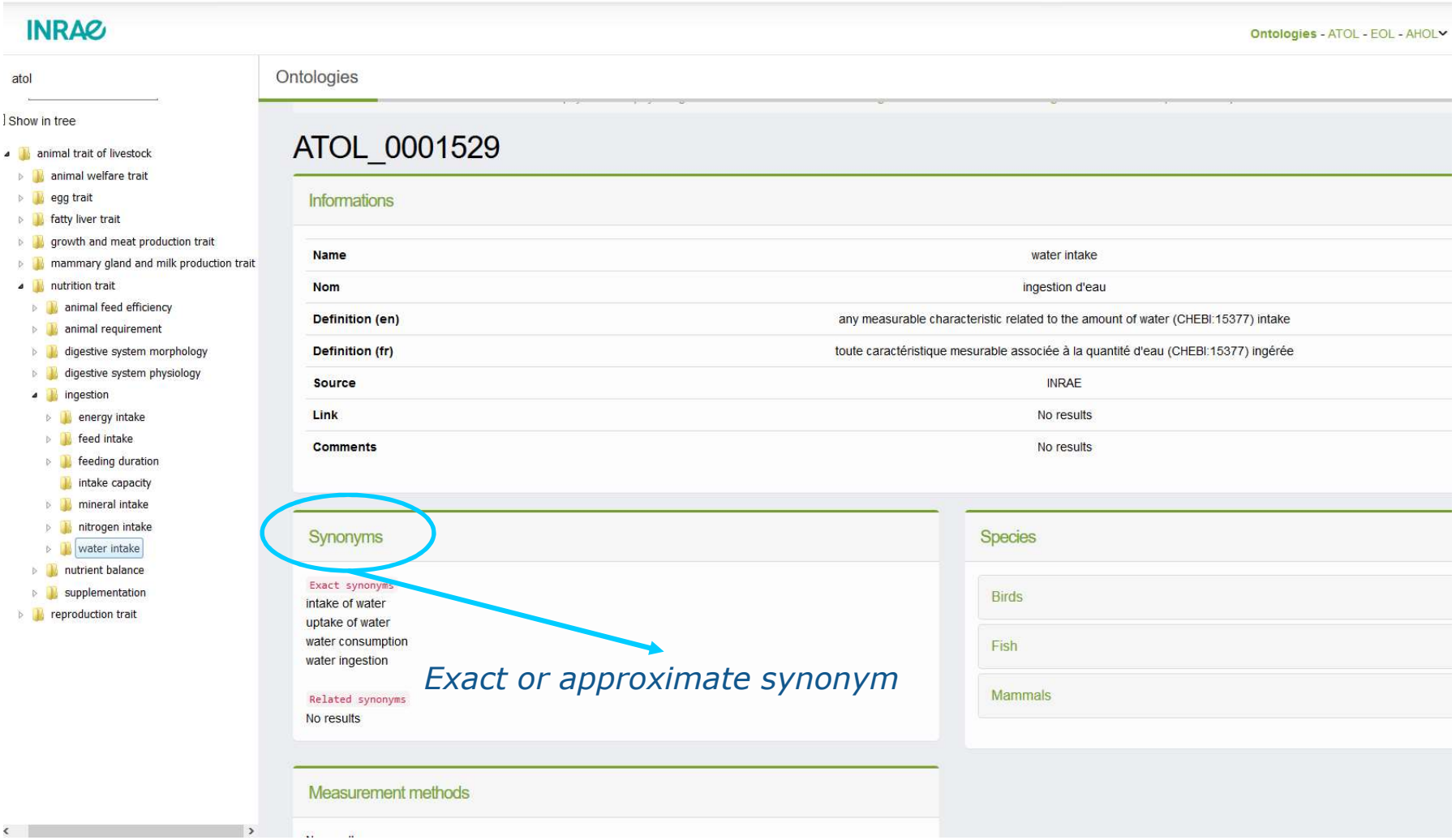
ATOL: structure of a trait « water intake »



ATOL: structure of a trait « water intake »

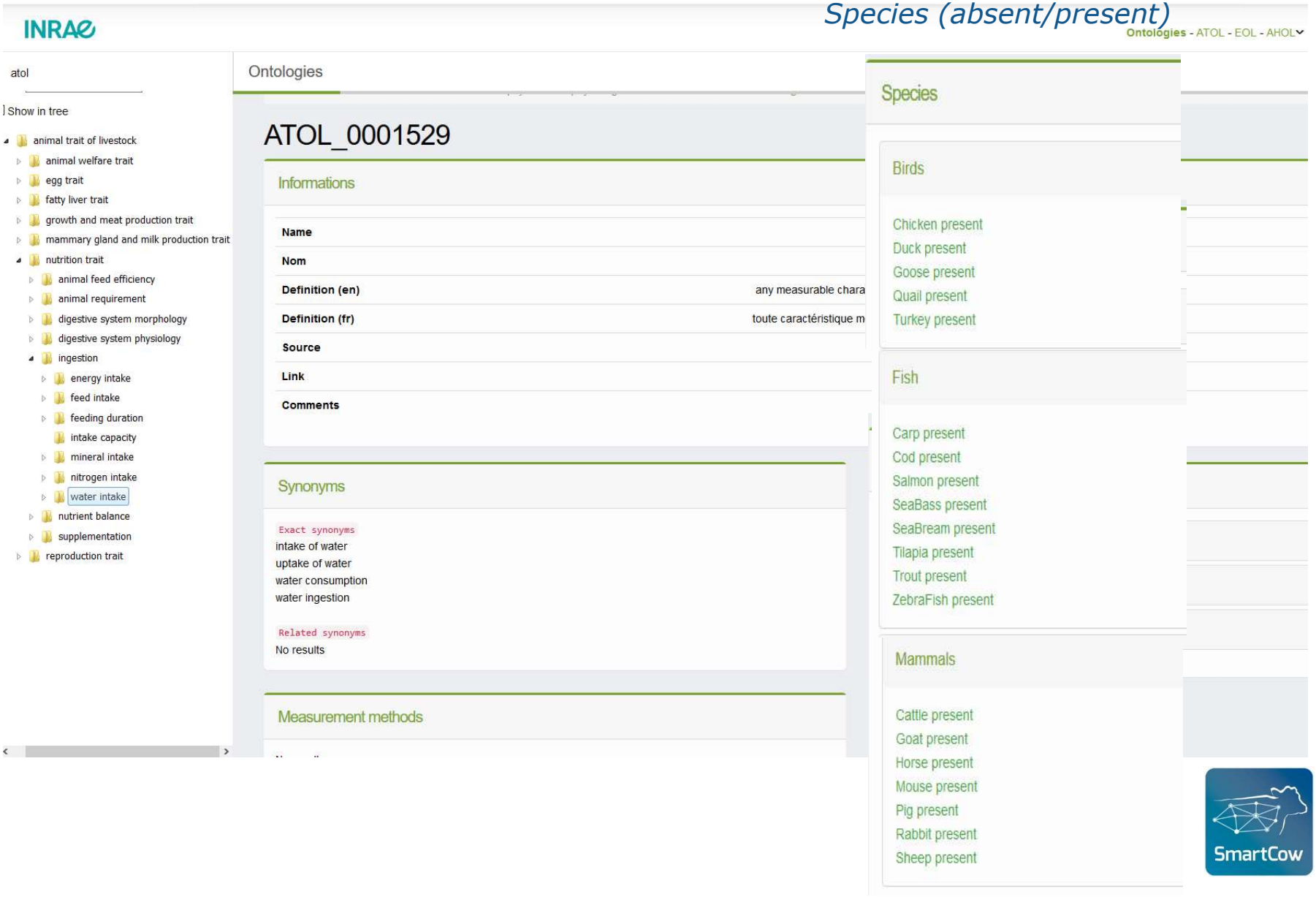


ATOL: structure of a trait « water intake »

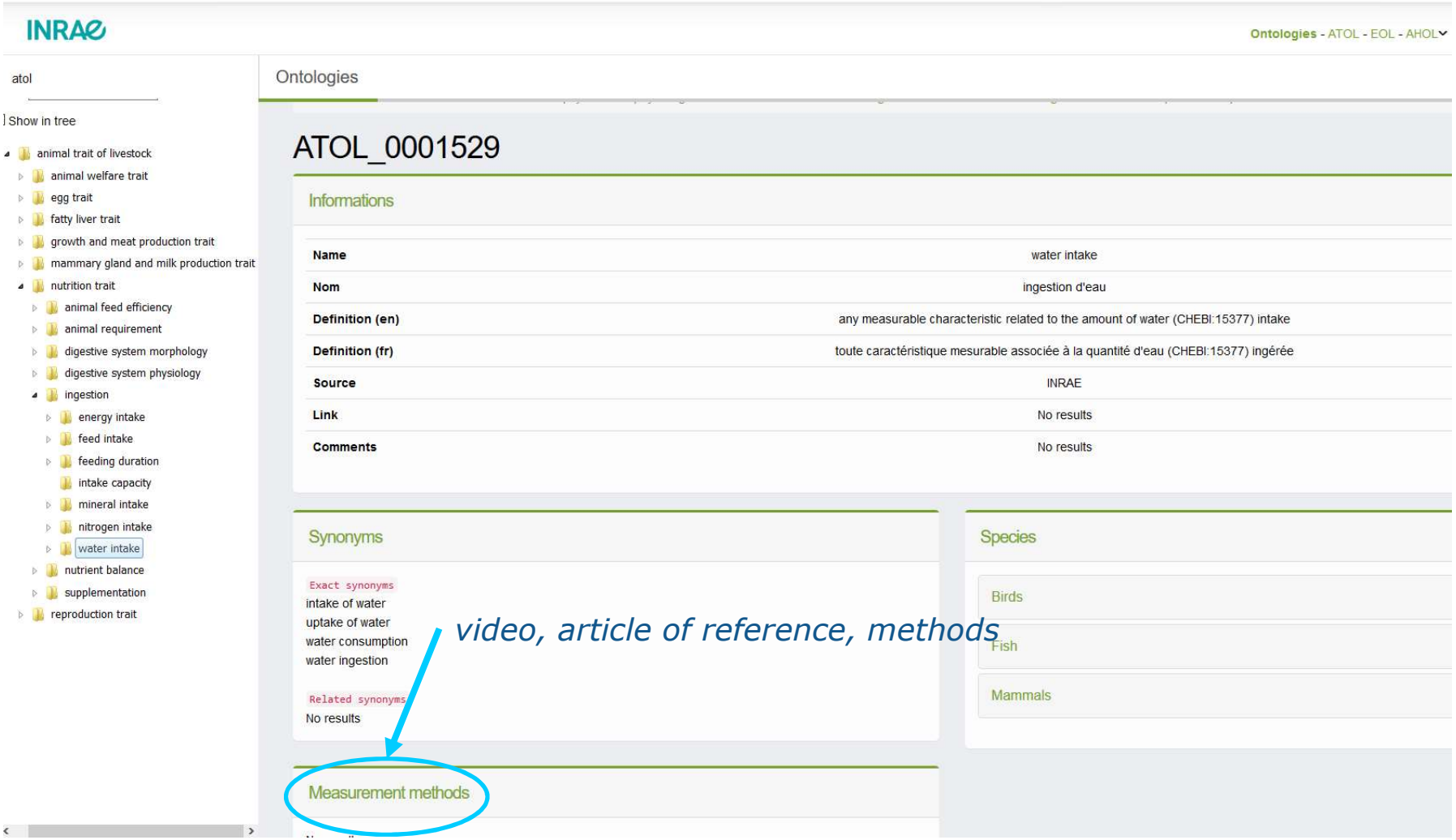


ATOL: structure of a trait « water intake »

Species (absent/present)



ATOL: structure of a trait « water intake »



Sadjad Danesh Mesgaran, Björn Kuhla
Institute of Nutritional Physiology "Oskar Kellner," Leibniz Institute for Farm Animal Biology (FBN), Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany

Guideline on measuring individual water intake in dairy cow (ATOL_0001529)

SmartCow

https://www.atol-ontology.com/en/erter-2/

SmartCow

an integrated infrastructure for increased research capability and innovation in the European cattle sector

Webinar:

Using ontologies to improve animal science research

8-9 October 2020

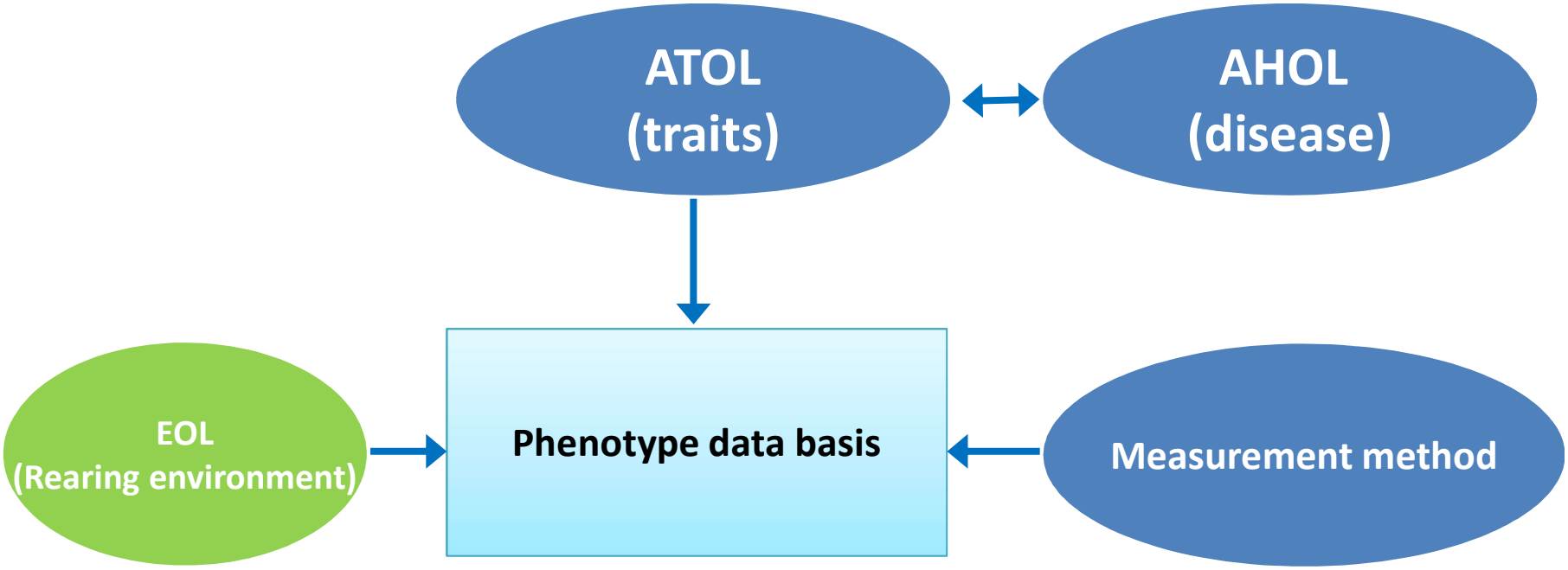
Jerôme BUGEON
INRAE LPGP Rennes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n°730924

Environment Ontology for Livestock: EOL

Why EOL ?

A phenotype depends on the genotype and the environnement
(*P = G + E + Epi*).



EOL: the hierarchy

Impact of the living/rearing environment on the phenotypic traits

INRAE

eol

Filter: Search...

Show in tree

- environment ontology for livestock
  - livestock farming environment
  - livestock farming structure
  - livestock farming system
  - livestock feeding

The proximal environnement

Rearing structure

Choice made by the farmer

Feeding (quantity, quality)

SmartCow

EOL: the hierarchy

The proximal environnement

- environment ontology for livestock
  - livestock farming environment
  - livestock farming structure
  - livestock farming system
  - livestock feeding

- environment ontology for livestock
  - livestock farming environment
    - rearing living environment
      - livestock abundance
        - animal biomass
        - animal density
      - livestock group characteristic
        - livestock group size
        - livestock group structure
      - rearing biocoenosis context
    - rearing physicochemical environme
      - air rearing environment
      - light rearing environment
      - soil environment
      - water rearing environment

What's around me ?

EOL: the hierarchy

- environment ontology for livestock
  - livestock farming environment
  - livestock farming structure
  - livestock farming system
  - livestock feeding

Rearing structure

- environment ontology for livestock
  - livestock farming environment
  - livestock farming structure
    - experimental structure equipment
      - respiration chamber
    - rearing structure equipment
      - animal handling facility
      - enrichment rearing equipment
    - rearing structure material
      - environmental material
    - rearing structure physical dimensio
      - aquaculture rearing structure ph
      - terrestrial rearing structure physi
    - rearing structure type
      - aquatic rearing structure
      - terrestrial rearing structure

What's around me ?

SmartCow

EOL: the hierarchy

- environment ontology for livestock
  - livestock farming environment
  - livestock farming structure
  - livestock farming system
  - livestock feeding

Choice made by the farmer

- livestock farming system
  - aquatic farming system
    - aquatic management system
    - aquatic rearing stage unit
    - biome support
    - product oriented aquatic rearing
    - species association system
  - precision livestock farming system
    - animal material
    - body evaluation material
  - terrestrial farming system
    - biome support
    - product oriented terrestrial rearir
    - species association system
    - terrestrial management housing
    - terrestrial rearing stage unit

How I conceived and managed my farm ?

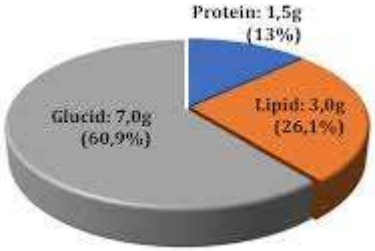
EOL: the hierarchy

Feeding (quantity, quality)

- environment ontology for livestock
  - livestock farming environment
  - livestock farming structure
  - livestock farming system
  - livestock feeding

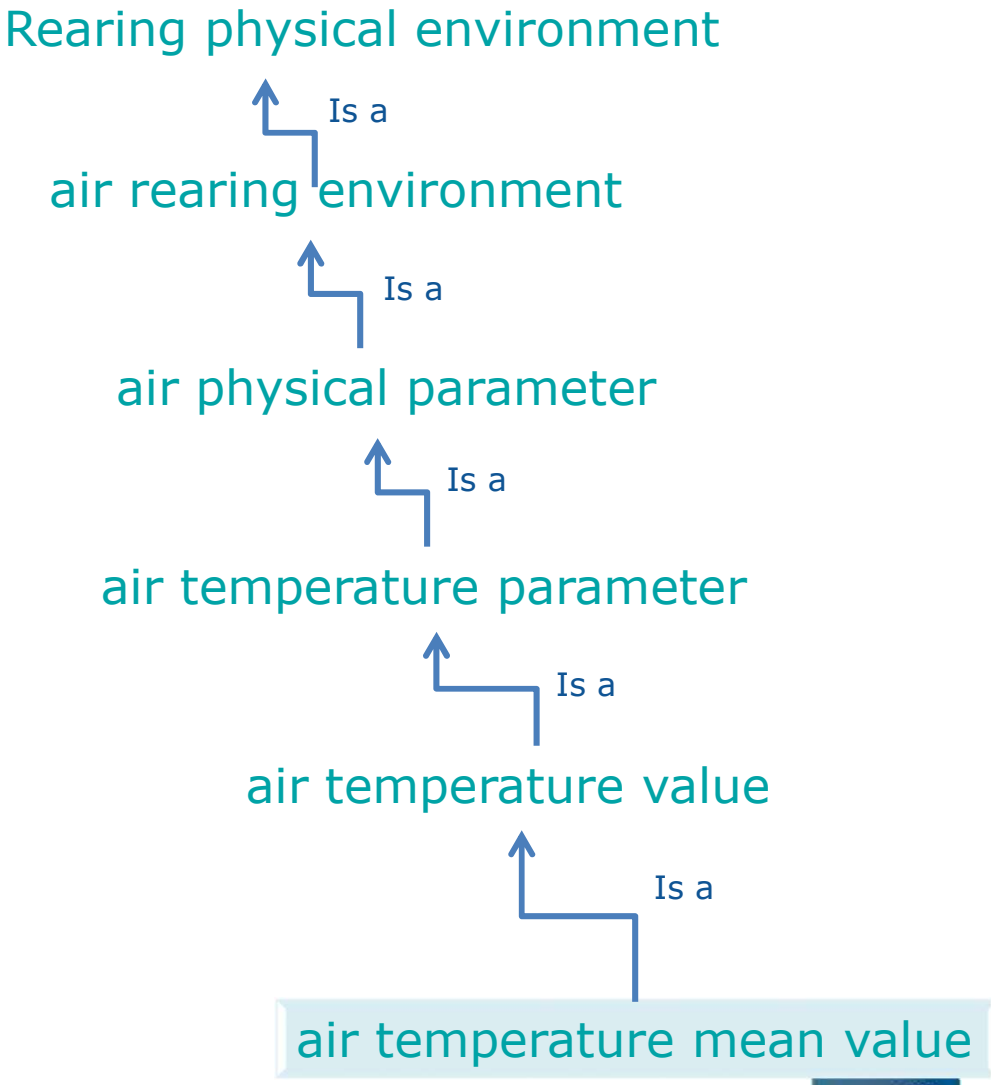
- livestock feeding
  - diet change
  - diet characteristic
    - dietary relative composition
  - diet delivery
    - feed delivery
    - milk delivery
    - water delivery
  - feed characteristic
    - feed buffer ability
    - feed chemical composition
    - feed energy content
    - feed nature
    - feed pH
    - feed physical characteristic
    - feed preservative content

How I conceived and managed my farm ?

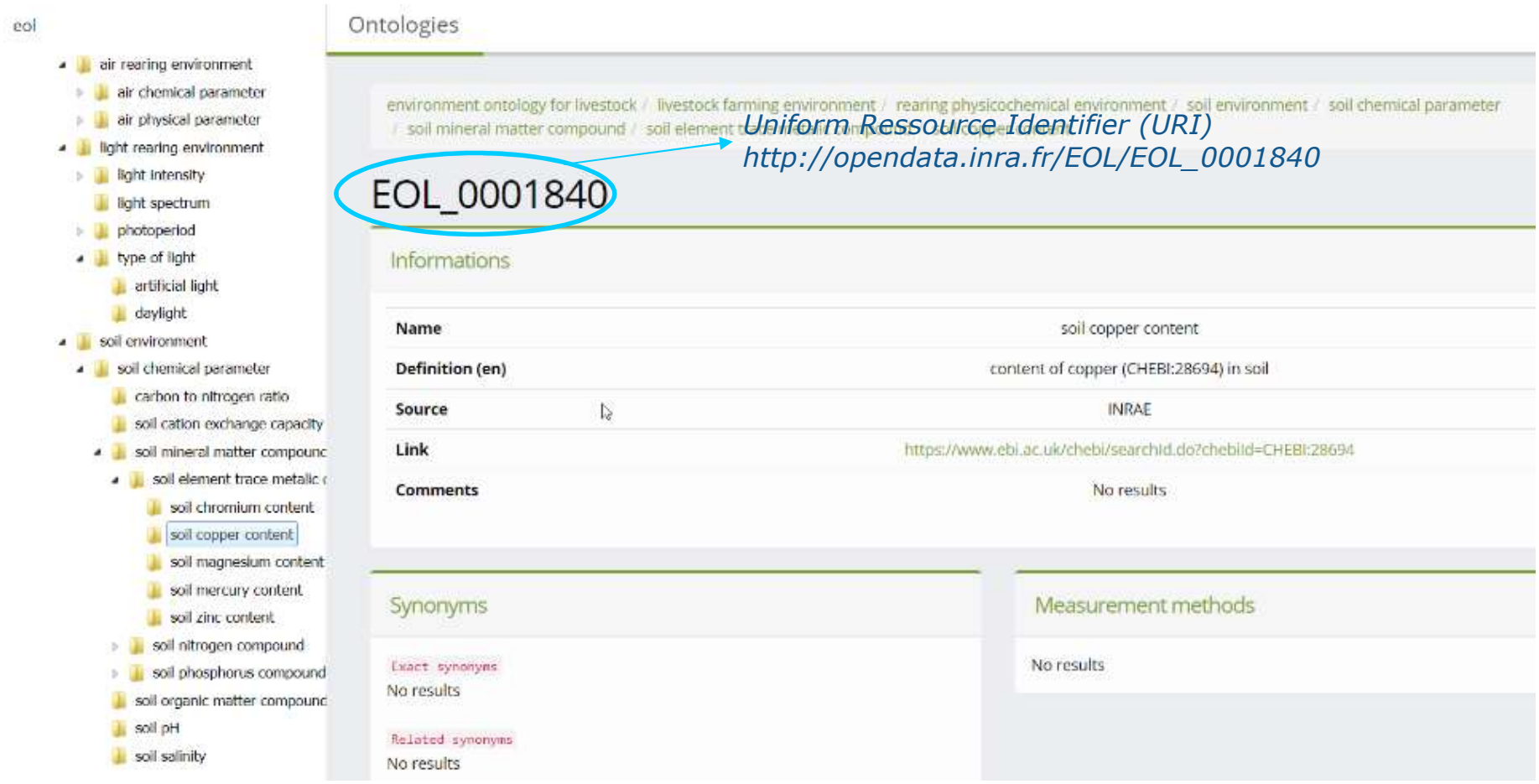


EOL: structure of a trait « air temperature mean value »

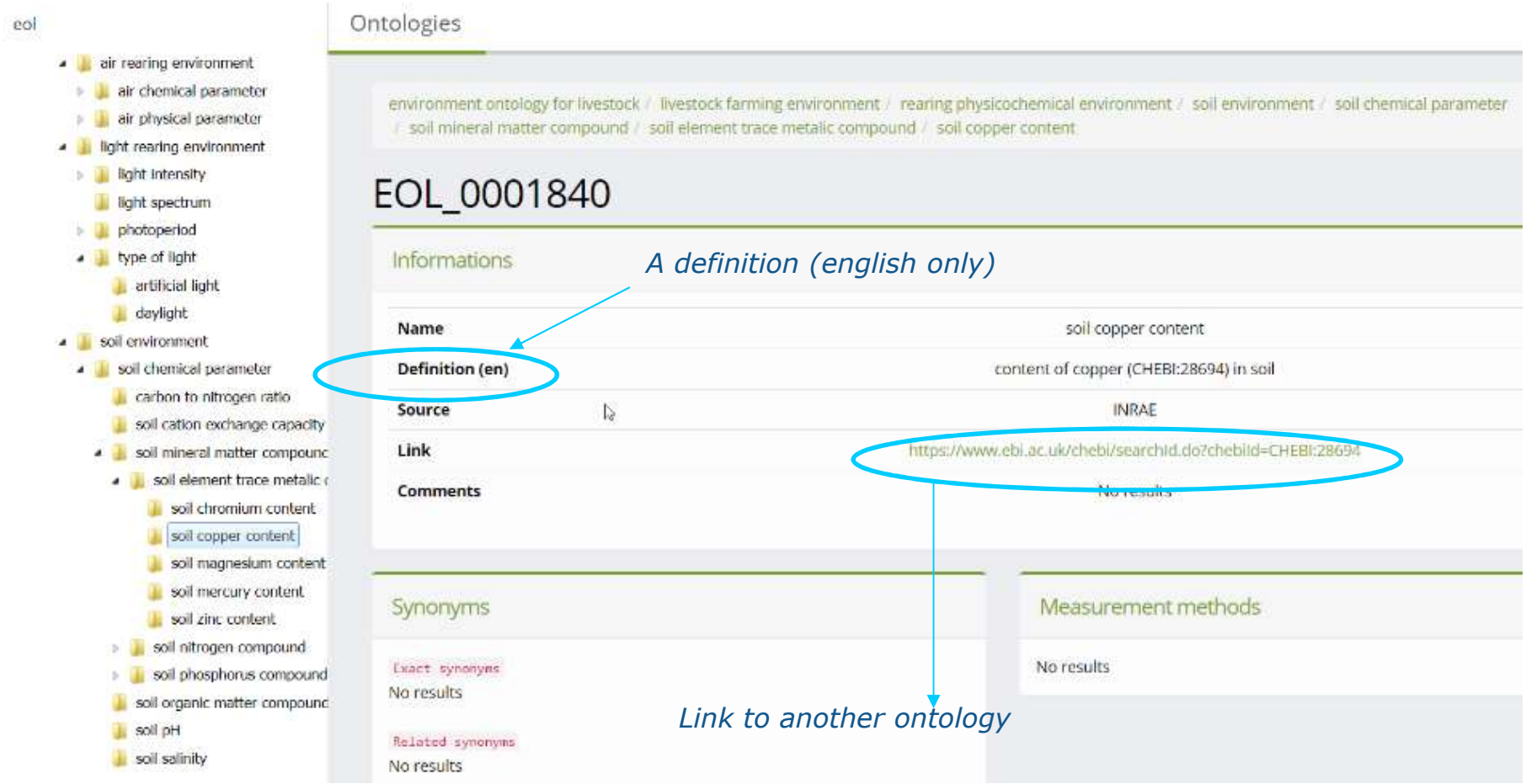
- rearing physicochemical environment
  - air rearing environment
    - air chemical parameter
    - air physical parameter
      - air density
      - air enthalpy
      - air movement
      - air particle content
      - air pressure
      - air temperature parameter
        - air dew temperature
        - air fluctuation temperature
        - air temperature value
          - air temperature maximum value
          - air temperature mean value
          - air temperature minimum value
          - instantaneous air temperature



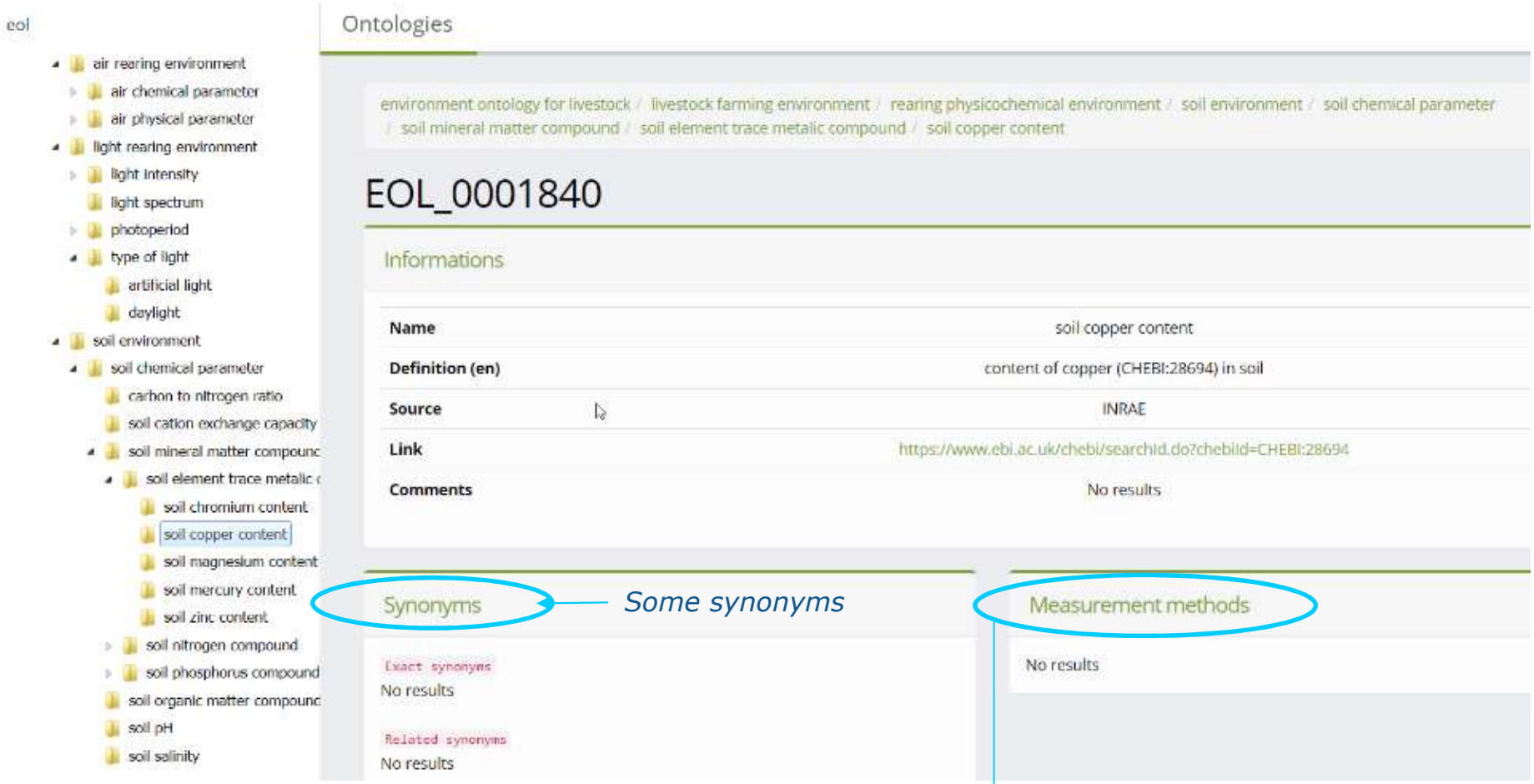
EOL: structure of a trait « soil copper content»



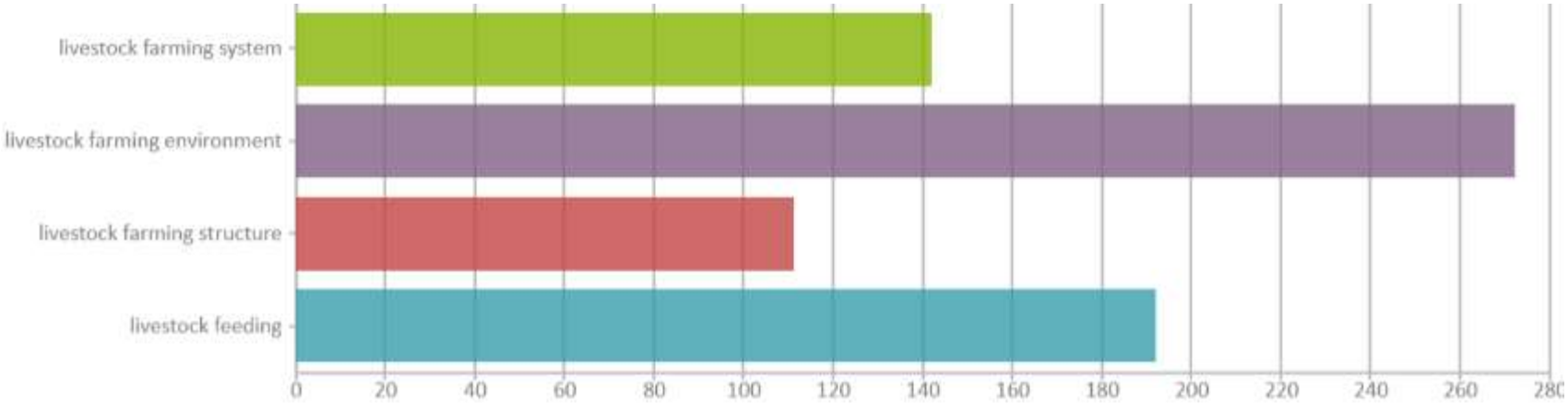
EOL: structure of a trait «soil copper content»



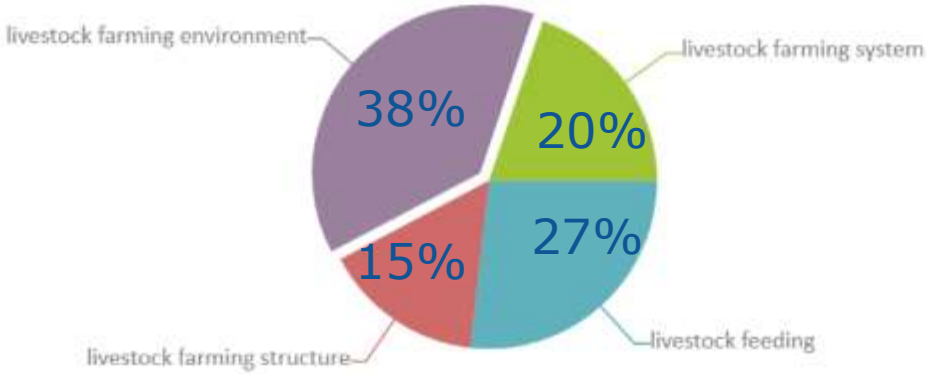
EOL: structure of a trait «soil copper content»



EOL: number of parameters



Traits



More than 700 parameters in December 2020

SmartCow

THANK YOU
FOR YOUR
ATTENTION

INRAE **Livestock Ontologies**

ATOL – EOL – AHOL

ONTOLOGIES ATOL STRUCTURE ACTORS BIBLIOGRAPHY

Eol

The EOL ontology describes environmental conditions of livestock farms. More specifically, it describes the feeding modalities, the environment, the structure of livestock farms and rearing systems.

SmartCow

SmartCow

an integrated infrastructure for increased research capability and innovation in the European cattle sector

Webinar:

Animal Health Ontology for livestock AHOL

MC Meunier-Salaün

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n°730924

AHOL Ontology

Objectives

- fine description of the health status of individuals and flocks in different species of animal production

➢ multispecies focus

➢ production disease : agent or factor involved
  * pathogen agent (infectious diseases)
  * environmental factors

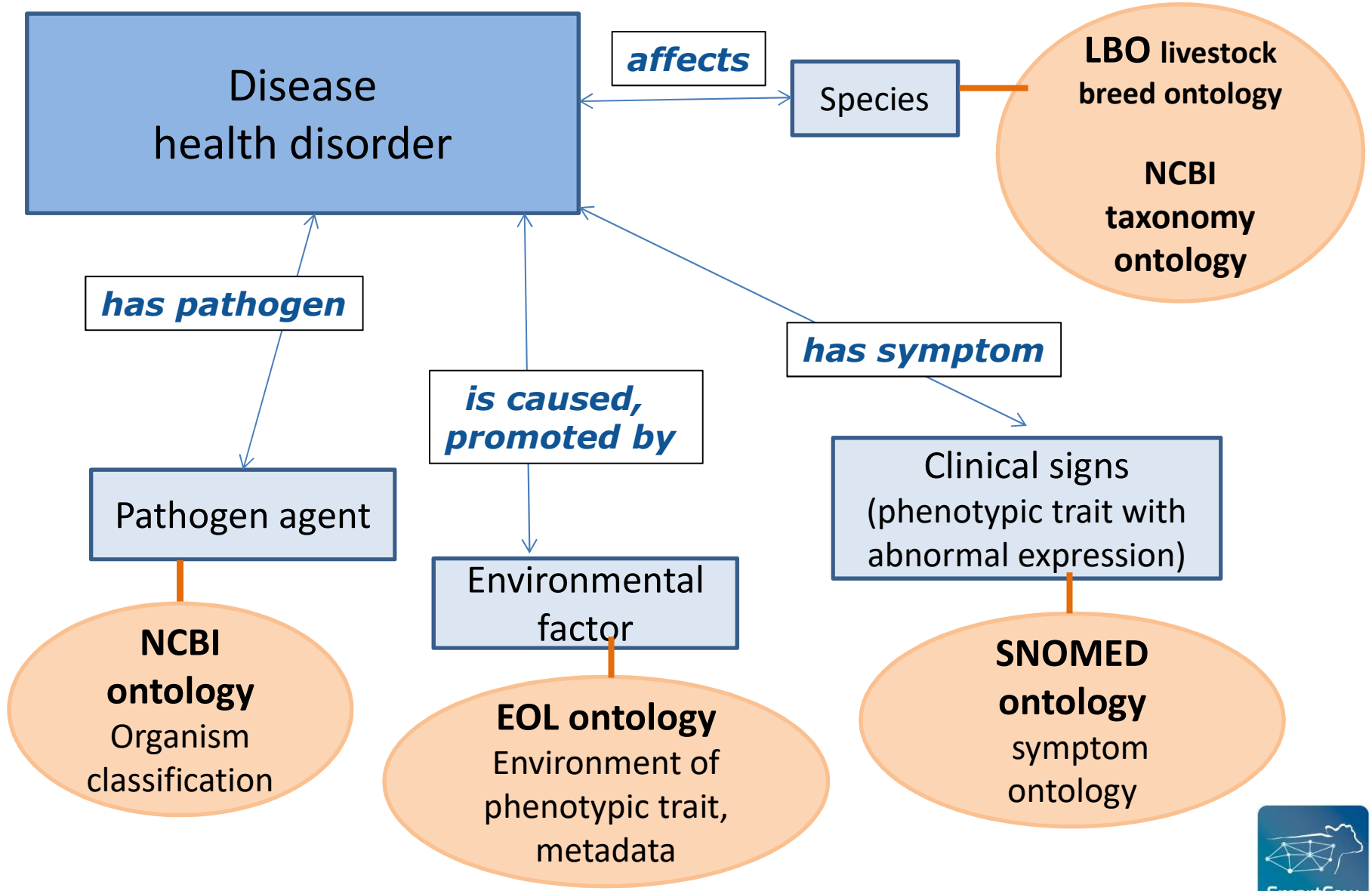
➢ clinical signs related to disease or health disorder

- Further : applications using ontologies dedicated to the animal production sector : health management of herd, diagnosis tool

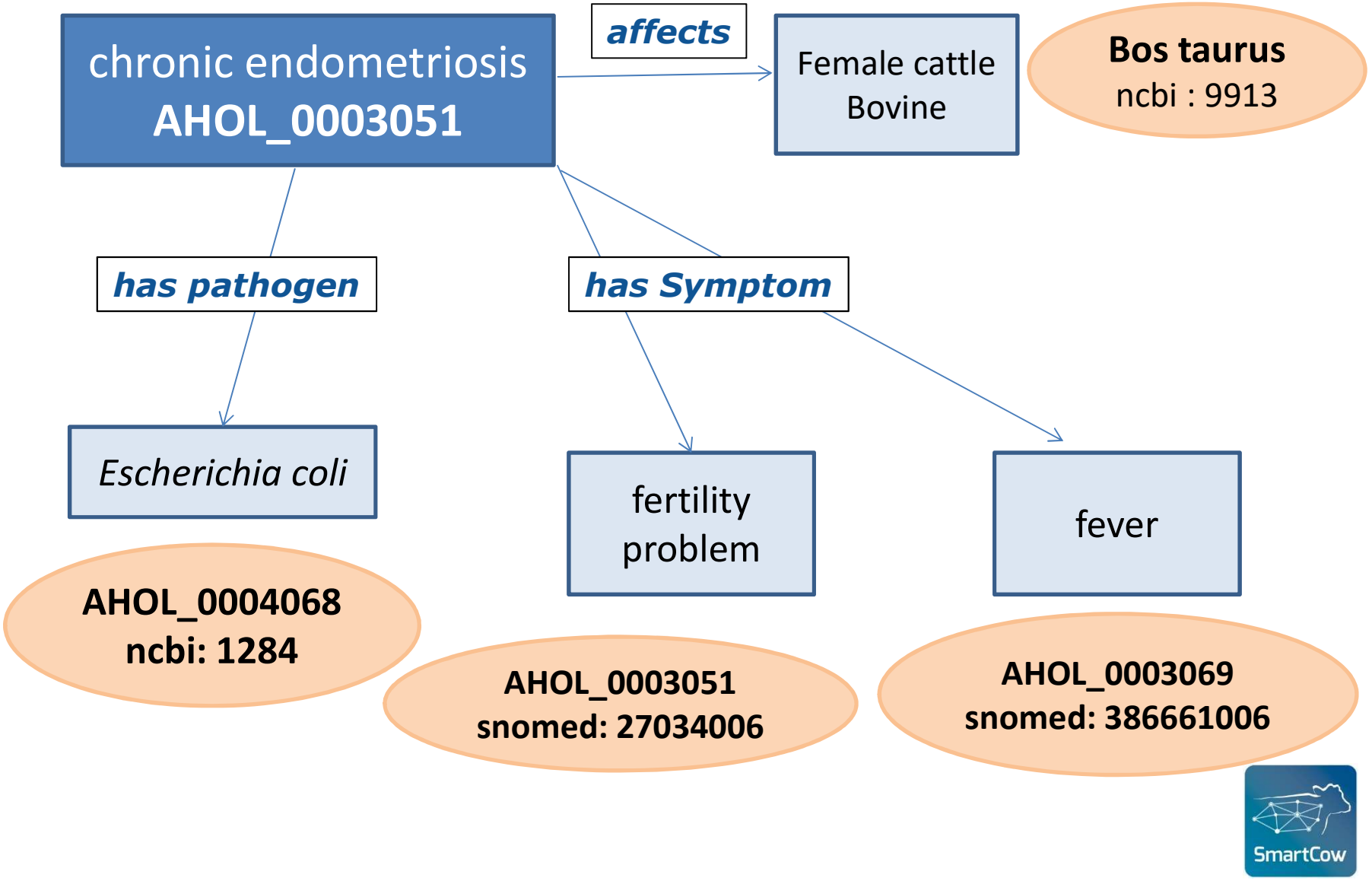
SmartCow

AHOL – perimeter focus on the disease/health disorder

properties : specie affected, causes, consequences

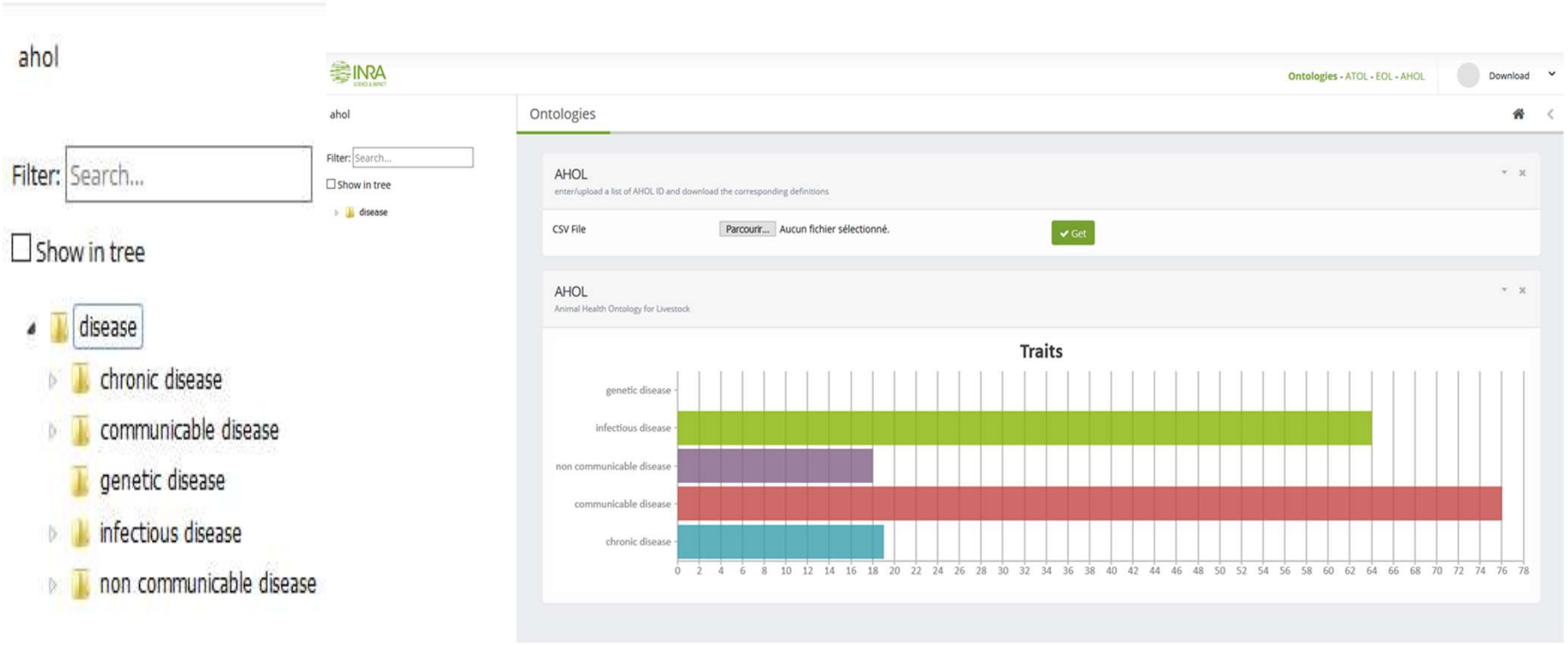


AHOL – *Example*



AHOL – hierarchy

main branches 177 maladies

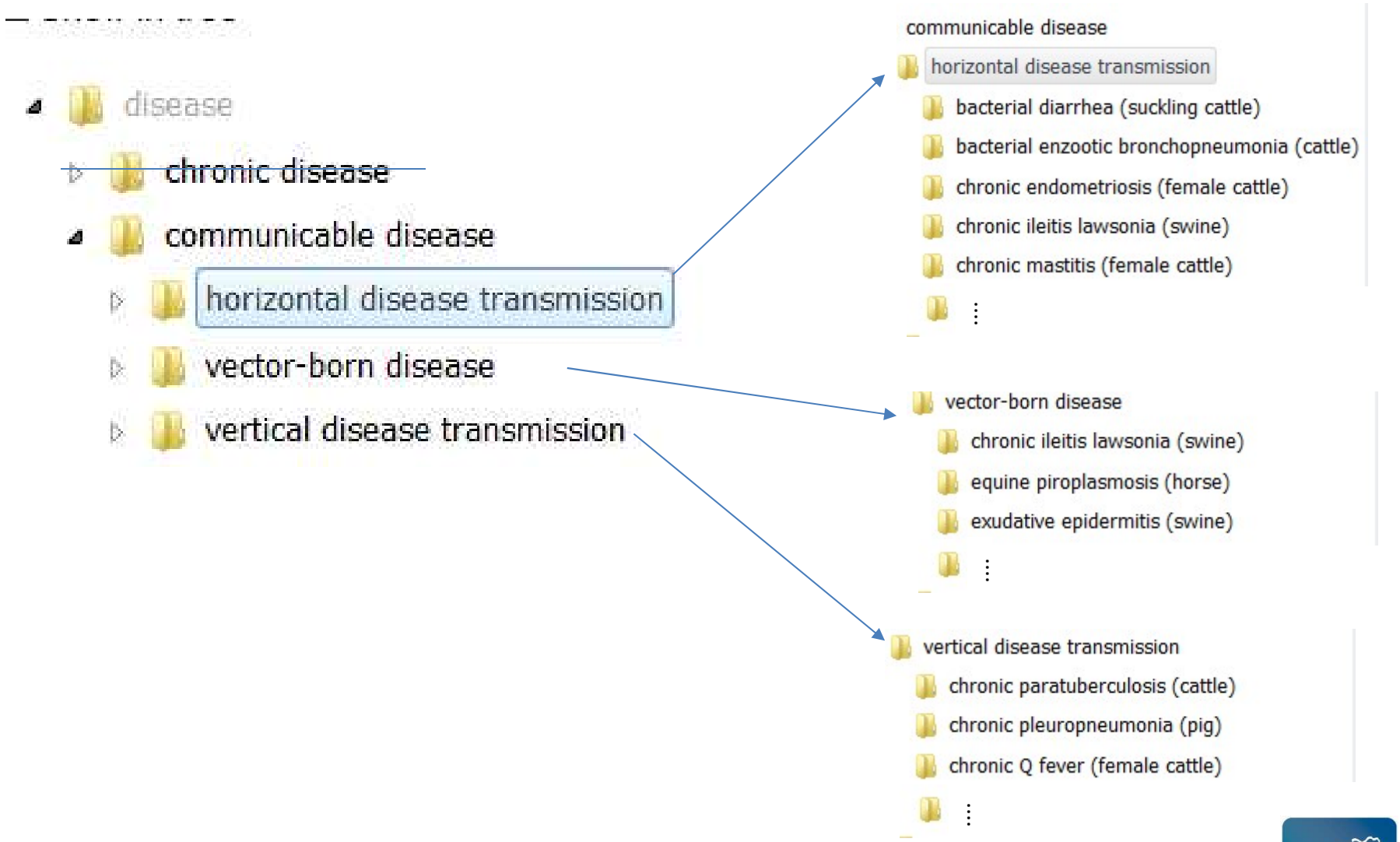


On going work to
*** enrich the hierarchy with new branches**
*** add new diseases**

especially genetic disease – non communicable disease

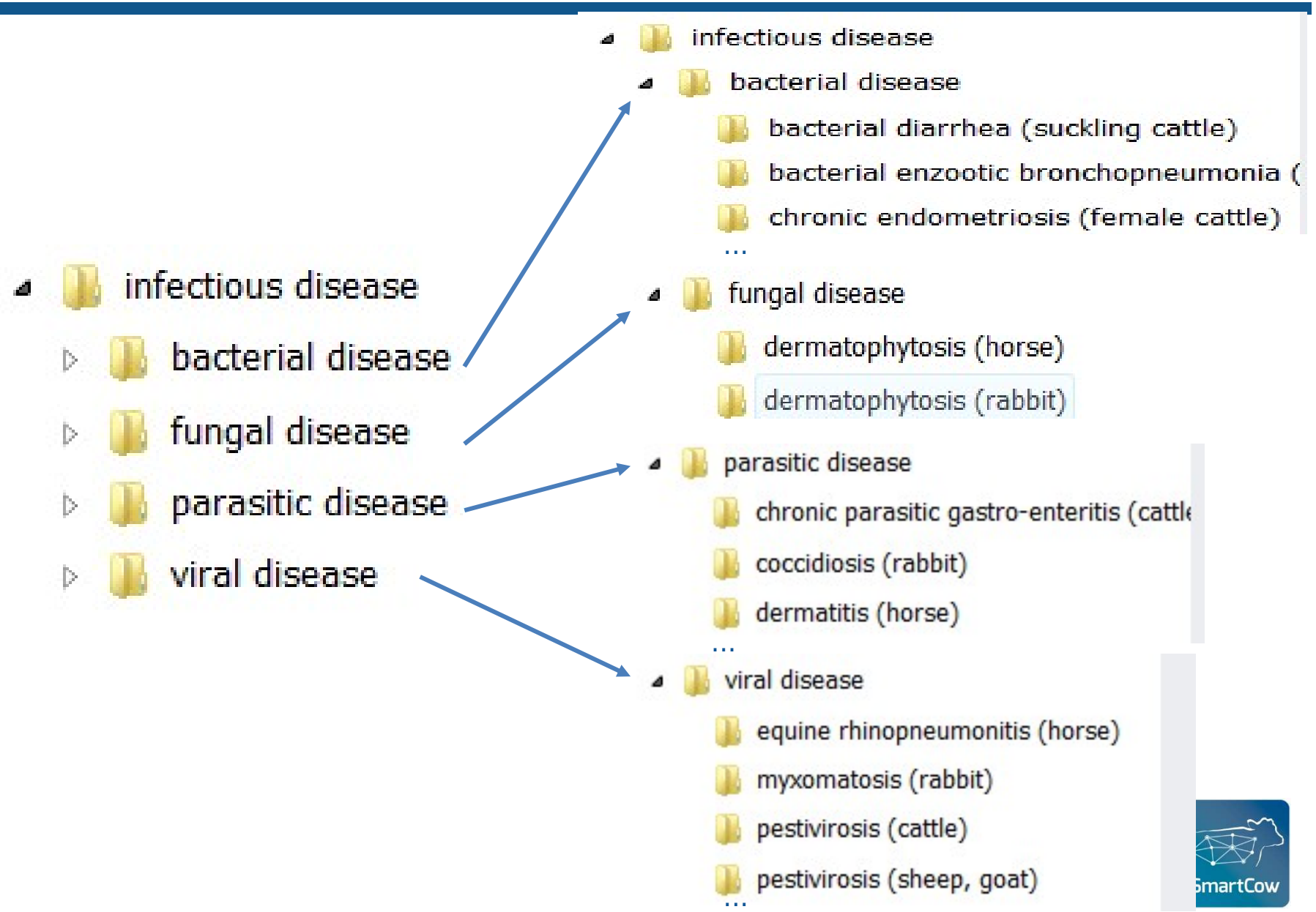
SmartCow

AHOL – communicable disease branch



SmartCow

AHOL – infectious disease branch

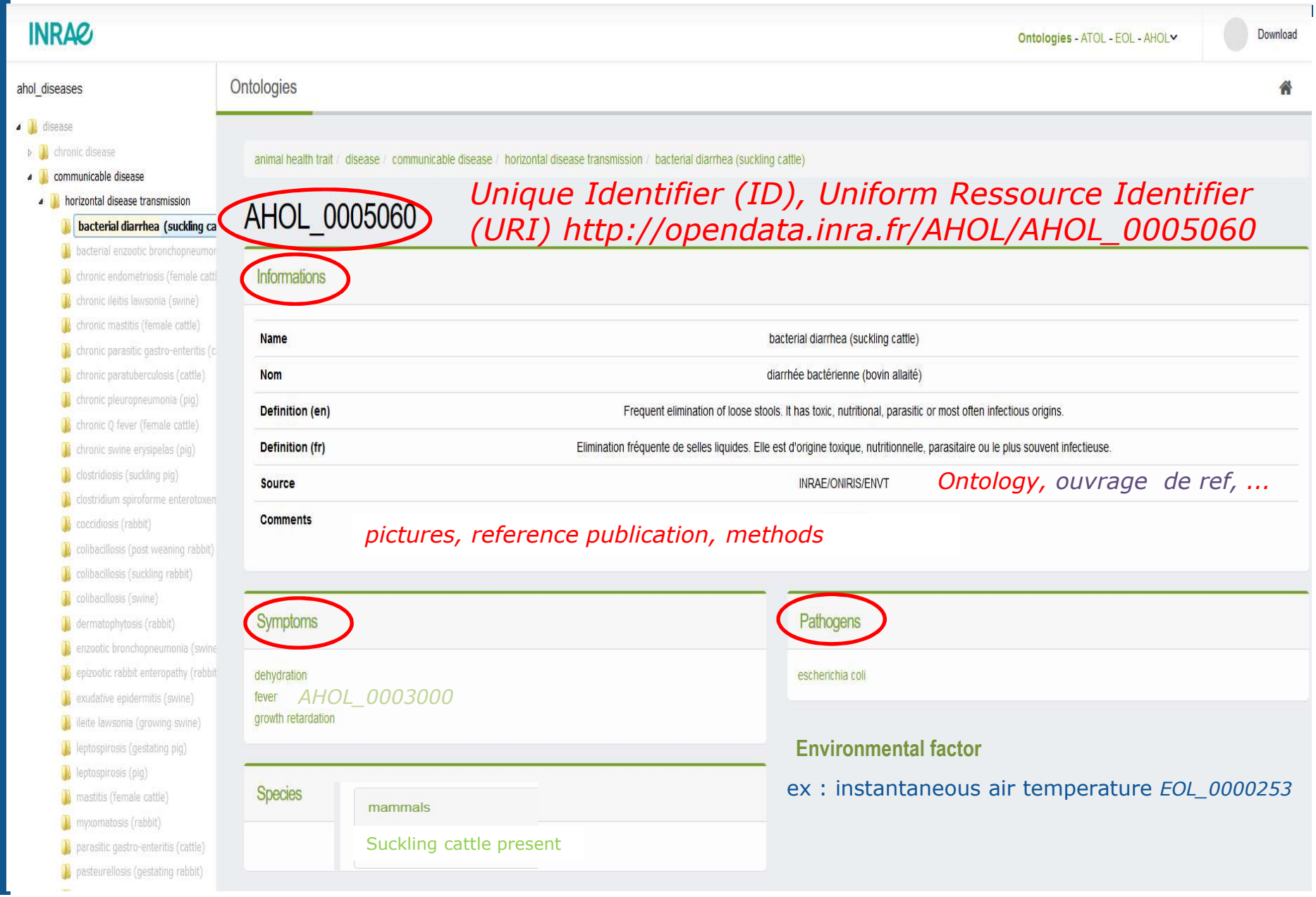


AHOL – non communicable disease branch

- non communicable disease
  - allergic disorder
  - metabolic disorder
  - mycotoxin-related disorder
  - nutritionnal disorder
    - nutriment-related disorder
    - micronutriment-related disorder
  - physical disorders
    - skin disorder
    - musculsqueletal disorder
    - neurological disorder
  - psychological disease
    - cognitive disorder
    - developmental disorder
  - toxic-related disorder
  - tumor cancer disease

AHOL – properties of disease

Example bacterial diarrhea



SmartCow – AHOL

Exercise

https://www.atol-ontology.com/ahol/

Request on disease / pathogen / symptom

bacterial diarrhea (suckling cattle) 2 matches

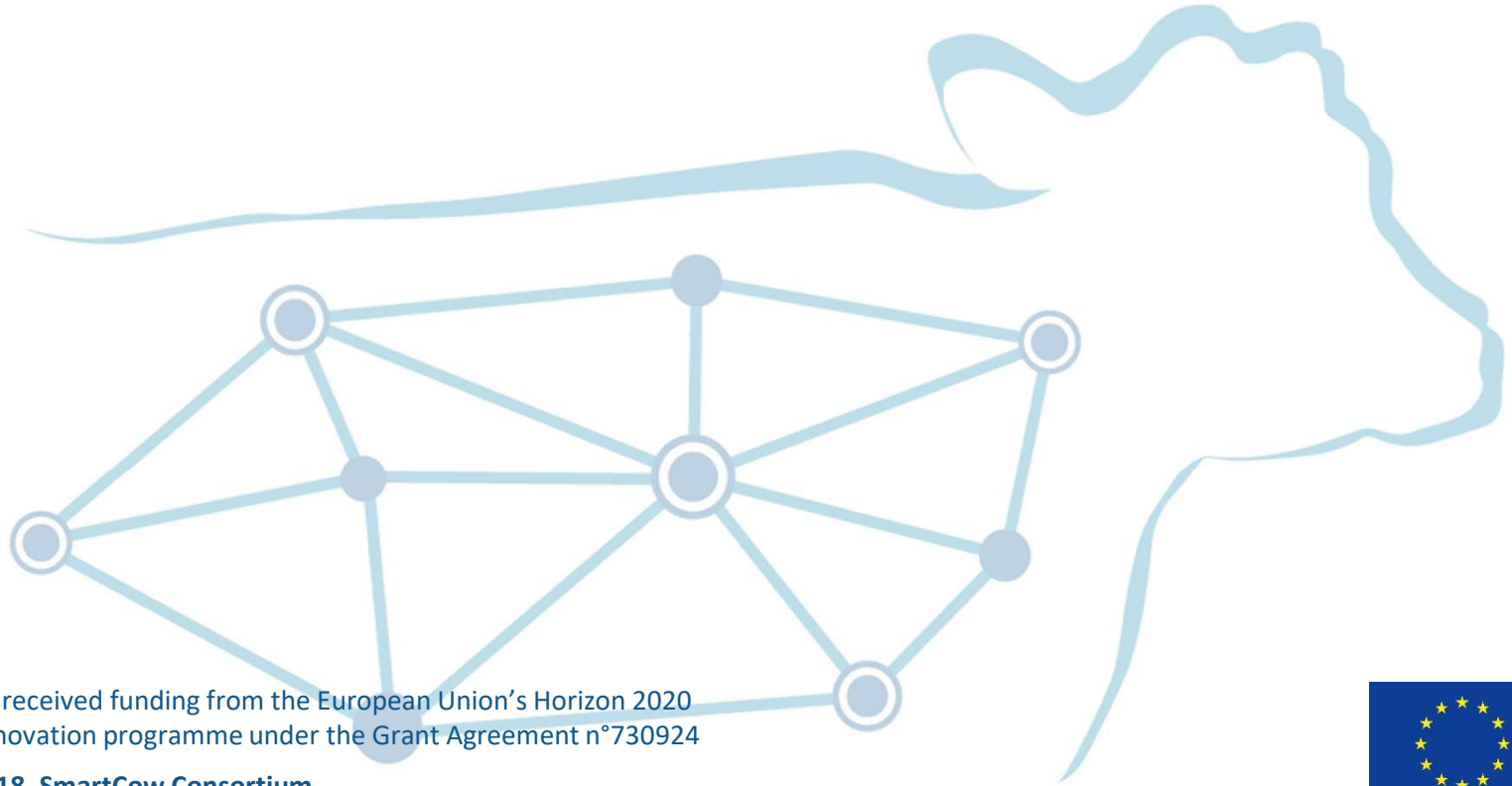
AHOL_0005060

- communicable disease
- horizontal disease transmission
- infectious disease
- bacterial disease

.....

SmartCow

Thanks for your attention



This project has received funding from the European Union's Horizon 2020
research and innovation programme under the Grant Agreement n°730924

SmartCow

an integrated infrastructure for increased research capability and innovation in the European cattle sector

Webinar:

Applications of the ontology AHOL

MC Meunier-Salaün

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n°730924

Applications of AHOL Ontology

Objectives

- Creation of centralized databases with computer interface with ontologies => annotated databases on health events

- Diagnosis tool using knowledge of the ontology

SmartCow

Applications of AHOL Ontology

annotated on farm databases

- annotated databases on health events
- query and exploit data on health events in correspondence with ontologies identifiers via a formal language *(ex SPARQL query)*
- assess the health status of animal or herd from queries on the frequency of diseases, the presence of pathogens, the expression of clinical signs
- allow interoperable datasets using the same reference identifier

SmartCow

Applications of AHOL Ontology
principle of the interface with ontology

Experimental/farm Unit

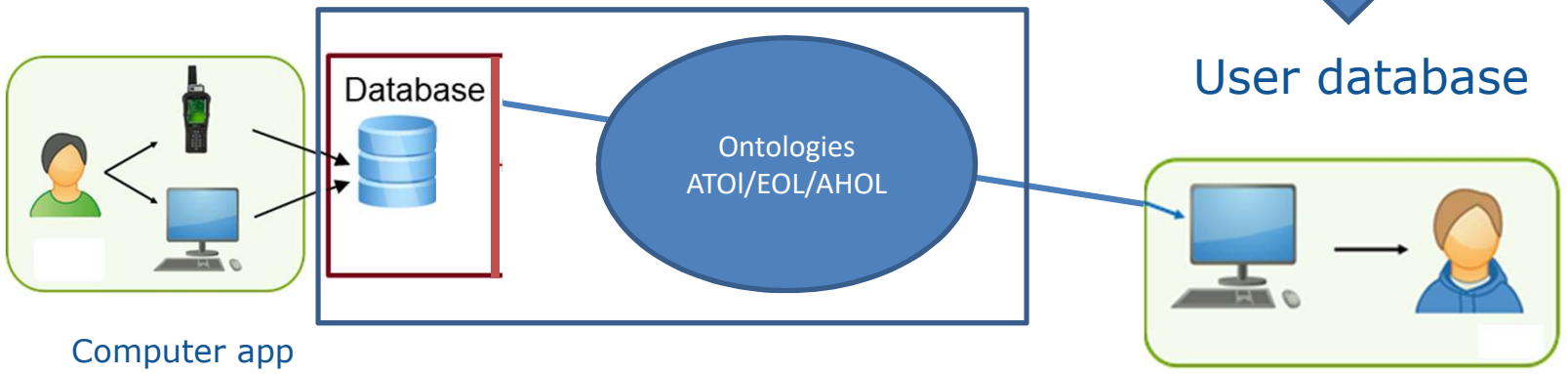
Data collectors at the animal level

On farm database

Users (scientist/ manager)

Data extraction
from queries of the farm database

User database



Computer app

Web app

Applications of AHOL Ontology
principle of the interface with ontology

Experimental/farm Unit

Users (scientist/ manager)

Data collectors at the animal level

Data extraction
from queries of the farm database

On farm database

Items on

Informatic interfaces

User database

Items on

* animal — animal /all (farm level)

* recording time — Time period of recording

* health events :
  - disease
  - treatment
  - pathogen
  - clinical signs

Ontology AHOL

- disease
- pathogen
- clinical signs

SmartCow

Annotated files from data extraction of the user database

Data extraction Query items - animals from the experiment
- recording period : 01/01/2020 to 01/10/2020
- on event clinical signs : lameness

data file with various format

Copy | Excel | CSV | PDF | Print

Animal	Start	End	Treatment	Event family	Event	Id AHOL
805471	01/01/2020	01/10/2020	MELOVEM	members	lameness	AHOL_0003029
060032	01/01/2020	01/10/2020	-	members	lameness	AHOL_0003029
....						

➔ shared datasets for health status of several farms
➔ meta analysis on experimental health dataset

a significant added value for standardized registration, optimal operation, and sharing of health data, and broader phenotypic data (ATOL) and their context (EOL) collected on animals

SmartCow

Applications of AHOL Ontology : diagnostic tool

- Query the ontology via an app (web, smartphone) to confirm health diagnosis

- disease suspected : check the pathogen or clinical symptom related to

- pathogen found : check potential diseases related to

- clinical signs recorded : check potential disease related to

➔ Potential tool to check, to improve knowledge on health

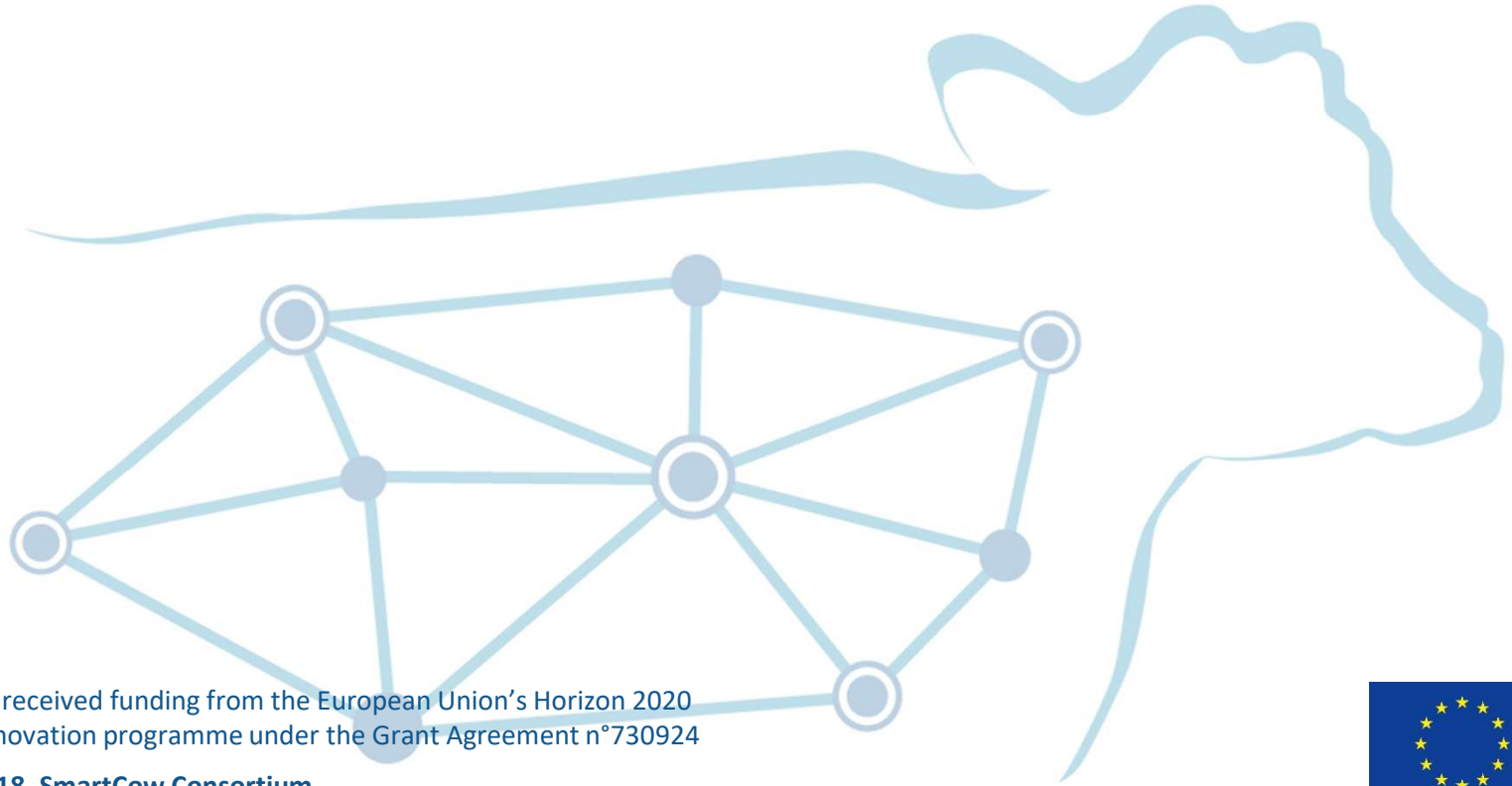
Applications of AHOL Ontology : diagnostic tool

- Assess the health status of animal or herd from queries on the frequency of diseases, the presence of pathogens, the expression of clinical signs

- Allow interoperable datasets using the same reference identifier

SmartCow

Thanks for your attention



This project has received funding from the European Union's Horizon 2020
research and innovation programme under the Grant Agreement n°730924

SmartCow

an integrated infrastructure for increased research capability and innovation in the European cattle sector

Webinar

Using ontologies to improve animal science research

8-9 October 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n°730924

Text-based annotation with ontologies

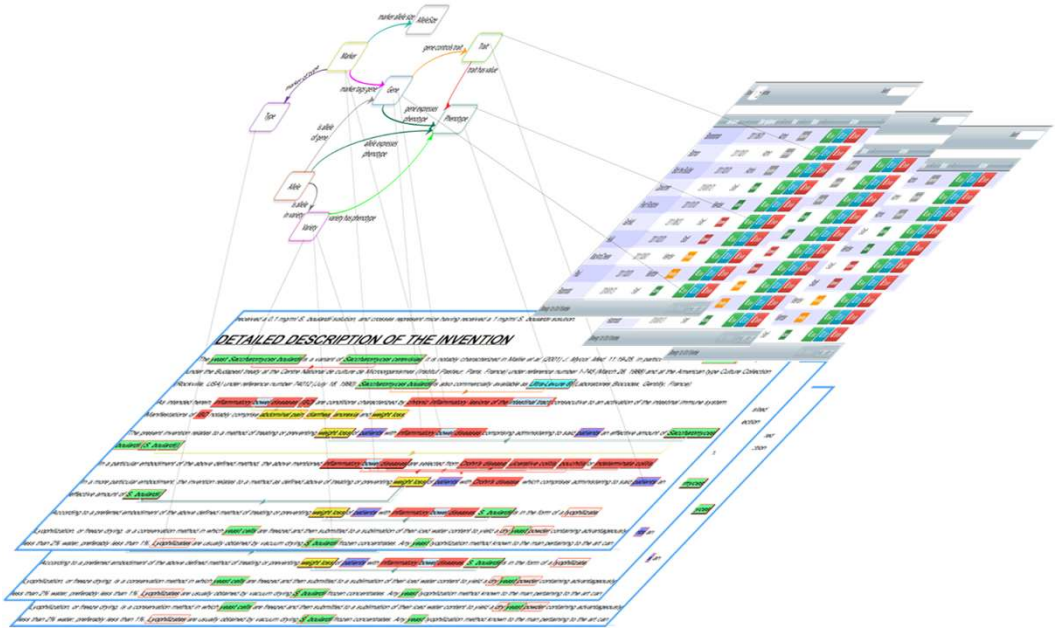
Claire Nédellec
MaIAGE, INRAE
Université Paris-Saclay

Using ontologies to combine data from multiple sources

Objective
Centralize, structure and standardize information
For better access and reuse

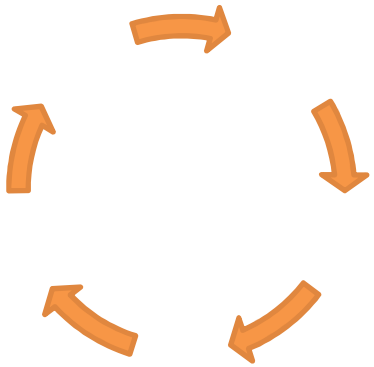
Free text data are many

Text mining
transforms
unstructured data
into structured data



Combining data from multiple sources

Information from bibliography



- Supports research hypotheses
- Supports experimental condition settings
- Confirms experiments results by previous published work
- Meta-analysis

Ontology for data annotation

mapping

ATOL

- animal trait of livestock
  - mammary gland and milk production trait
    - mammary gland production trait
      - milk trait
        - milk composition trait
          - milk organic acid trait
            - milk citric acid concentration

Experimental data

Parameter	Group 1‡	Group 2	Group 3
Number of cows	49	23	18
Milk yield/milking (kg)	$11.16^{a} \pm 0.37$	$10.92^{ab} \pm 0.59$	$9.37^{b} \pm 0$
Fat (%)	$4.77^{a} \pm 0.13$	$4.79^{a} \pm 0.21$	$5.08^{a} \pm 0$
Total protein (%)	$3.55^{a} \pm 0.04$	$3.51^{a} \pm 0.07$	$3.55^{a} \pm 0$
Casein (%)	$2.61^{a} \pm 0.03$	$2.57^{a} \pm 0.05$	$2.56^{a} \pm 0$
Whey protein (%)	$0.94^{a} \pm 0.02$	$0.94^{a} \pm 0.03$	$0.98^{a} \pm 0$
Casein number	$0.74^{a} \pm 0.00$	$0.73^{ab} \pm 0.01$	$0.72^{b} \pm 0$
Lactose (%)	$4.55^{A} \pm 0.03$	$4.51^{ab} \pm 0.04$	$4.39^{c} \pm 0$
Citric acid (%)	$0.16^{a} \pm 0.00$	$0.16^{a} \pm 0.01$	$0.16^{a} \pm 0$
Log SCC (cells/ml)	$4.50(32^{\dagger})^{A} \pm 0.05$	$5.20(160^{\dagger})^{B} \pm 0.07$	$5.89(784^{\dagger})^{C} \pm 0$

Ontology for annotation of textual data

milk organic acid trait
└ milk citric acid concentration

The **citric acid in whole milk was measured** at different intervals

Different parameters of **milk samples** (fat, total protein, casein, lactose, total solids, solids non-fat, **citric acid**, [..]

The effect of subclinical mastitis on the **citric acid concentration of cow's milk** was studied

Milk citrate concentration of affected animals was significantly lower

Factors influencing **milk citric acid concentrations in dairy cows**

Choline, carnitine, **citric acid** and lactose **in milk** were correlated with

Ontology for annotation of textual data

milk organic acid trait
- milk citric acid concentration

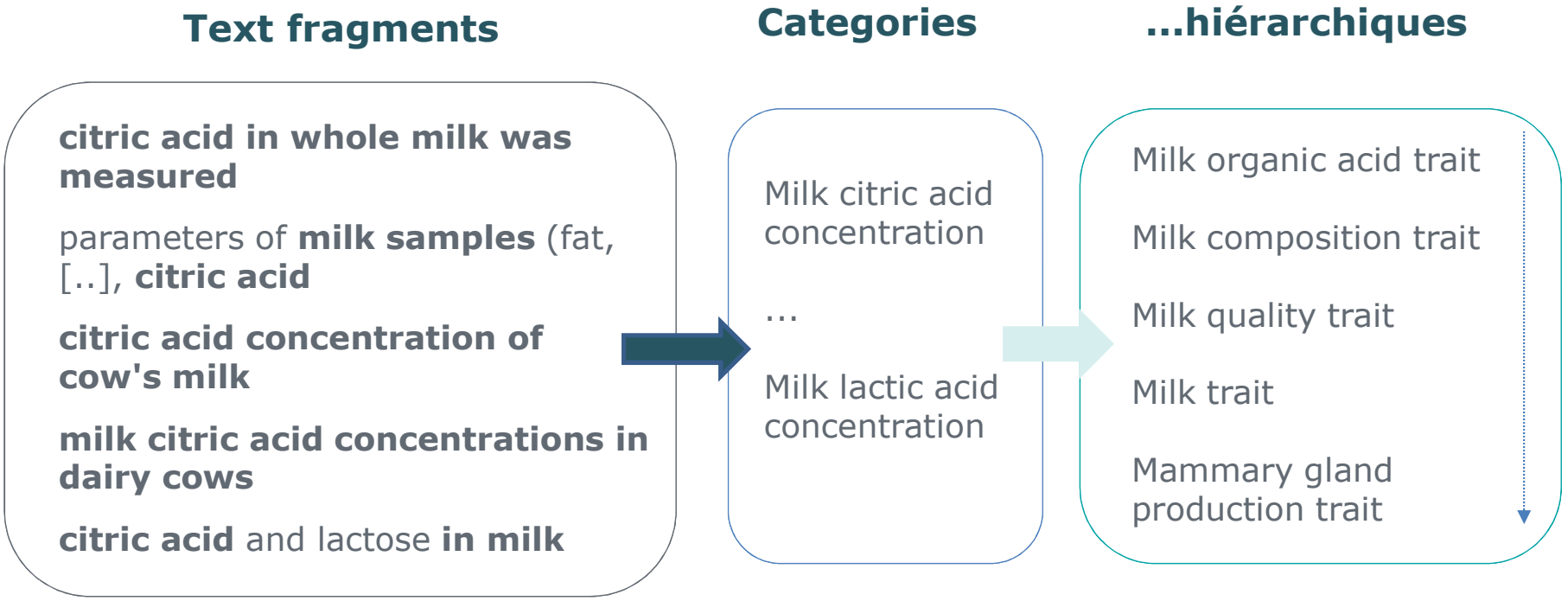
measure of | citric acid | milk | cattle

From concept → Lexicon, syntax → To text

Factors influencing **milk citric acid concentrations in dairy cows**

The **citric acid in whole milk was measured** at different intervals

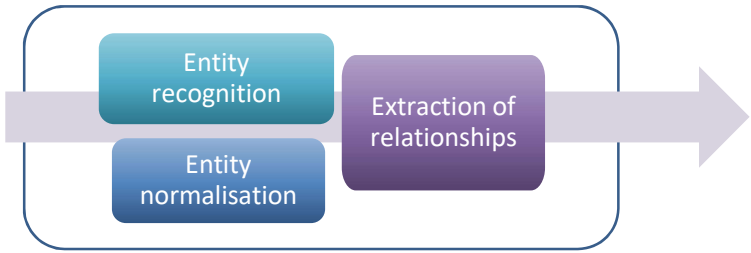
How to find information on milk composition, citric acid and other components

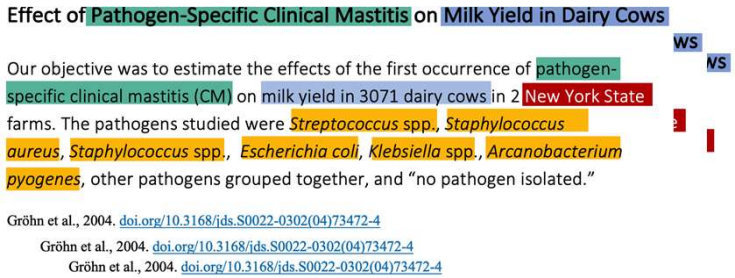


Enable conceptual abstraction = By climbing the hierarchy

Text mining

Automatic extraction of information





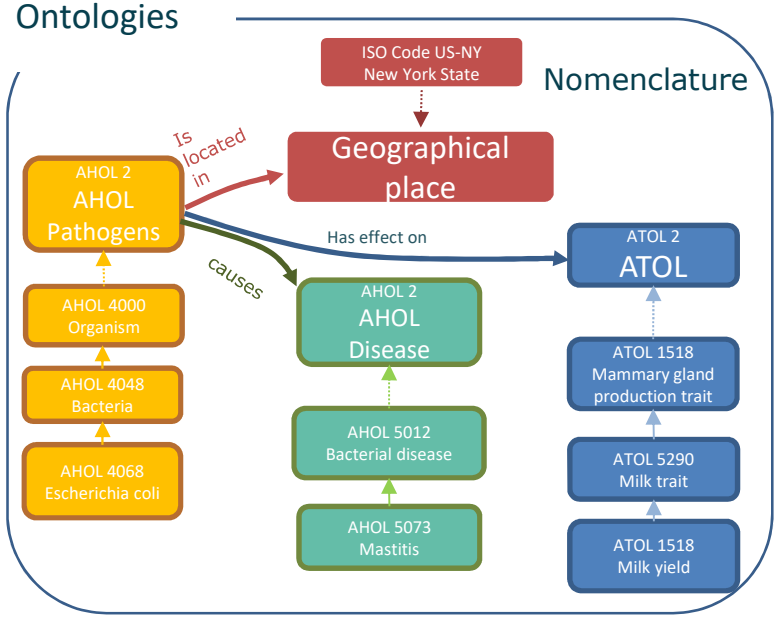
Effect of Pathogen-Specific Clinical Mastitis on Milk Yield in Dairy Cows

Our objective was to estimate the effects of the first occurrence of pathogen-specific clinical mastitis (CM) on milk yield in 3071 dairy cows in 2 New York State farms. The pathogens studied were *Streptococcus* spp., *Staphylococcus aureus*, *Staphylococcus* spp., *Escherichia coli*, *Klebsiella* spp., *Arcanobacterium pyogenes*, other pathogens grouped together, and "no pathogen isolated."

Gröhn et al., 2004. doi.org/10.3168/jds.S0022-0302(04)73472-4

Gröhn et al., 2004. doi.org/10.3168/jds.S0022-0302(04)73472-4
Gröhn et al., 2004. doi.org/10.3168/jds.S0022-0302(04)73472-4

Ontologies



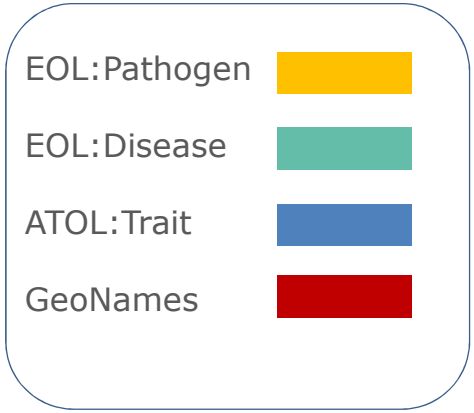
Entity recognition

Entities types and boundaries

Effect of Pathogen-Specific Clinical Mastitis on Milk Yield in Dairy Cows

Our objective was to estimate the effects of the first occurrence of pathogen-specific clinical mastitis (CM) on milk yield in 3071 dairy cows in 2 New York State farms. The pathogens studied were *Streptococcus* spp., *Staphylococcus aureus*, *Staphylococcus* spp., *Escherichia coli*, *Klebsiella* spp., *Arcanobacterium pyogenes*, other pathogens grouped together, and "no pathogen isolated."

Gröhn et al., 2004. doi.org/10.3168/jds.S0022-0302(04)73472-4



Mapping of Ontology concepts to text

Dealing with variation

Typographic variation	meat colour / meat color crude fiber / crude fibre
Abbreviation	Lactobacillus spp. / Lactobacillus subspecies
Acronym	Colostrum somatic cell count / Colostrum SCC
Morphological variation	testicular size / testis size / testes size
Morpho-syntactic variation	weight seen at birth / weight at birth activity for LDH / LDH activity citric acid and lactose in milk / milk citric acid
Semantic variation	adipose tissue / fat tissue

Entity regognition

methods

Dictionary mapping

- For strict nomenclature
- Suffers from homonymy and ambiguity

Rules with criteria on context and entity composition

- Good precision, poor recall

Supervised machine Learning from training examples

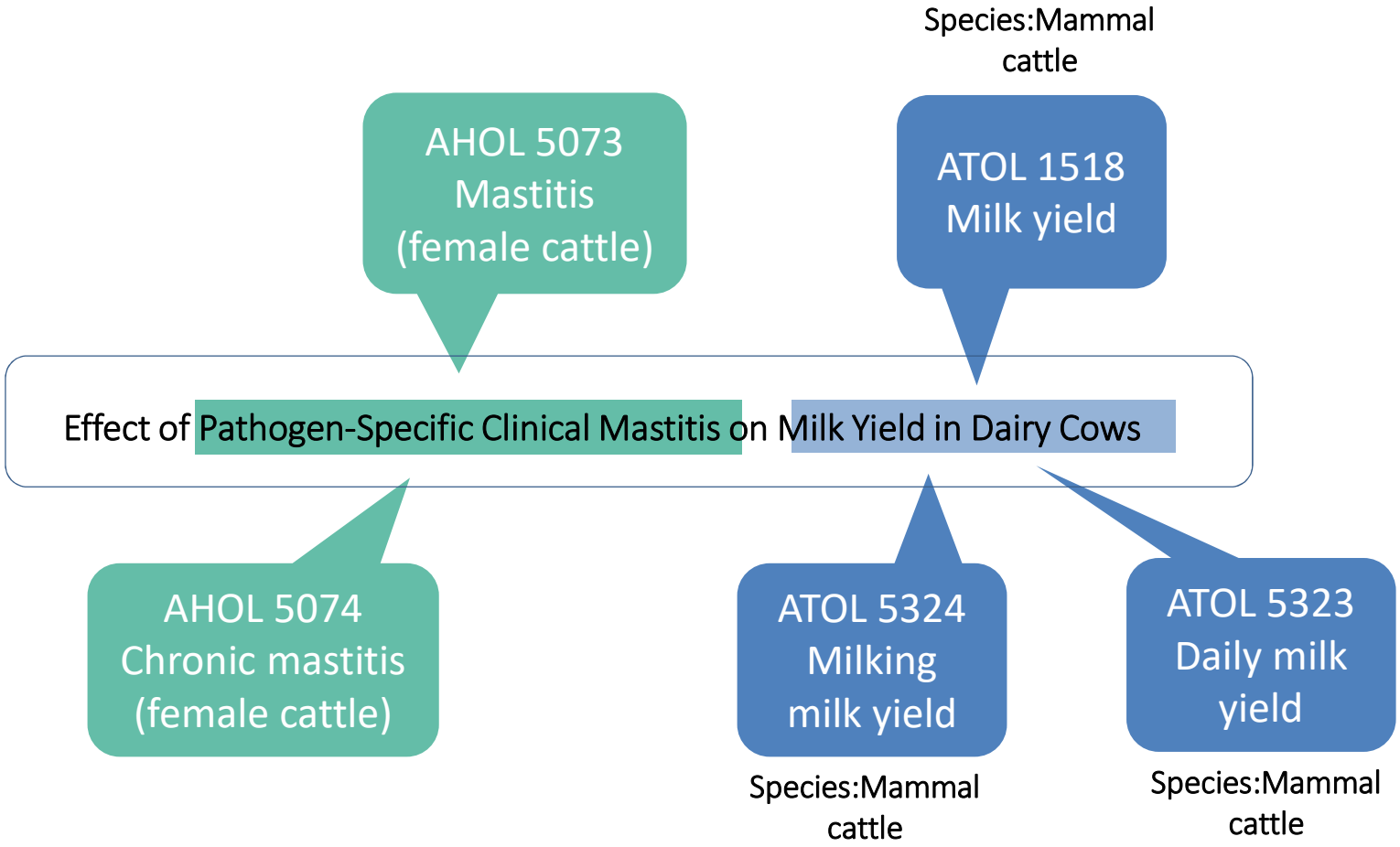
- Probability of a word to belong to an entity of a given type
- Required manual annotation
- To be combined with word embeddings

Term extraction and entity type prediction by machine learning

- Good precision of the boundaries
- To be combined with distant supervision

Entity normalization

Example



Entity normalization

methods

Similarity criteria between the entity and the class

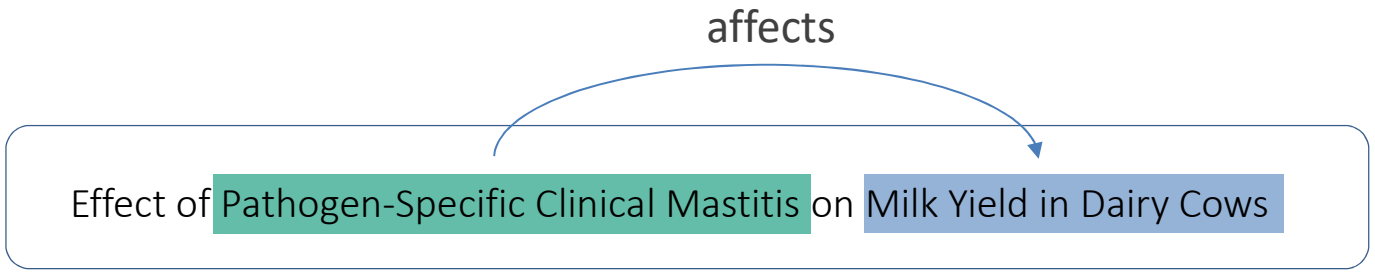
- Common words
- Same « head »
- Vector (word embedding) distance

Supervised machine Learning from examples

- Required manual annotation
- Zero-shot learning problem
- To be combined with word embeddings
- Using the structure of the ontology

Relationships

Examples and methods



Given the entities of the right type, predicts the relation ships

- Rules
- Supervised Machine Learning from positive and negative training examples of the relation

Applications

- **For animal research**
  - Semantic search of papers.
  - Combining data, experimental data and bibliographical results
  - Meta-analysis based on semantic annotation of experimental data tables
- **For decision support**
  - Cartography of main and secondary topics.
  - Who is working on what? Expert profiling
- **For Ontology extension**
  - Add or revise classes through bibliography analysis

Semantic search engine

Classical bibliography search

- Researchers face an overwhelming amount of publications.
- Technical terms are not known by researchers conducting interdisciplinary studies.
- Word-based search misses relevant papers.
- Terminology variation hinders efficient searching.

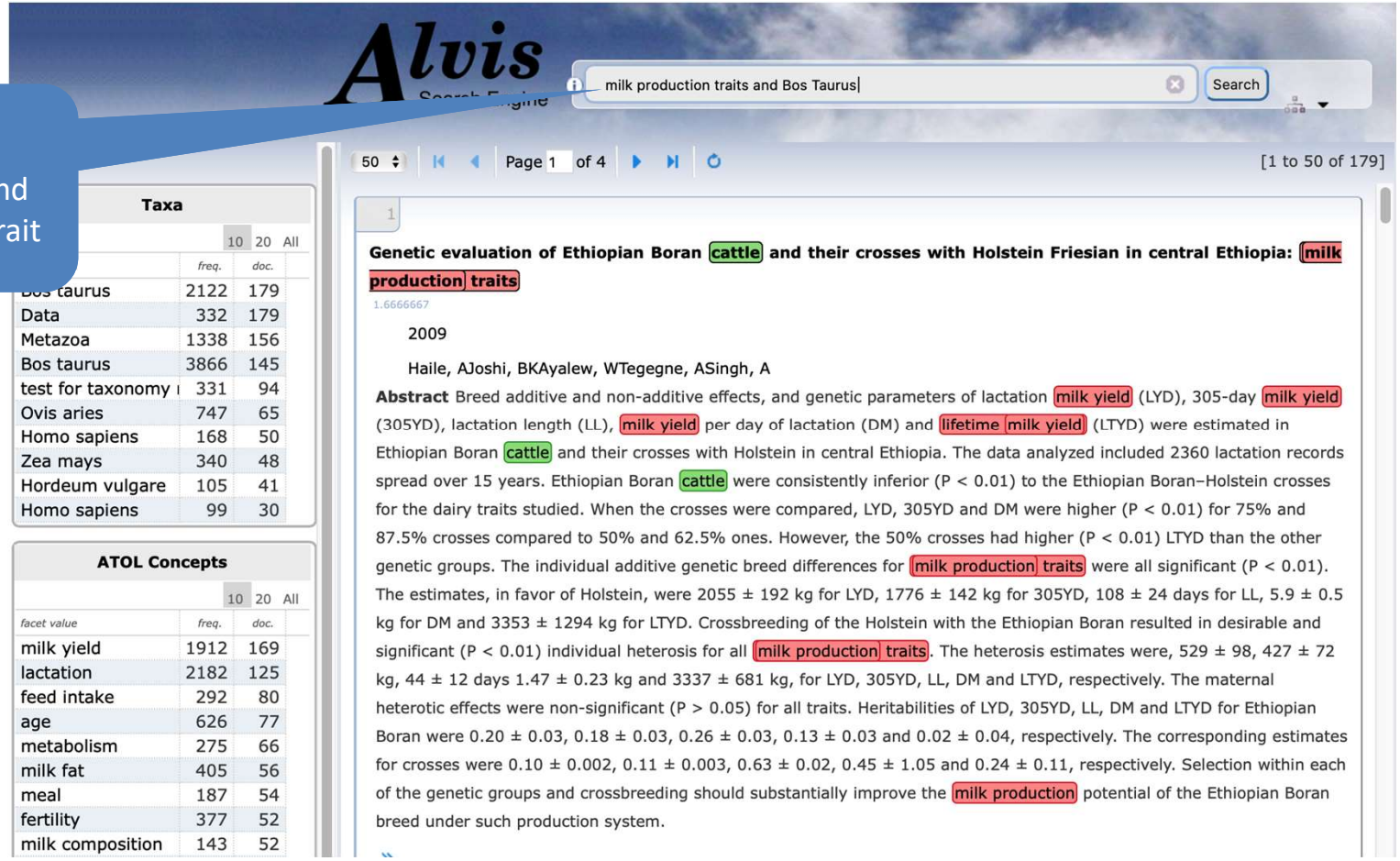
Semantic search principle. Search the full-text of documents for the *ontology concepts of the user query*. Not only user query words.

Benefits.

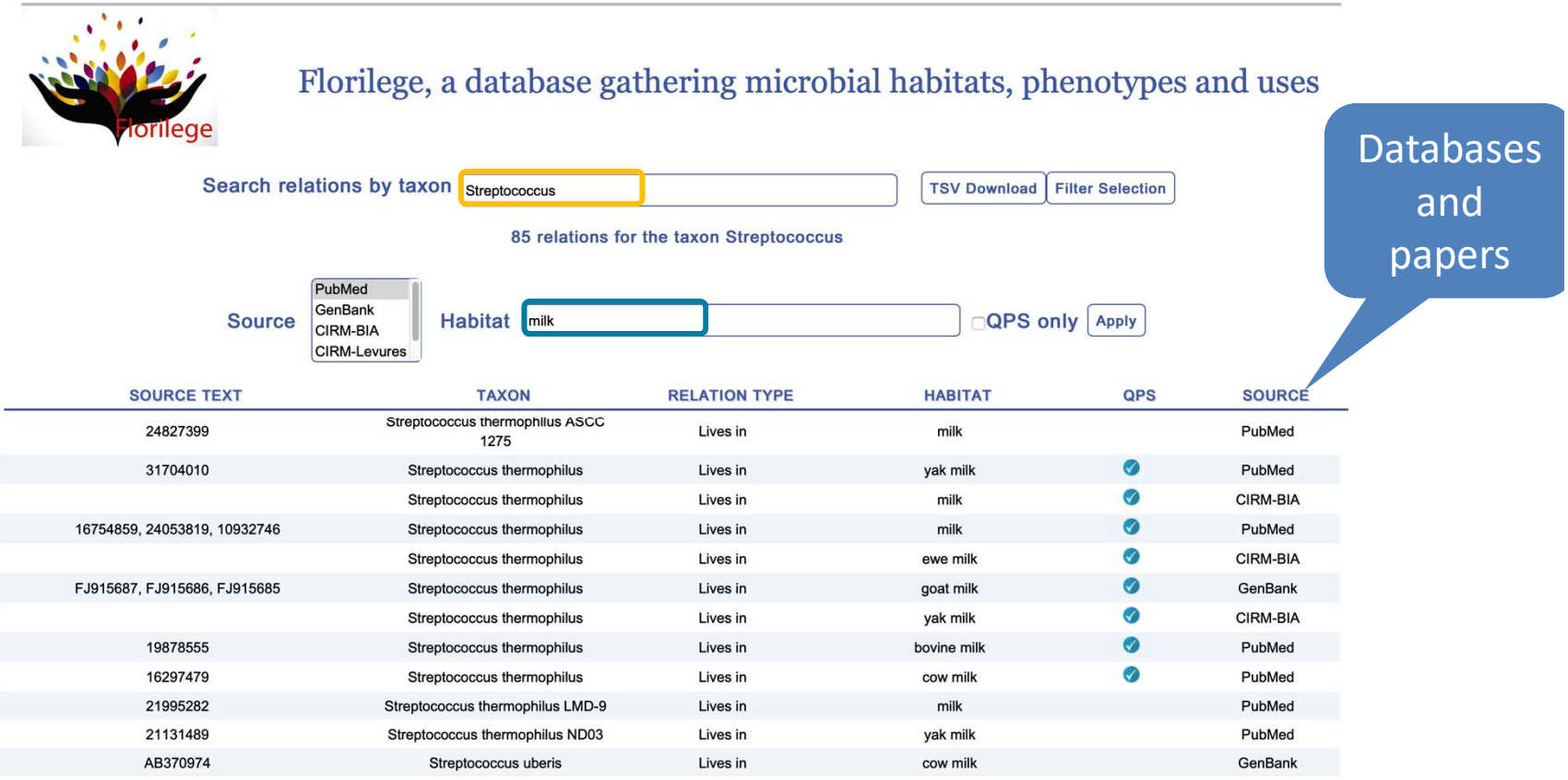
- Efficient search without the knowledge of the relevant keywords,
- Large collections can be automatically indexed and kept up to date.
- An alternative to time-consuming manual indexing of the papers

Semantic search engine - AlvisIR

ATOL 5
Mamary gland and milk production trait



Combining experimental information and bibliography



SOURCE TEXT	TAXON	RELATION TYPE	HABITAT	QPS	SOURCE
24827399	Streptococcus thermophilus ASCC 1275	Lives in	milk		PubMed
31704010	Streptococcus thermophilus	Lives in	yak milk	✓	PubMed
	Streptococcus thermophilus	Lives in	milk	✓	CIRM-BIA
16754859, 24053819, 10932746	Streptococcus thermophilus	Lives in	milk	✓	PubMed
	Streptococcus thermophilus	Lives in	ewe milk	✓	CIRM-BIA
FJ915687, FJ915686, FJ915685	Streptococcus thermophilus	Lives in	goat milk	✓	GenBank
	Streptococcus thermophilus	Lives in	yak milk	✓	CIRM-BIA
19878555	Streptococcus thermophilus	Lives in	bovine milk	✓	PubMed
16297479	Streptococcus thermophilus	Lives in	cow milk	✓	PubMed
21995282	Streptococcus thermophilus LMD-9	Lives in	milk		PubMed
21131489	Streptococcus thermophilus ND03	Lives in	yak milk		PubMed
AB370974	Streptococcus uberis	Lives in	cow milk		GenBank

Combining experimental information and bibliography

Genbank database

Streptococcus uberis gene for 16S ribosomal RNA, partial sequence, str
362B-1

GenBank: AB370974.1

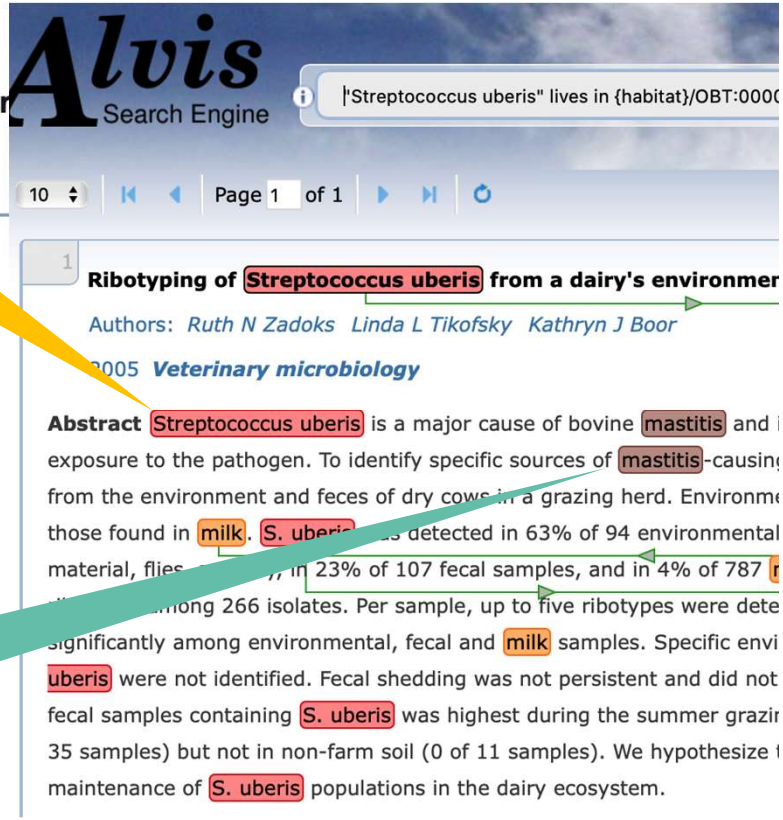
Go to:

```
LOCUS       AB370974               1481 bp    DNA
DEFINITION  Streptococcus uberis gene for 16S ribosoma
            strain: 1-362B-1.
ACCESSION   AB370974
VERSION     AB370974.1
FEATURES             Location/Qualifiers
     source          1..1481
                     /organism="Streptococcus uberis"
                     /mol_type="genomic DNA"
                     /strain="1-362B-1"
                     /isolation_source="mastitis milk of cow"
                     /db_xref="taxon:1349"
                     /country="Japan: Hokkaido,Ebe
                     /collection_date="2006"
     rRNA            <1..>1481
                     /product="16S ribosomal RNA"
```

AHOL 4036
Streptococcus uberis

AHOL 5073
Mastitis
(female cattle)

Pubmed bibliography



Combining experimental information and bibliography

Extend the Information System to deal with semantic queries

Query interpretation
Result display

Query: all data about mammary gland production trait

Interpretation

ATOL

- mammary gland production trait
  - milk trait
    - milk composition trait
      - milk organic acid trait
        - milk citric acid concentration

ATOL 1765

Mapping

ATOL 1765

Animal	experiment Id	Sample Id	Citric acid %
805471	87GH9	UOI98	0,16
060032	87GH9	UOI99	0,11
....			

Experimental data

Combining experimental information and bibliography

Requirements

Extend the Information System to deal with semantic queries

- Query interpretation
- Result display

Alignment of the attributes of the database schema with the ontology classes

- Select the relevant attributes of the database (e.g. trait, disease)
- Manually align the attributes with the ontology classes
- Re index the data

Combining experimental information and bibliography

Requirements

Extend the Information System to deal with semantic queries

- Query interpretation
- Result display

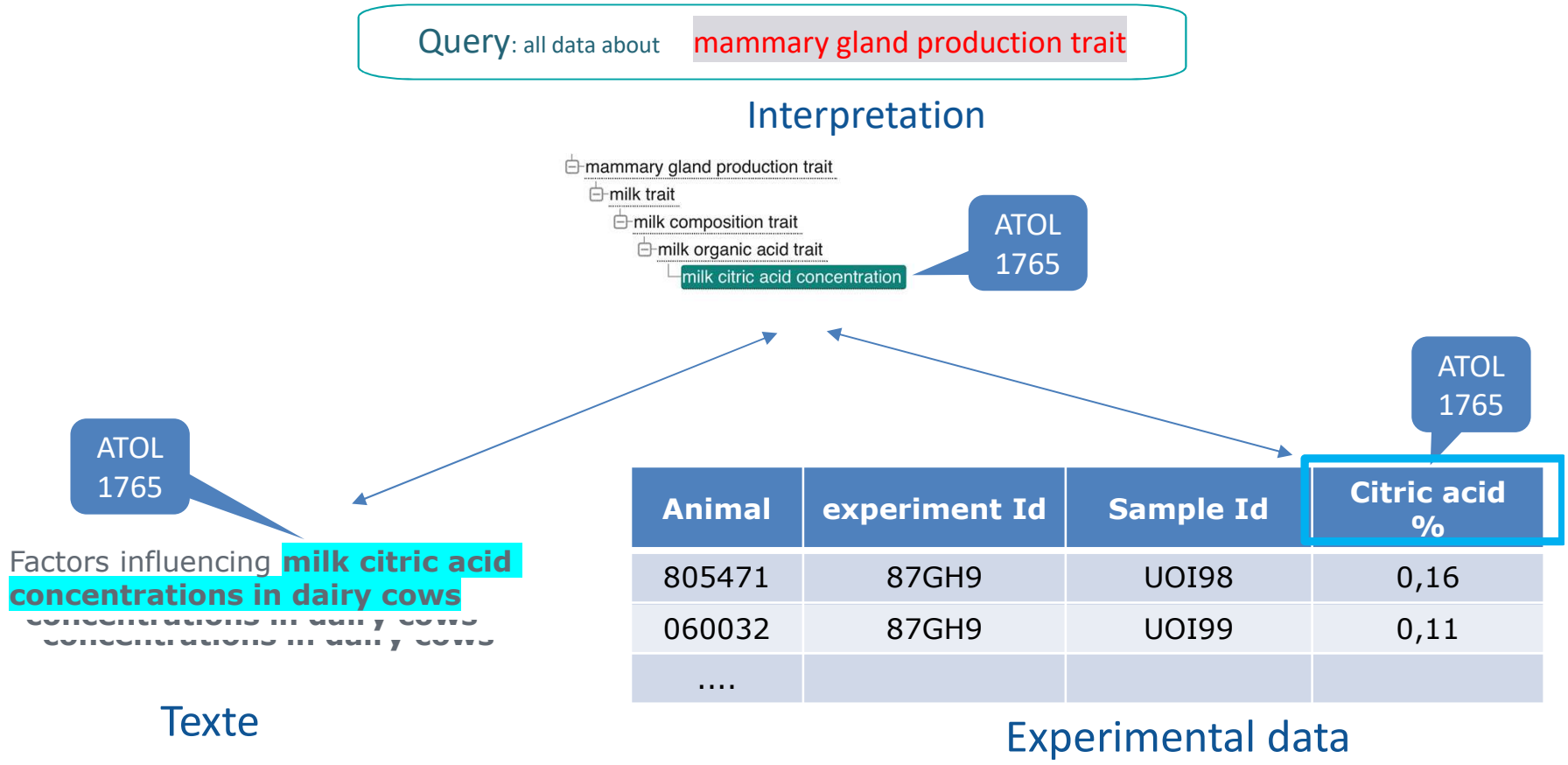
Alignment of the attributes of the database schema with the ontology classes

- Select the relevant attributes of the database (e.g. trait, disease)
- Manually align the attributes with the ontology classes
- Re index the data

Semantic analysis of the text

- Identify relevant nomenclatures and ontologies
- Train the methods to recognize entites of the relevant class types
- Index the textual data with the trained methods

Combining experimental information and bibliography



Conclusion

Text mining tranforms unstructured text into structured data

- Extracts information
- Relates it to domain knowledge by normalisation and extraction of relationships

Integration into a unified information system

- Mapping of the data schemata based on ontology
- May require more complex alignment rules (*e.g.* different measure units)

Next: align data, not only concepts

- Extraction of numbers from the text (e.g. tables)
- Transform numerical values into qualitative values (e.g 0,15 = low)