A unified Knowledge Graph based Approach to Exploring Scientific Text Corpora

Catherine Faron

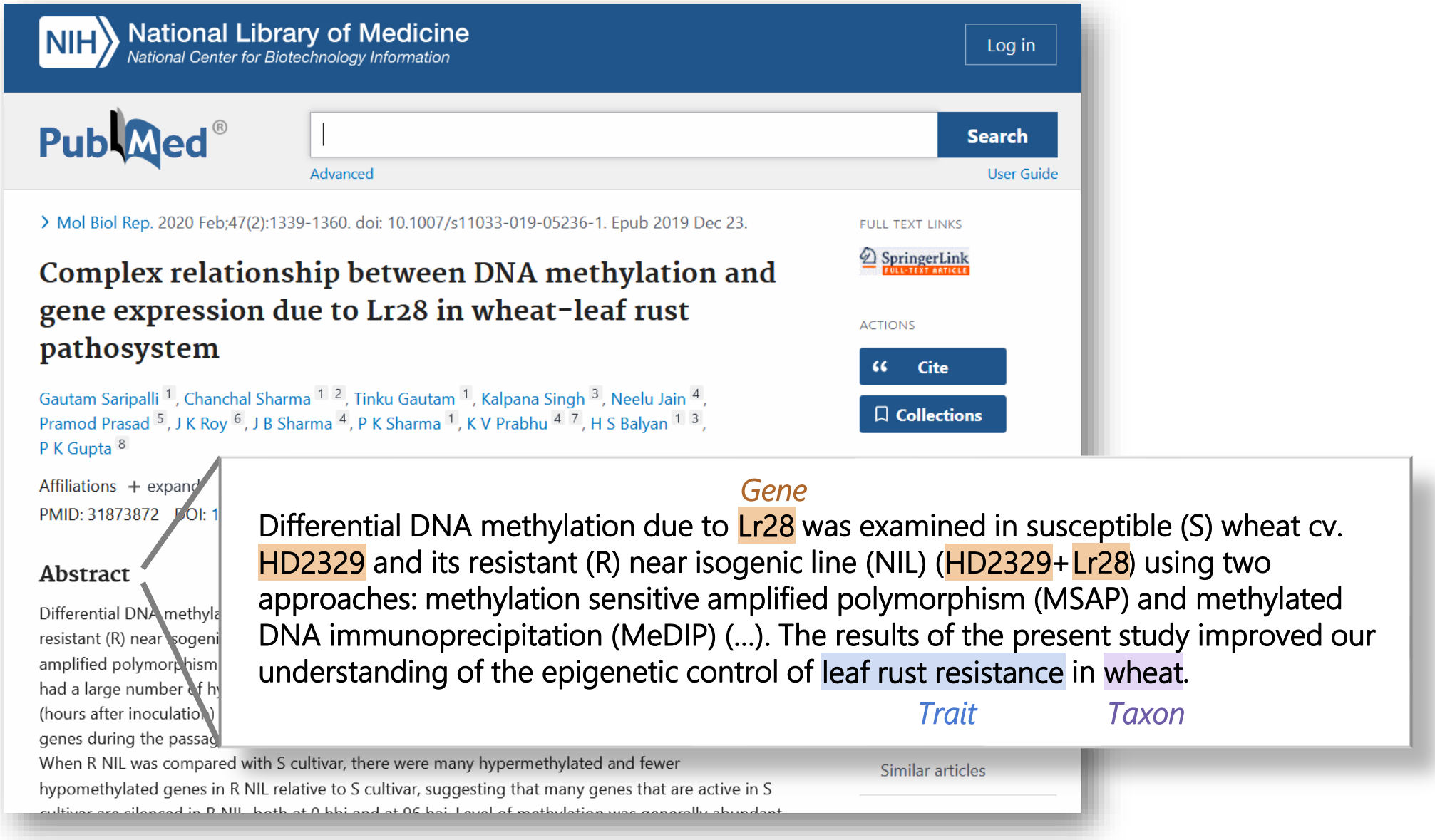
UNIVERSITÉ CÔTE D'AZUR cnrs Inria

séminaire HisINum, Académie 5, 25/09/2024

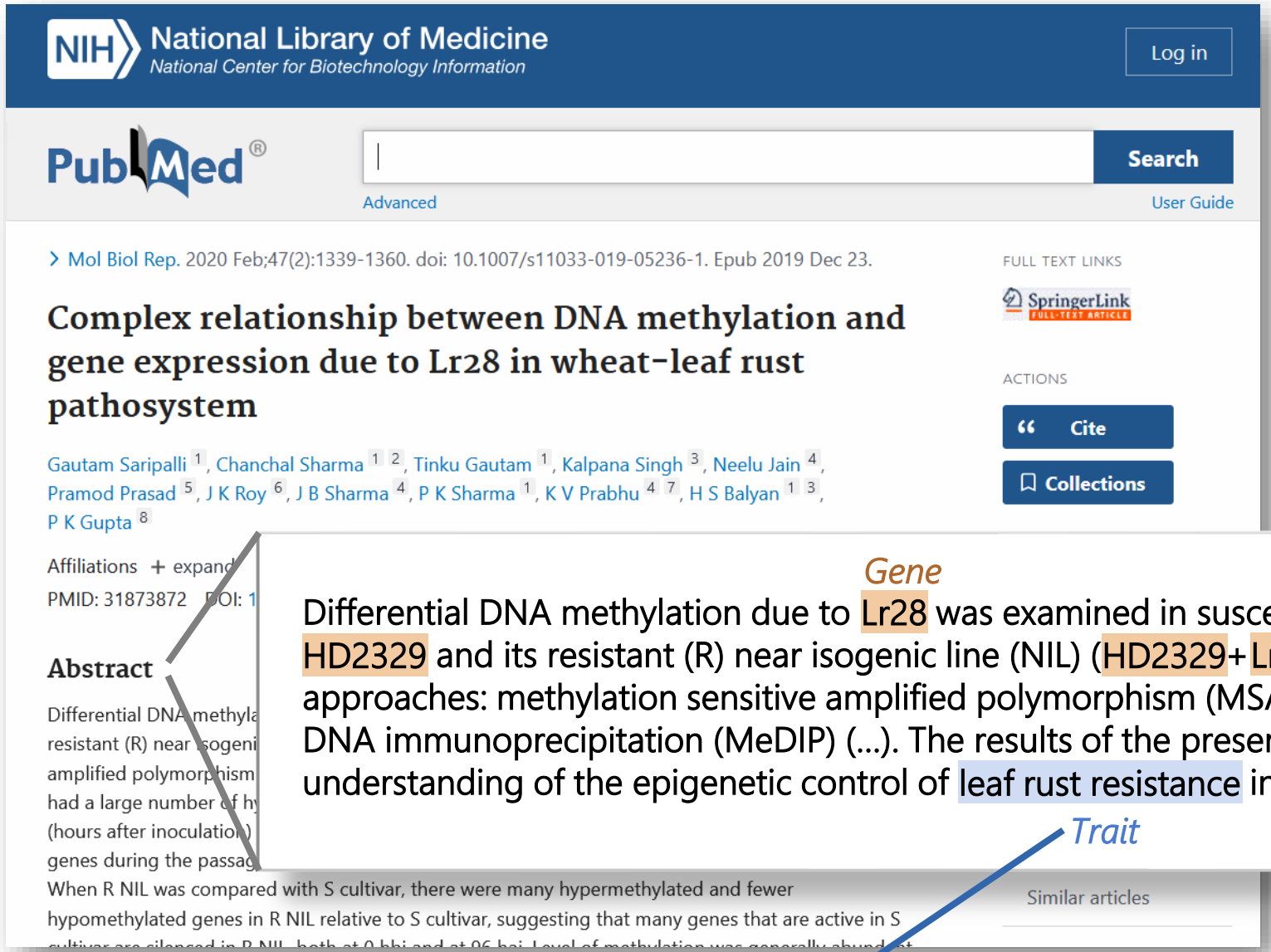
Wheat Genomics Scientific Literature Knowledge Graph

Data to Knowledge in Agronomy and Biodiversity
2019-2024
ANR grant ANR-18-CE23-0017
http://www.d2kab.org

Researchers need help to make sense of scientific literature


National Library of Medicine
National Center for Biotechnology Information
Log in
PubMed
Search
Advanced
User Guide
Mol Biol Rep. 2020 Feb;47(2):1339-1360. doi: 10.1007/s11033-019-05236-1. Epub 2019 Dec 23.
Complex relationship between DNA methylation and gene expression due to Lr28 in wheat–leaf rust pathosystem
Gautam Saripalli [1], Chanchal Sharma [1 2], Tinku Gautam [1], Kalpana Singh [3], Neelu Jain [4], Pramod Prasad [5], J K Roy [6], J B Sharma [4], P K Sharma [1], K V Prabhu [4 7], H S Balyan [1 3], P K Gupta [8]
Affiliations + expand
PMID: 31873872 DOI: 1
Abstract
Differential DNA methyla
resistant (R) near isogeni
amplified polymorphism
had a large number of hy
(hours after inoculation)
genes during the passag
When R NIL was compared with S cultivar, there were many hypermethylated and fewer hypomethylated genes in R NIL relative to S cultivar, suggesting that many genes that are active in S
FULL TEXT LINKS
SpringerLink
FULL-TEXT ARTICLE
ACTIONS
Cite
Collections
Similar articles
Gene
Differential DNA methylation due to Lr28 was examined in susceptible (S) wheat cv. HD2329 and its resistant (R) near isogenic line (NIL) (HD2329+Lr28) using two approaches: methylation sensitive amplified polymorphism (MSAP) and methylated DNA immunoprecipitation (MeDIP) (...). The results of the present study improved our understanding of the epigenetic control of leaf rust resistance in wheat.
Trait
Taxon

Extracting & Linking Named Entities



Gene

Differential DNA methylation due to Lr28 was examined in susceptible (S) wheat cv. HD2329 and its resistant (R) near isogenic line (NIL) (HD2329+Lr28) using two approaches: methylation sensitive amplified polymorphism (MSAP) and methylated DNA immunoprecipitation (MeDIP) (...). The results of the present study improved our understanding of the epigenetic control of leaf rust resistance in wheat.

Trait

Taxon

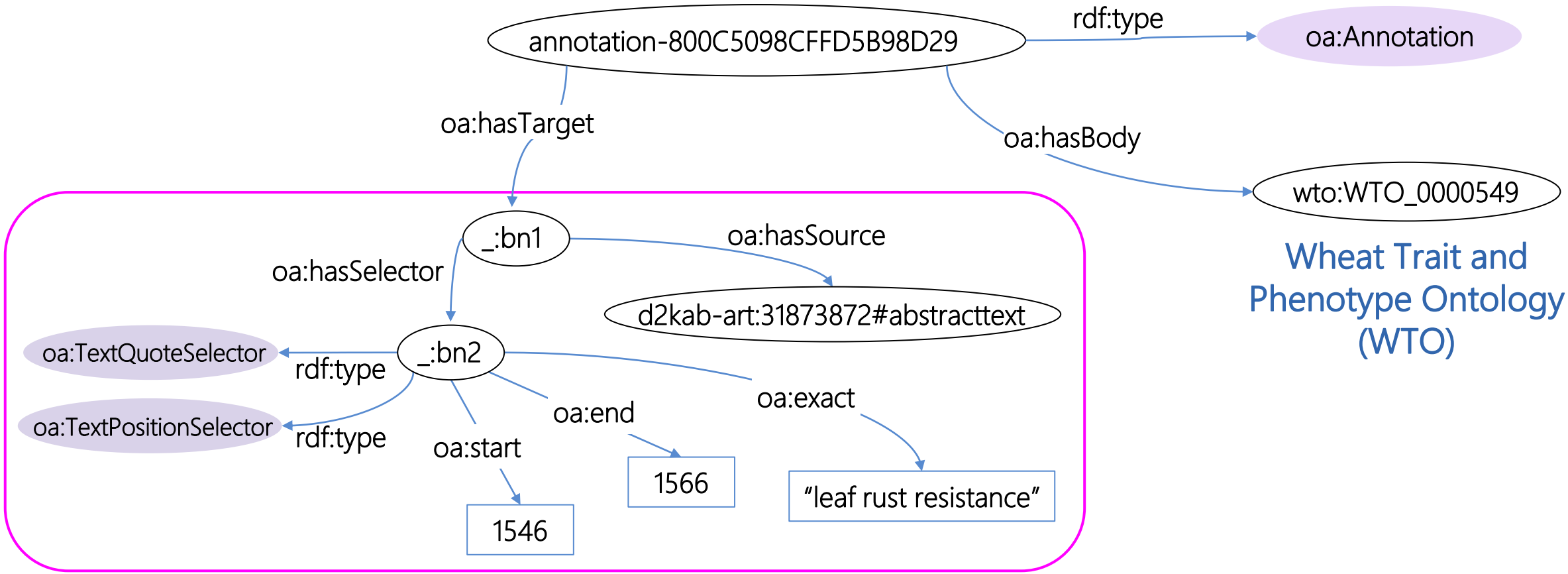
Wheat Trait and Phenotype Ontology (WTO)

NCBI Taxonomy

Tool: AlvisNLP/ML - https://maiage.inrae.fr/fr/node/626

Modeling mentions of terms as annotations

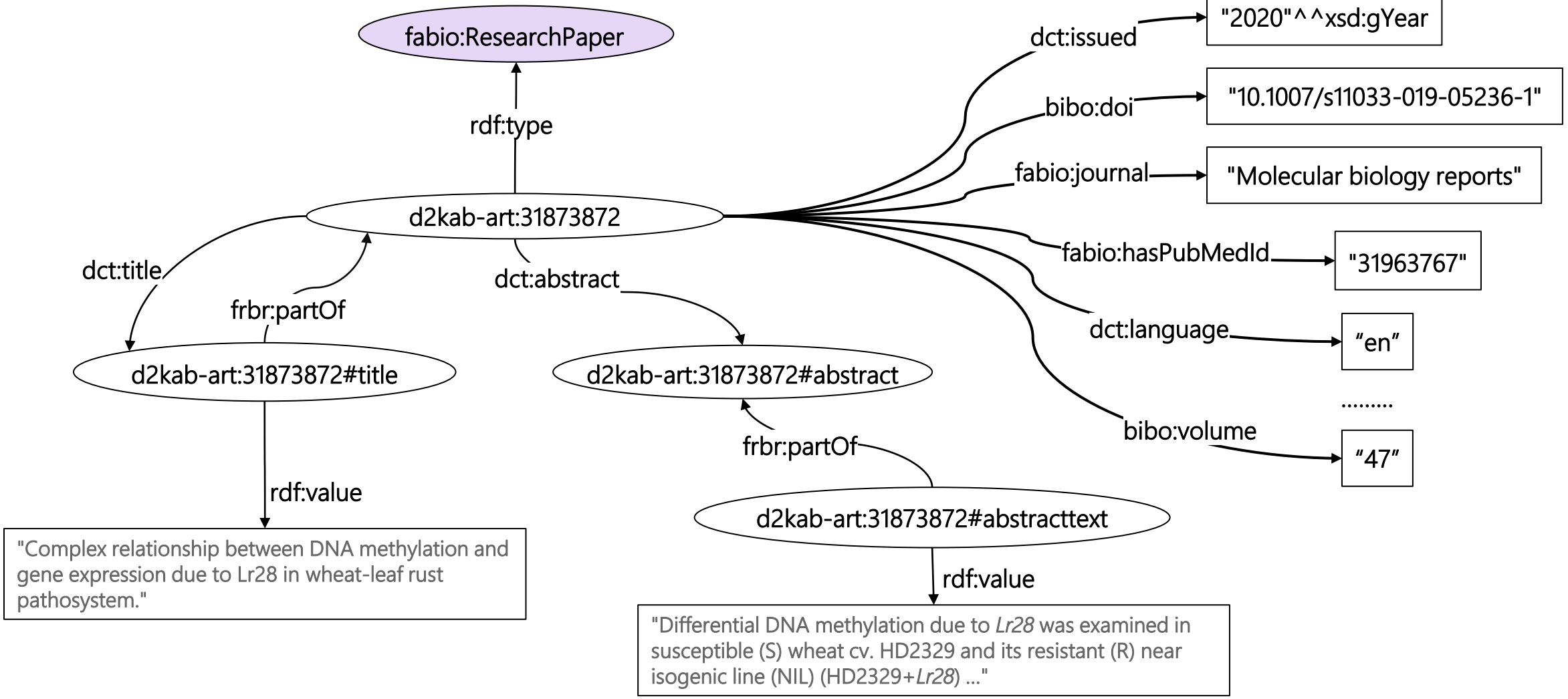
Web Annotations Ontology



Wheat Trait and Phenotype Ontology (WTO)

Tool: Morph-xR2RML - https://github.com/frmichel/morph-xr2rml/

Modeling articles metadata



Tool: SPARQL micro-service - https://github.com/frmichel/sparql-micro-service/

WheatGenomicsSLKG statistics

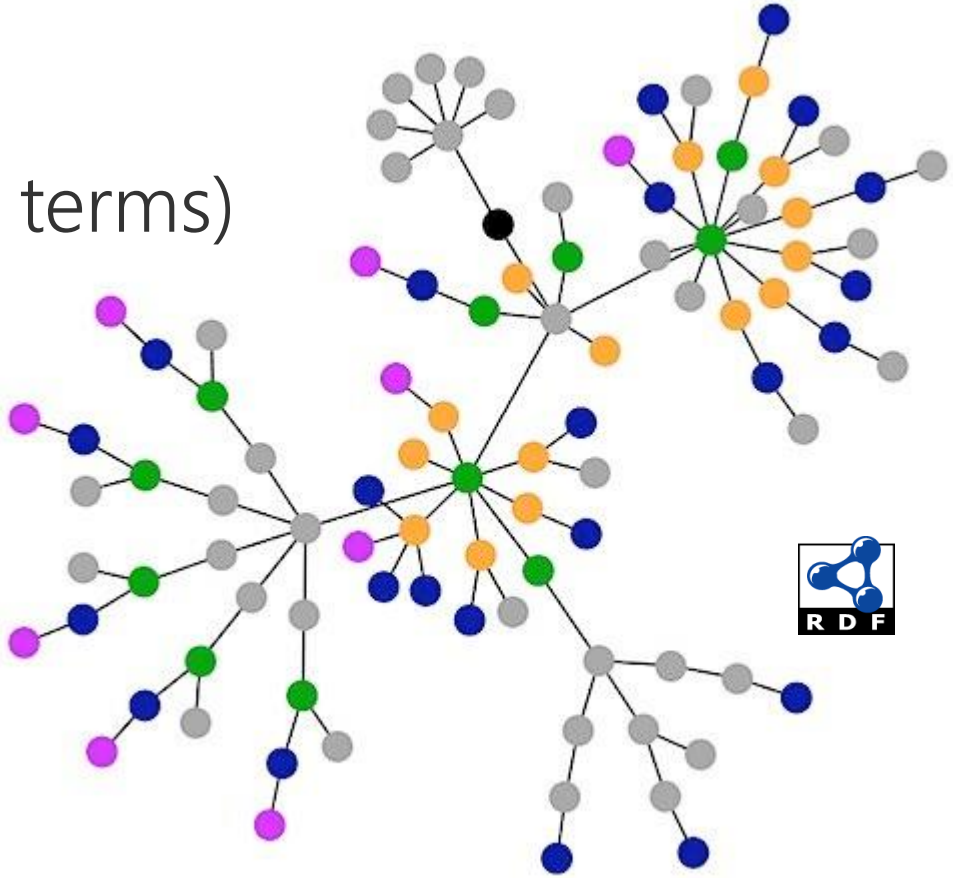
1,191,867 triples

8,496 scientific articles

88,880 annotations (mentions of terms)

4,319 unique entities:

- 2463 taxa
- 98 traits
- 77 varieties
- 1160 genes
- 521 genetic markers



Wheat Genomics SLKG

Search

Search interface for the Wheat Genomics Scientific Literature Knowledge Graph.

Search articles by named entities

Enter text and select a suggestion

Search

rust resistance × Lr34 ×

Results for only the selected named entities

Development of multiplex PCR to detect slow rust resistance genes Lr34 and Lr46 in w

QTL for spot blotch resistance in bread wheat line Saar co-locate to the biotrophic dis
Theoretical and applied genetics. Theoretische und angewandte Genetik.

The Lr34 adult plant rust resistance gene provides seedling resistance in durum whea

Postulation of rust resistance genes in Nordic spring wheat genotypes and identificatio
Bariana H, Lillemo M [...]. 2016. *Journal of applied genetics.*

Gene expression patterns in near isogenic lines for wheat rust resistance gene lr34/yr

Results for the selected named entities or their sub-entities

resistance to Stem rust Lr34
Leaf rust resistance gene lr34 associated with nonsuppression of stem rust resistance

resistance to Stripe rust Lr34
New slow-rusting leaf rust and stripe rust resistance genes Lr67 and Yr46 in wheat ar
[...]. 2011. *TAG. Theoretical and applied genetics. Theoretische und angewandte Gen*

Wheat Genomics SLKG

Search

Search interface for the Wheat Genomics Scientific Literature Knowledge Graph.

Development of multiplex PCR to detect slow rust resistance genes Lr34 and Lr46 in wheat.

Kwiatek M, Nawracała J, Skowrońska R, Tomkowiak A. 2019. Development of multiplex PCR to detect slow rust resistance genes Lr34 and Lr46 in wheat.. *Journal of applied genetics.*

Language: eng
Read the article

Abstract

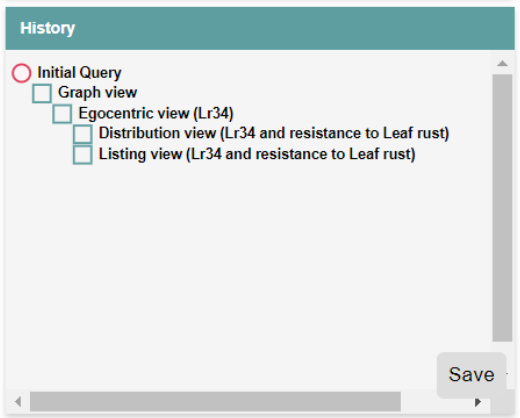
Leaf rust caused by Puccinia triticina belongs to one of the most dangerous fungal diseases of wheat (Triticum aestivum L.) and is the cause of large yield losses every year. Here we report a multiplex polymerase chain reaction (PCR) assay, which was developed for detection of two important wheat slow rust resistance genes Lr34 and Lr46, using two molecular markers: csLV34 and Xwmc44, respectively. The presence of genes was analyzed in one winter wheat variety TX89D6435 and five spring wheat varieties: Pavon F76, Parula 'S', Rayon 89, Kern, Mochis 88. Both Lr34 and Lr46 genes were identified in variety TX89D6435, gene Lr34 was also identified in Parula 'S' and Kern varieties, and gene L46 occurs in Pavon F76 and Mochis 88 variety. None of the resistance genes tested was detected in the Rayon 89 variety. The use of the multiplex PCR method allowed to shorten the analysis time, reduce costs of analyses, and reduce the workload.

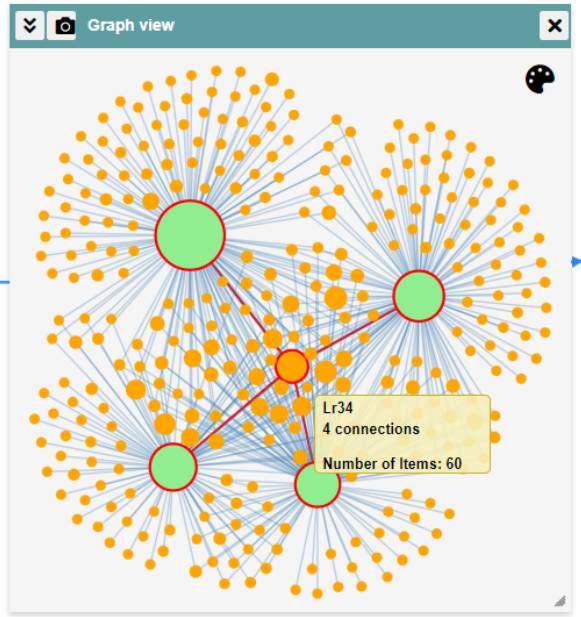
Hide named entities

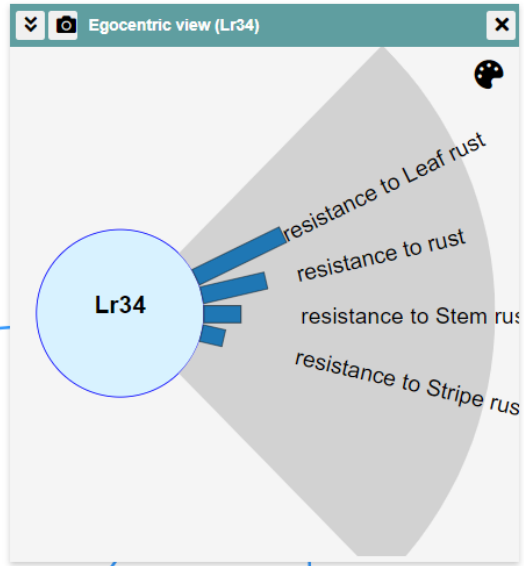
Linked Data Explorer

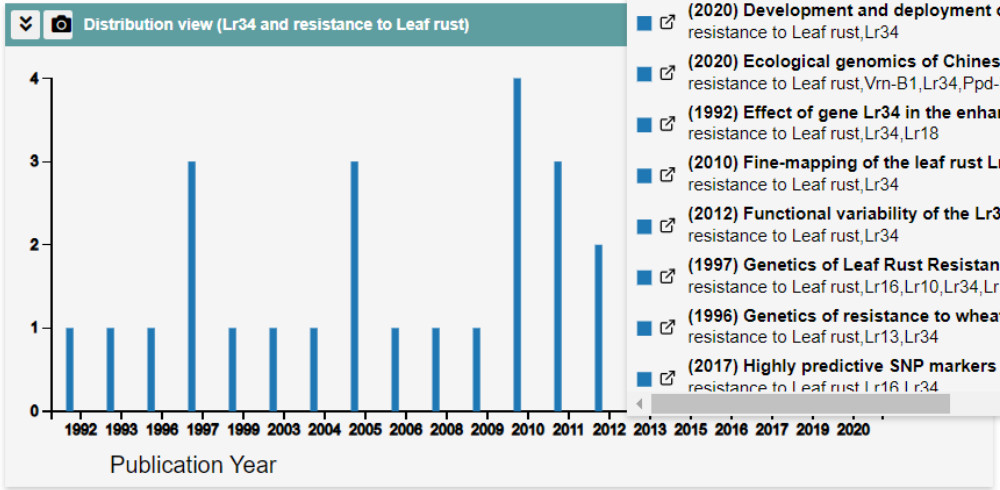
Query name:

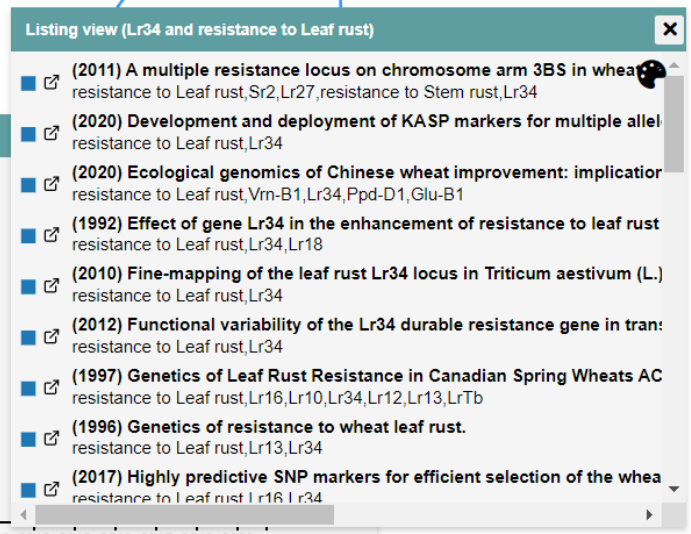
Wimmics team Inria, I3S, UCA, CNRS.











WheatGenomicsSLKG

Dataset (RDF dump)	10.5281/zenodo.10410742
SPARQL endpoint	http://d2kab.i3s.unice.fr/sparql
Documentation	https://github.com/Wimmics/WheatGenomicsSLKG
URIs namespace	http://ns.inria.fr/d2kab/
Web interfaces	http://d2kab.i3s.unice.fr/wheatgenomics/search https://dataviz.i3s.unice.fr/mgexplorer/dashboard with SPARQL endpoint http://d2kab.i3s.unice.fr/sparql
Publications	Yacoubi Ayadi N., Faron C., Michel F., Bossy R., Barbe A. **Construction d'un graphe de connaissances à partir des annotations d'articles scientifiques et de leur contenu en sciences de la vie.** *IC 2022 - PFIA 2022*. hal-03889968 Yacoubi Ayadi N., Bernard S., Bossy R., Courtin M., Happi Happi B.G., Larmande P., Michel F., Nédellec C., Roussey C., Faron C. **A unified approach to publish semantic annotations of agricultural documents as knowledge graphs.** Smart Agricultural Technology , vol 8, 2024

Semantic Indexing of a Scientific Archive in Agriculture

ISSA 2

Projet Lauréat

CollEx-Persée

https://issa.cirad.fr/

Slides from Franck Michel

Researchers need help to explore the Cirad open publication archive

Agritrop, the Cirad open archive

+120 000 bibliographical references including +42 000 in open access; +33 000 articles including +16 in open accesss

A manual indexation by specialized documentalists until the end of 2023

A reference thesaurus for indexing in Agritrop : Agrovoc (thematic and geographic descriptors)

Food and Agriculture Organization of the United Nations

Host plants associated with *Diatraea tabernella* Dyar (Lepidoptera: Crambidae) in sugarcane in Panama

Atencio Valdespino Randy, **Goebel François-Régis**, **Murillo Vielka**. 2018. Host plants associated with *Diatraea tabernella* Dyar (Lepidoptera: Crambidae) in sugarcane in Panama. *International Sugar Journal*, **120** (1438) : 786-791.
https://internationalsugarjournal.com/paper/host-plants-associated-with-diatraea-tabernella-dyar-lepidoptera-crambidae-in-sugarcane-in-panama/

Article de revue ; Article de recherche

Version publiée - Anglais
Accès réservé aux agents Cirad
Utilisation soumise à autorisation de l'auteur ou du Cirad.
ISJ_Oct18_Goebel_Atencio.pdf
Télécharger (273kB) | Demander une copie

Résumé : Within the genus Diatraea, studies of alternate host plants were mainly conducted on Diatraea saccharalis (Fab.). Such information doesn't exist for Diatraea tabernella Dyar. Therefore, the objective of this study was to determine the alternative host plants of D. tabernella in sugarcane plantations in Panama. From January 2016 to February 2017, a general inventory of alternate host plants was conducted in the sugarcane field and then plants were sampled among the most frequent 9 species found (Poaceae (8) and Cyperaceae (1)) in four areas the sugarcane field (near water source (Z1), within field (Z2), near mangrove (Z3) and field edge (Z4)). The species with the highest % borer infestation during the sugarcane growth period were Sorghum halepense (L.) Pers. (7.2%), Echinochloa colonum (L.) Link (6.4%) and Eleusine indica (L.) Gaertn. (5.8%). During the sugarcane harvest the species with the highest percentage of infested stems were Eleusine indica (L.) Gaertn. (21.6%), Dactyloctenium aegyptium (L.) Beauv. (17.1%) and Cenchrus echinatus (L.) (9.2%). On these host plants, species control measures may be applied to reduce stemborer populations.

Mots-clés Agrovoc : Canne à sucre, Ravageur des plantes, Lutte antiravageur, Désherbage, Croissance

Mots-clés géographiques Agrovoc : Panama

Mots-clés complémentaires : Diatraea tabernella

Mots-clés libres : HOST PLANTS, STALK BORERS, SUGARCANE, Integrated Pest Management, Weed management

Classification Agris : H10 - Ravageurs des plantes
H60 - Mauvaises herbes et désherbage
F62 - Physiologie végétale : croissance et développement

Auteurs et affiliations

- Atencio Valdespino Randy, CIRAD-PERSYST-UPR AIDA (FRA) - *auteur correspondant*
- Goebel François-Régis, CIRAD-PERSYST-UPR AIDA (FRA)
- Murillo Vielka, Universidad de Panama (PAN)

Source : Cirad-Agritrop (https://agritrop.cirad.fr/591124/)

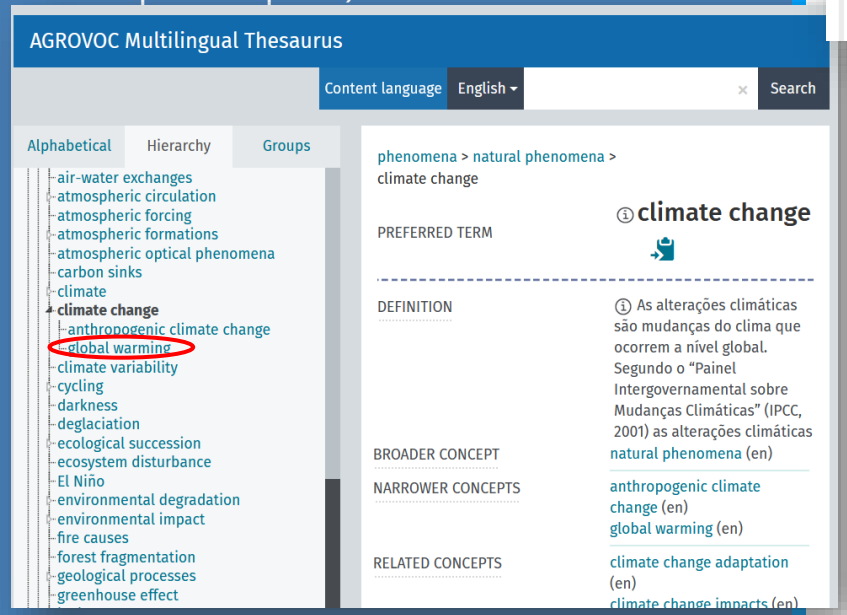
Indexation manuelle

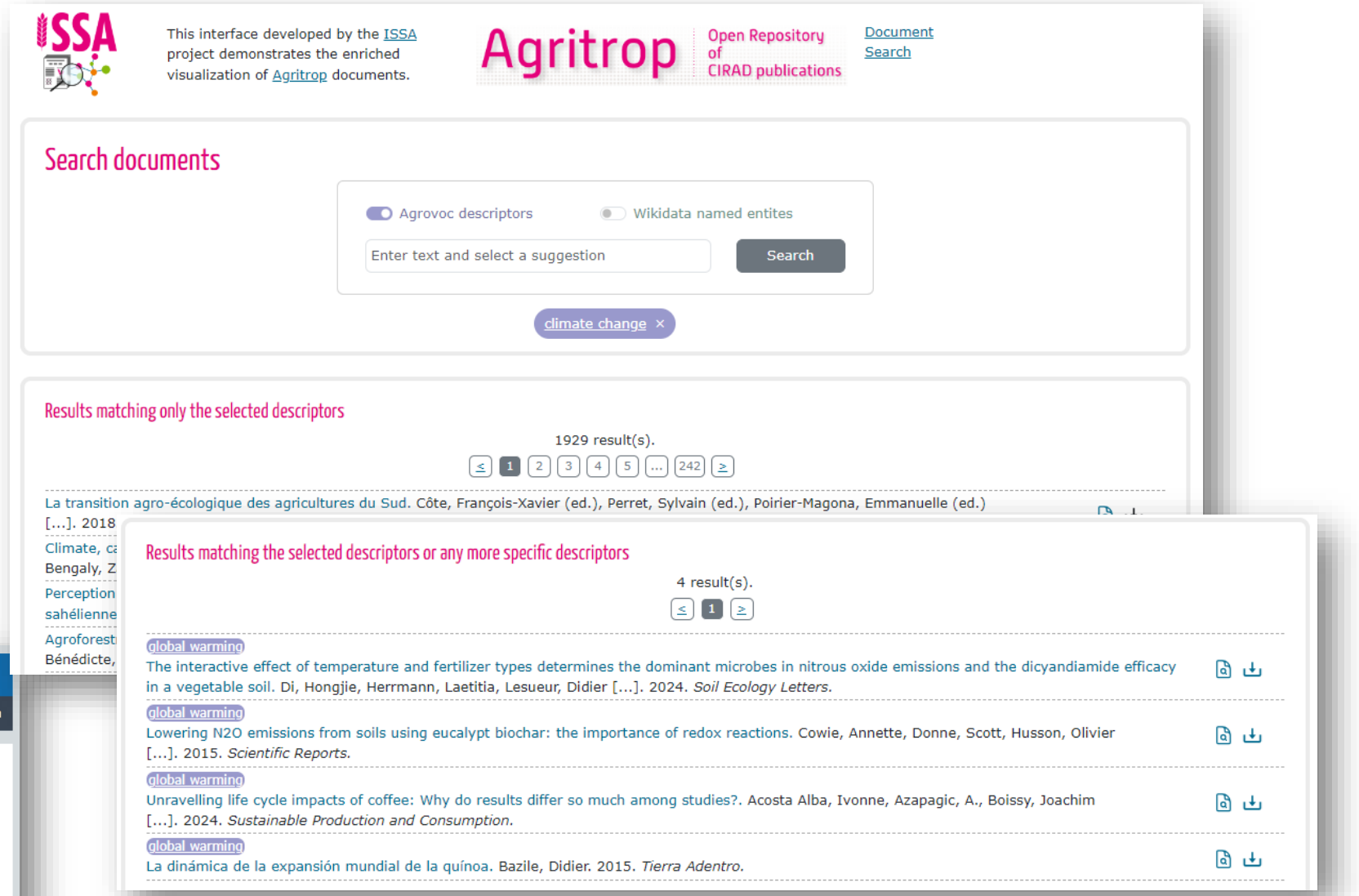
Concept search

- Descripteurs ou Entités nommées
- Multi-langue, synonymes

Concept search

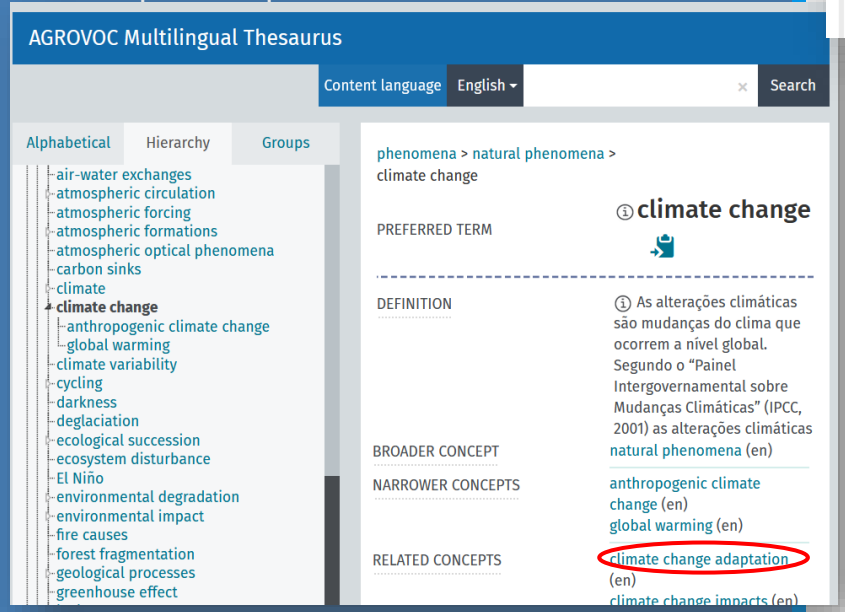
- Descripteurs ou Entités nommées
- Multi-langue, synonymes
- Hiérarchie (termes plus spécifiques)

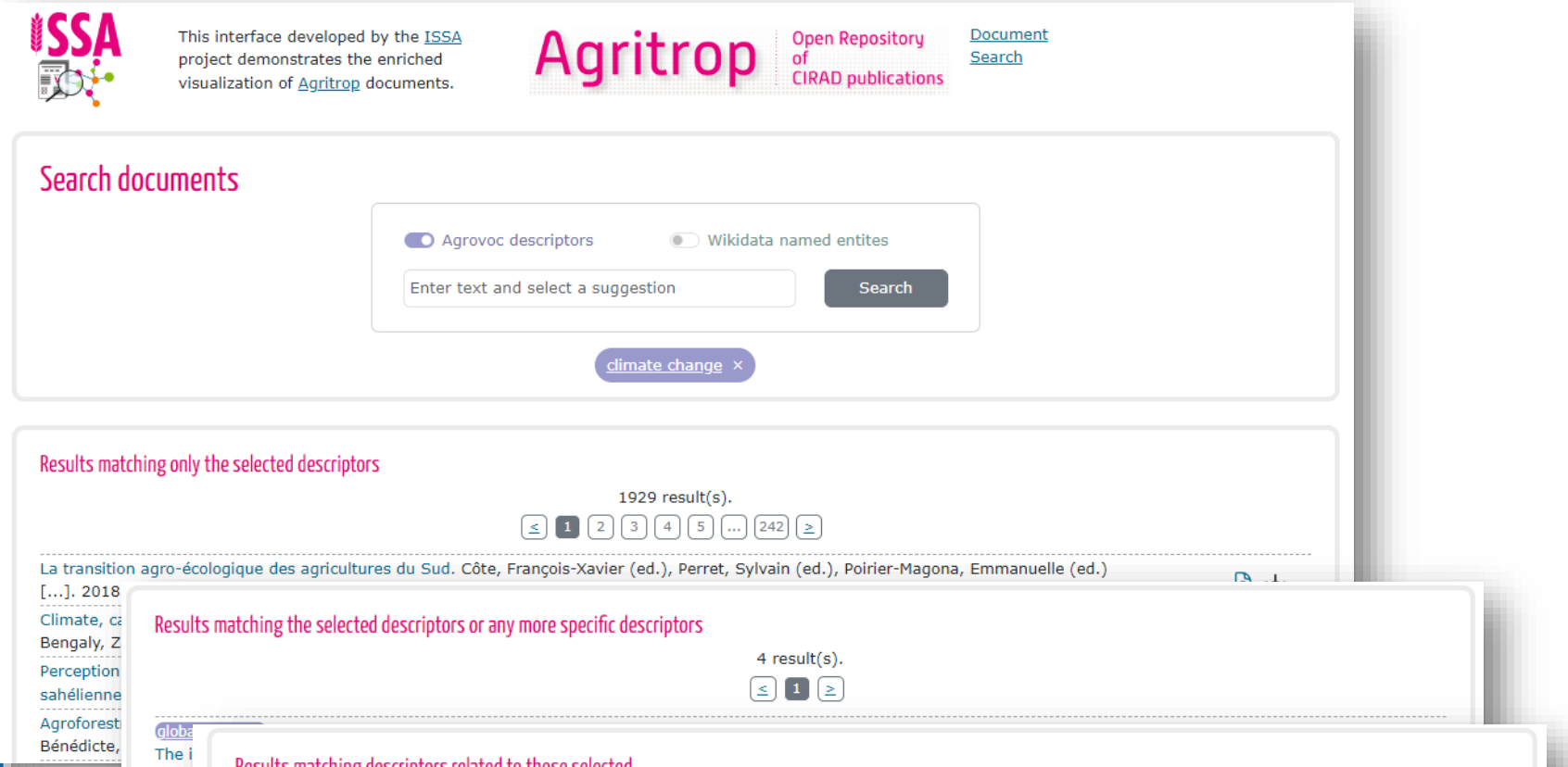


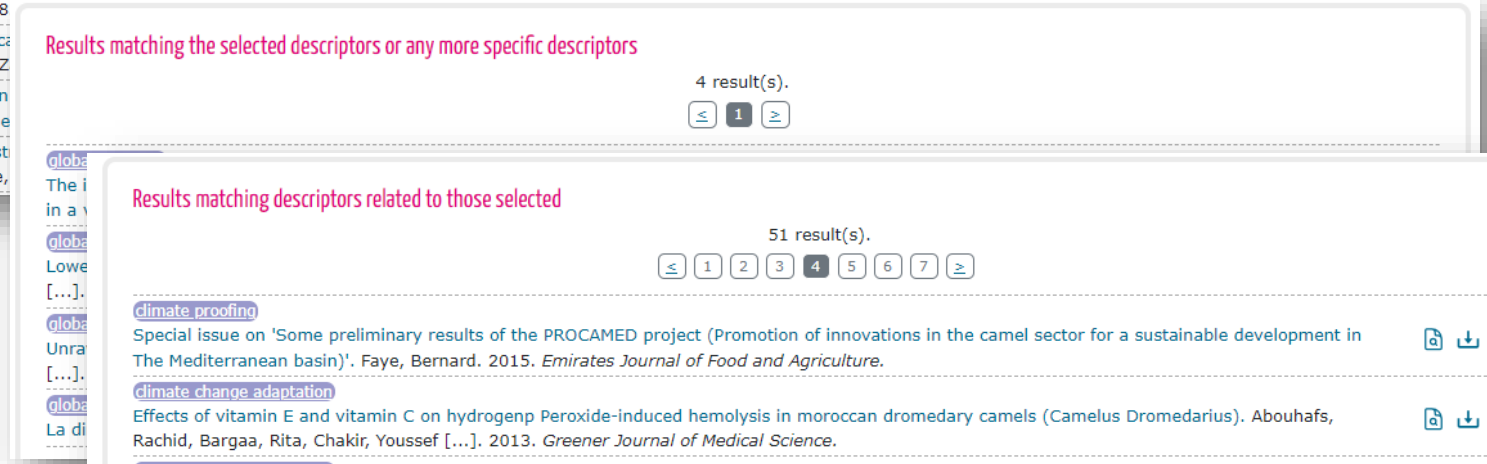


Concept search

- Descripteurs ou Entités nommées
- Multi-langue, synonymes
- Hiérarchie (termes plus spécifiques)

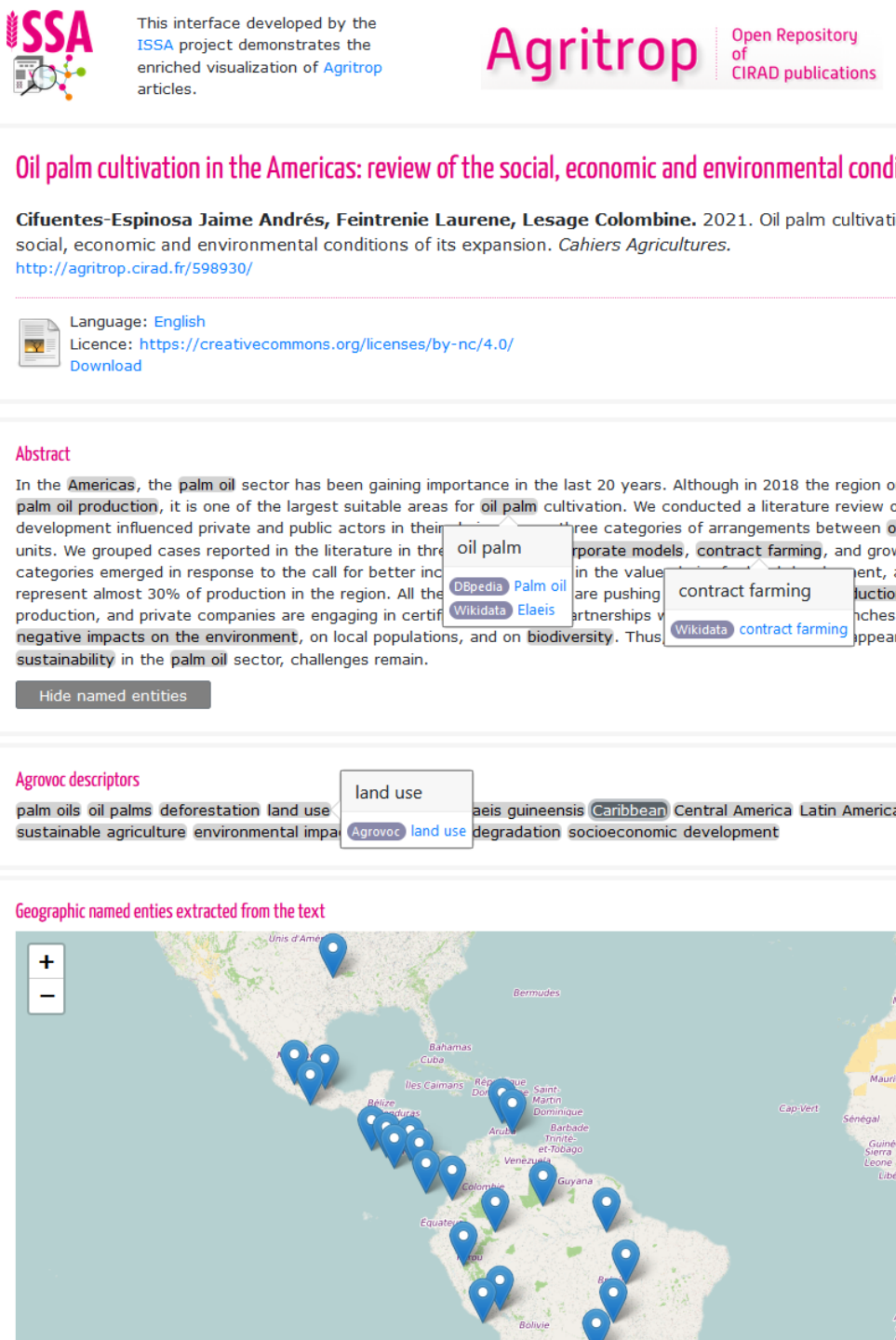






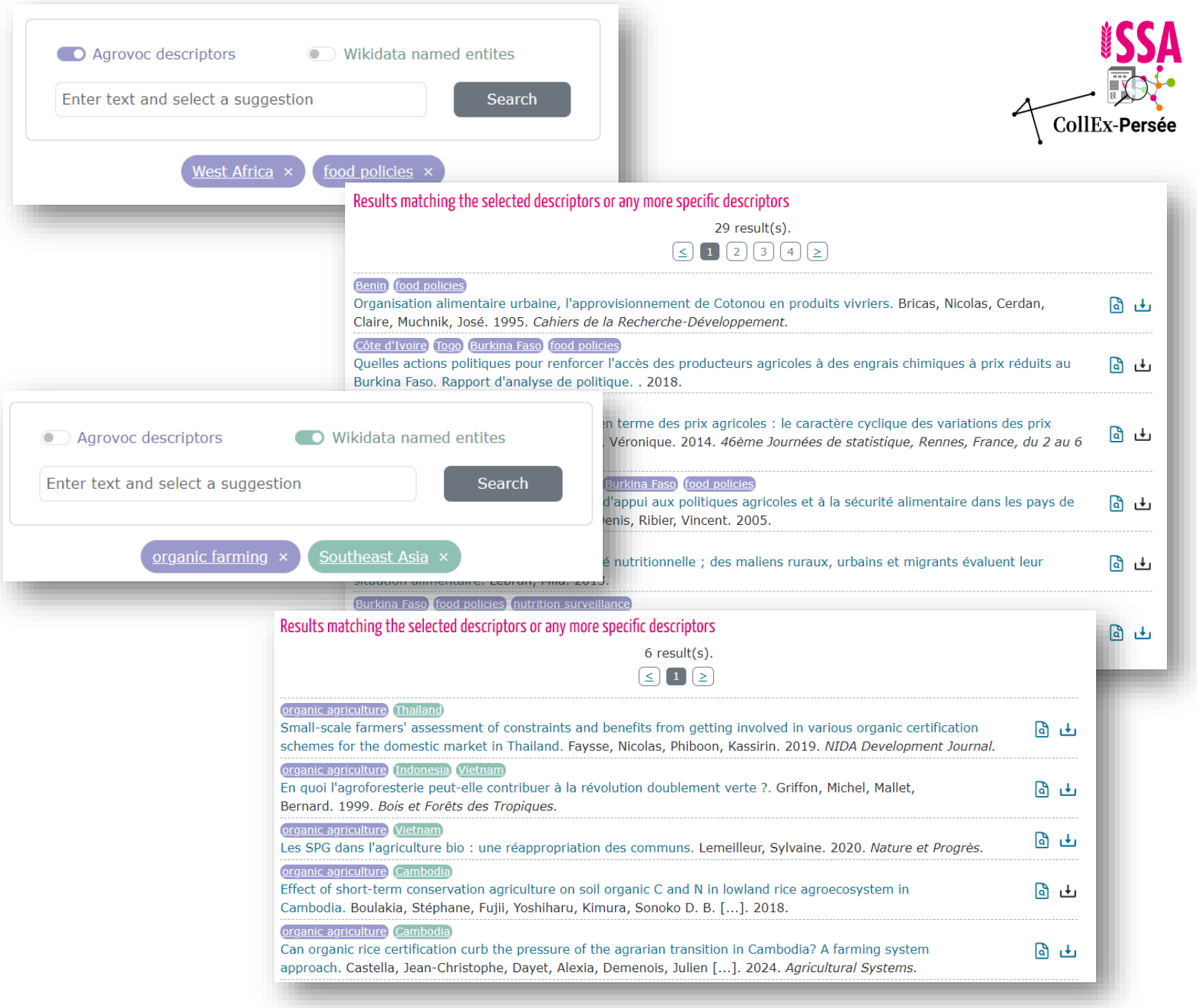
Enriched visualisation of Agritrop records

- Charte graphique d'Agritrop
- Métadonnées
- Entités nommées du résumé
- Descripteurs thématiques et géographiques (documentalistes et calculés)
- Entités nommées géographiques



Geographical search

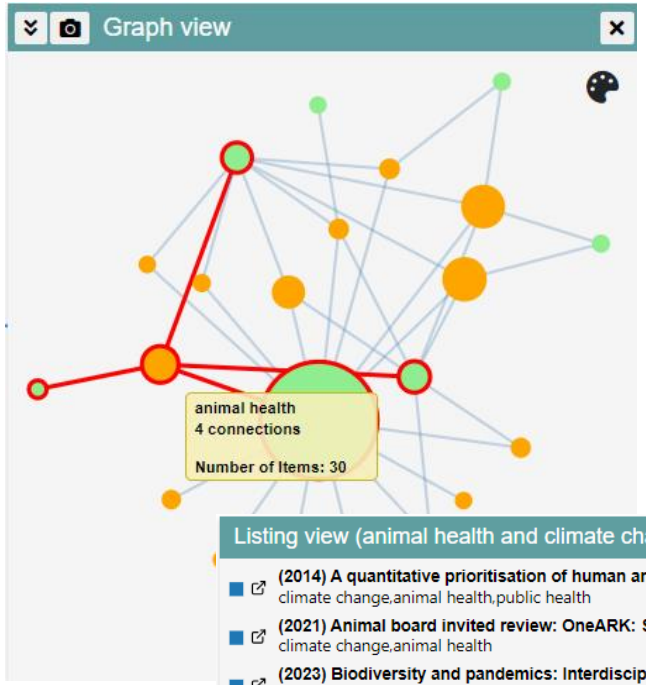
le référentiel contient des termes géographiques

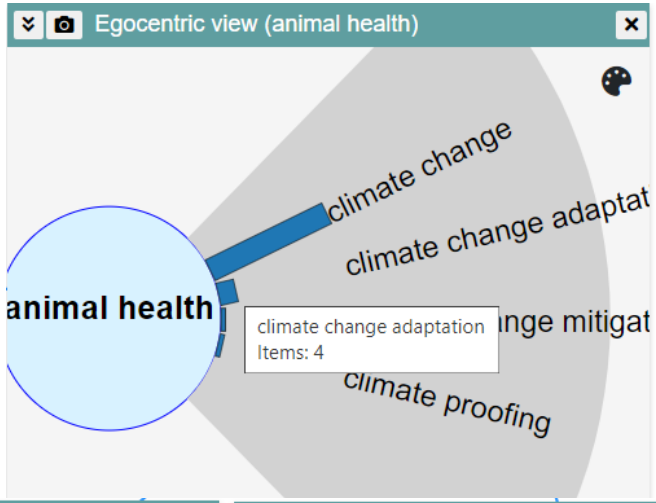


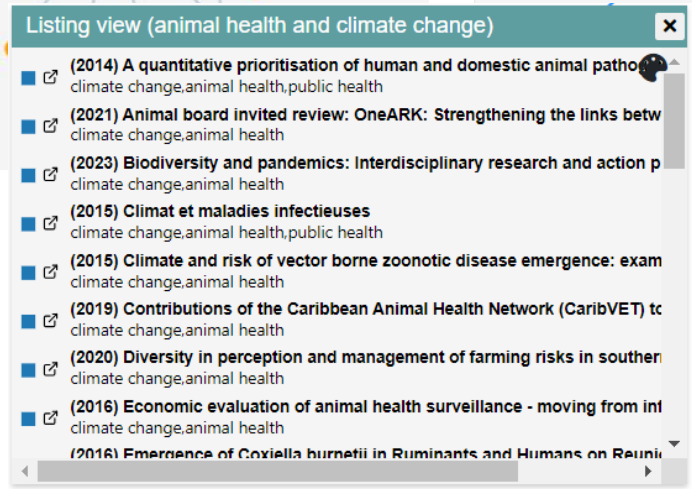
Archive exploration: articles on climate change and health ?

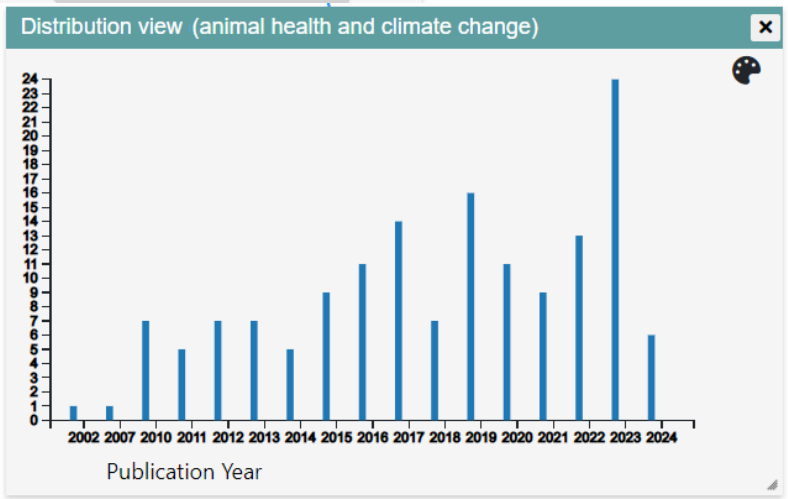
- climate change
  - anthropogenic climate change
  - global warming

- health
  - animal health
  - ecosystem health
  - forest health
  - human health
  - mental health
  - plant health
  - public health
  - rangeland health
  - reproductive health
  - seed health

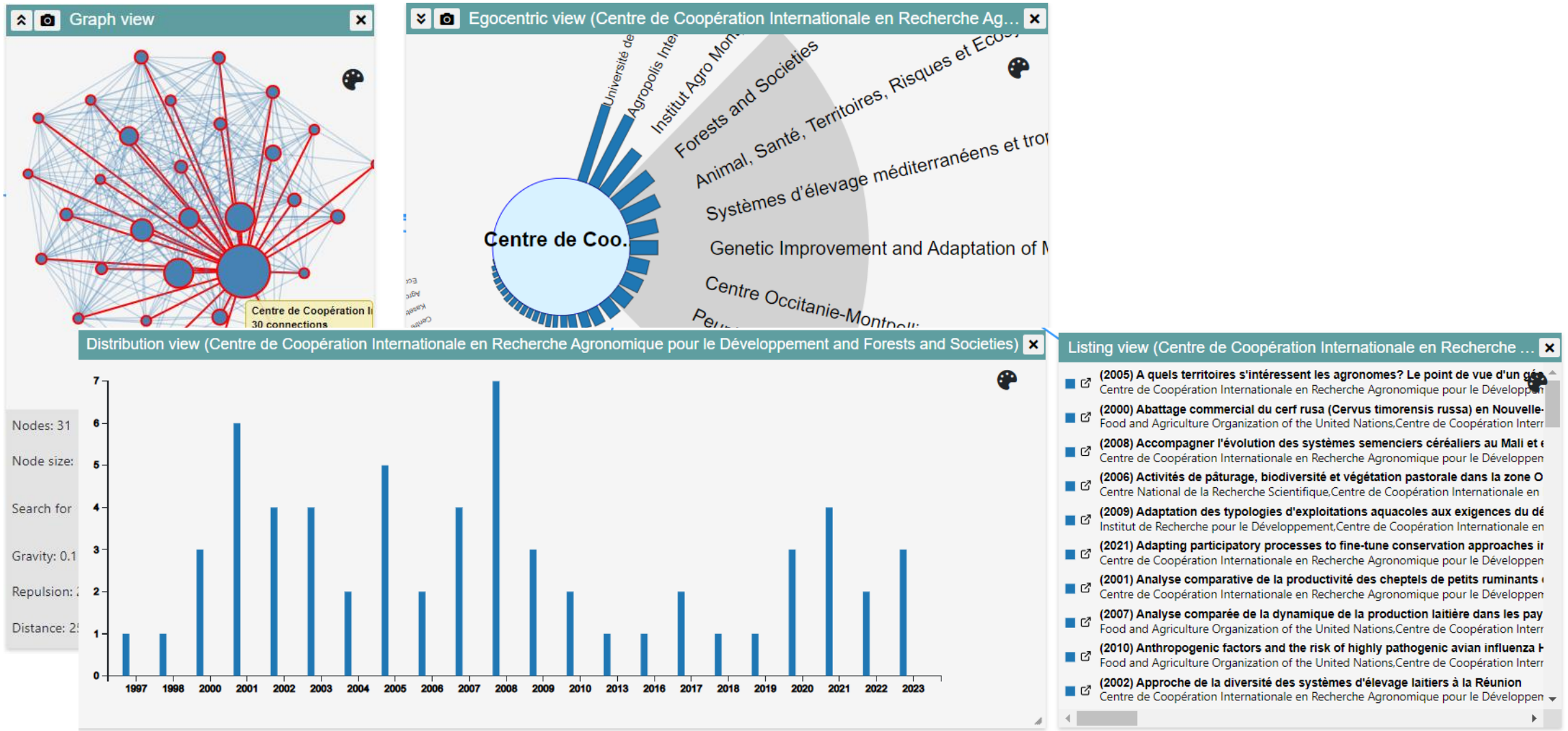








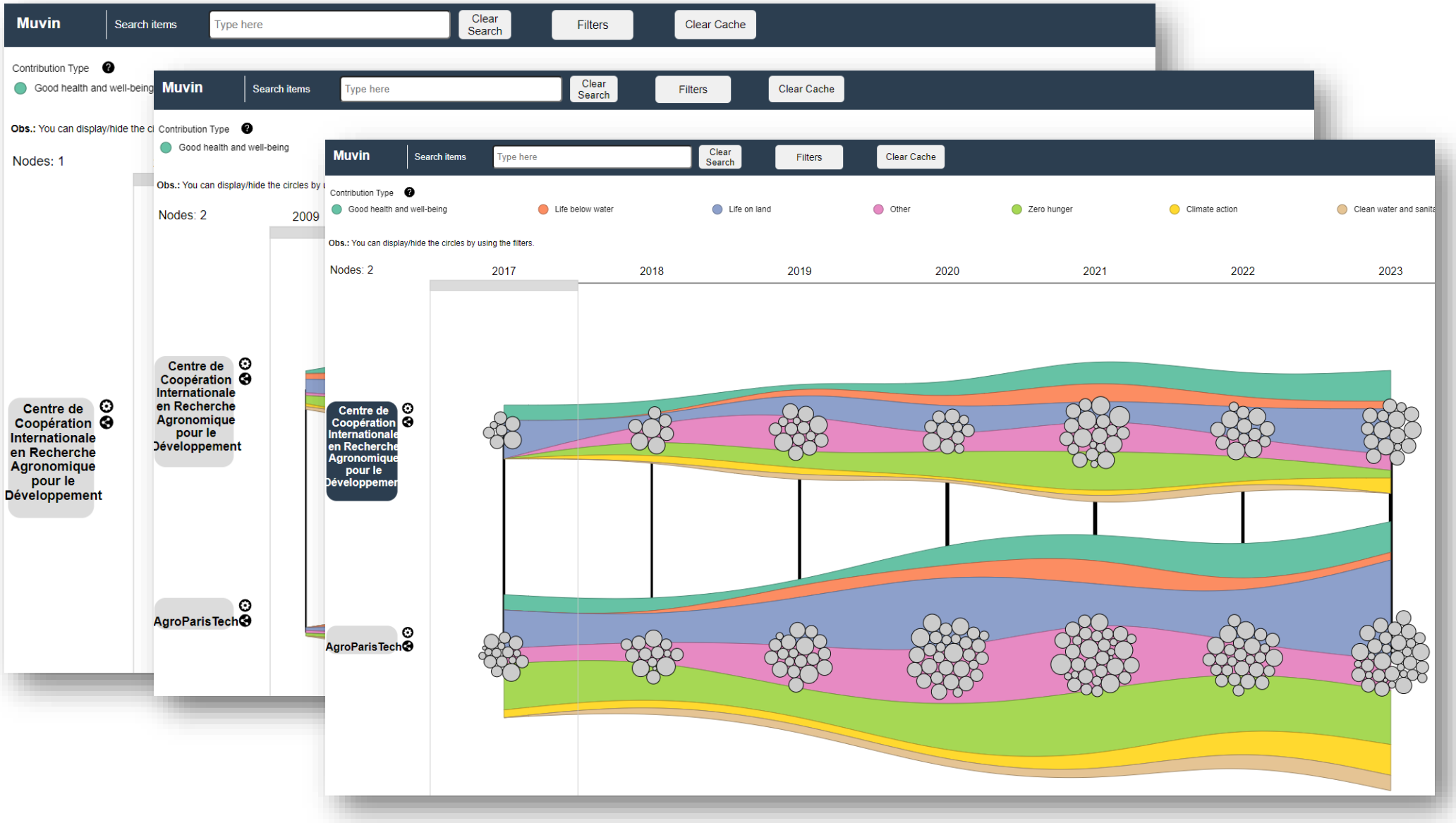
Archive exploration: collaborations with other institutions



Archive exploration: temporal distribution of publications

Sustainable development goals (SDG) in OpenAlex

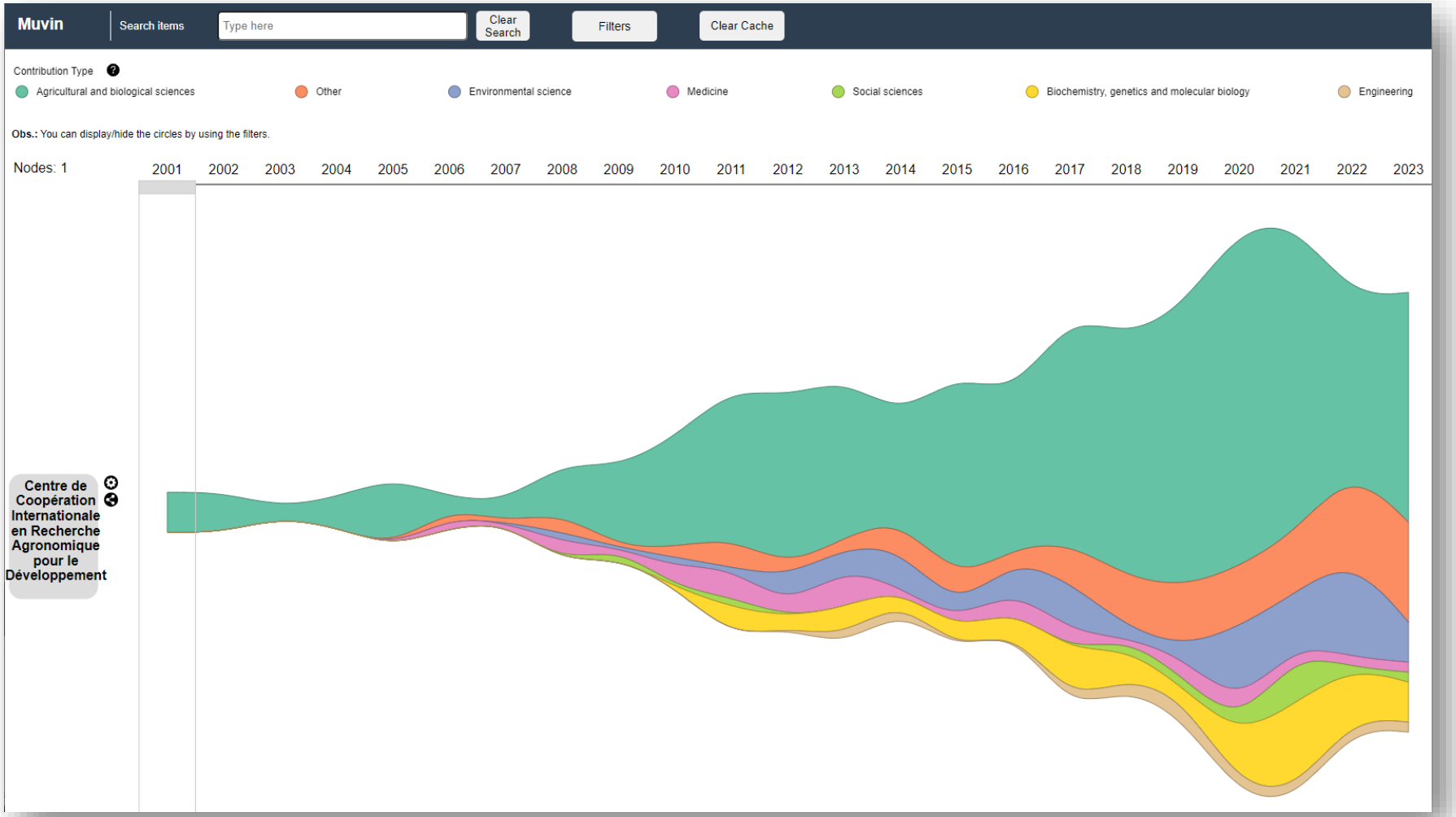
institutes



Archive exploration: temporal distribution of publications

Scientific field)
in OpenAlex

institutes

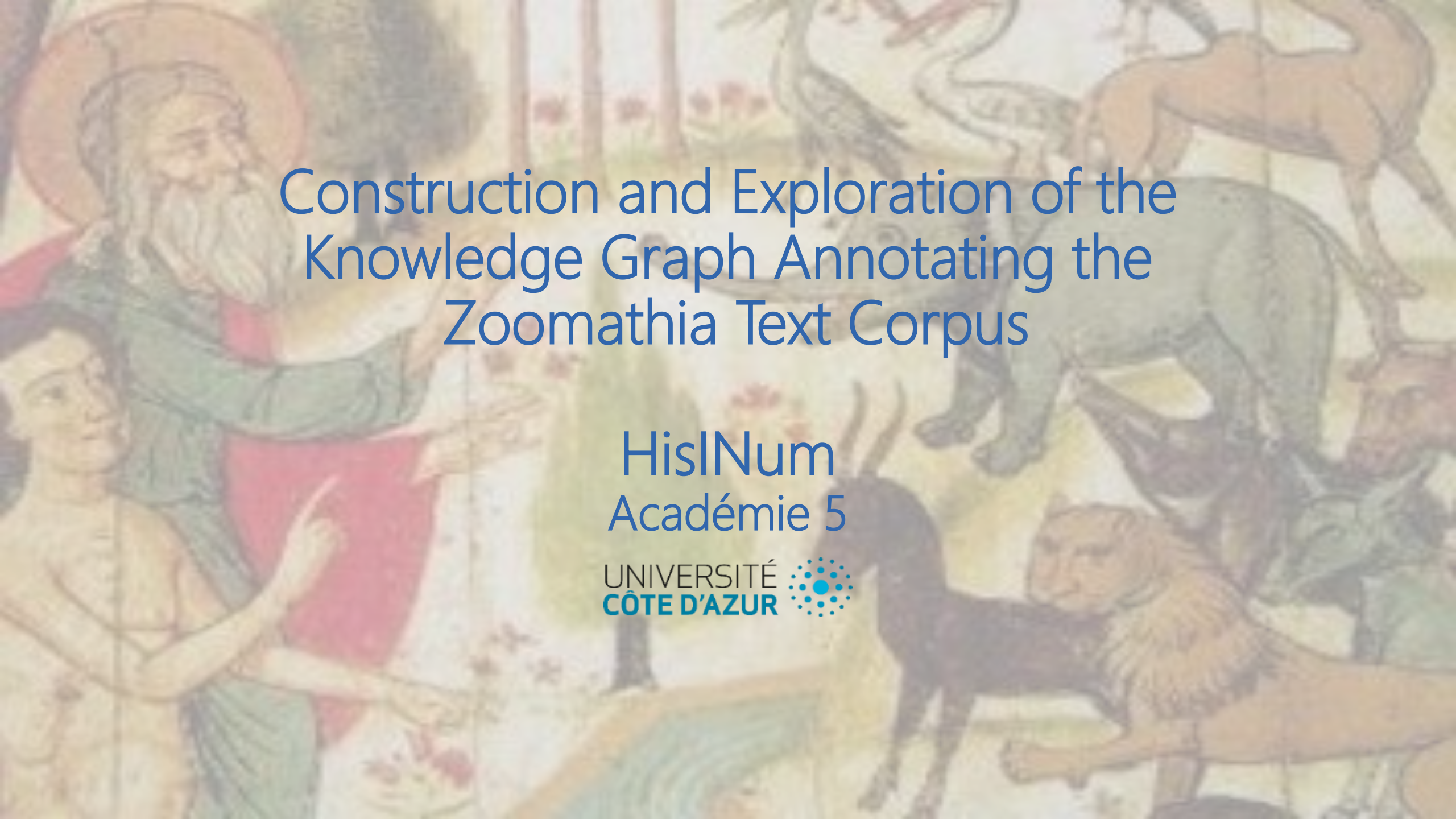


Semantic Indexing of a Scientific Archive in Agriculture

http://issa.i3s.unice.fr/search/

https://dataviz.i3s.unice.fr/mgexplorer/dashboard with SPARQL endpoint https://data-issa.cirad.fr/sparql

Toulet A., Michel F., Bobasheva A., Menin A., Dupré S. ISSA : un graphe de connaissances au service de la recherche bibliographique, EGC 2023



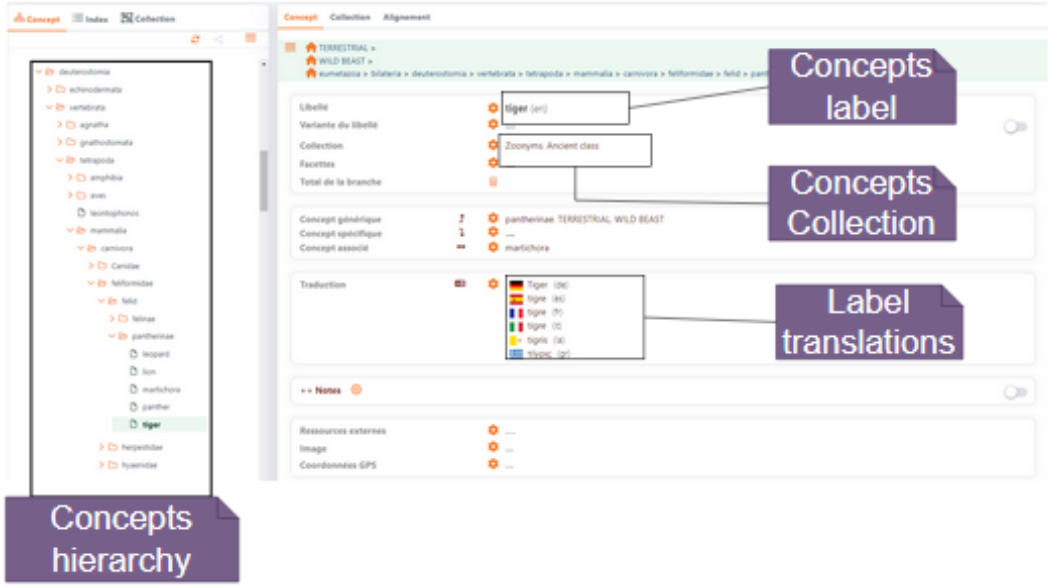
Construction and Exploration of the Knowledge Graph Annotating the Zoomathia Text Corpus

HisINum
Académie 5

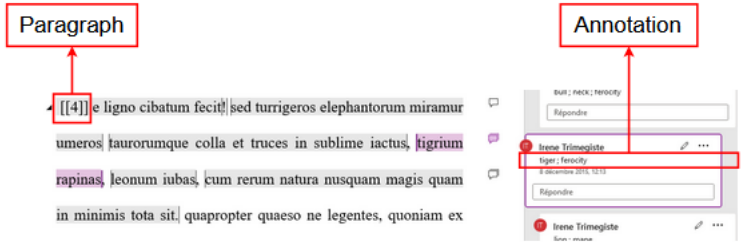
UNIVERSITÉ CÔTE D'AZUR

Semantic annotation of Historia Naturalis of Pliny the Elder

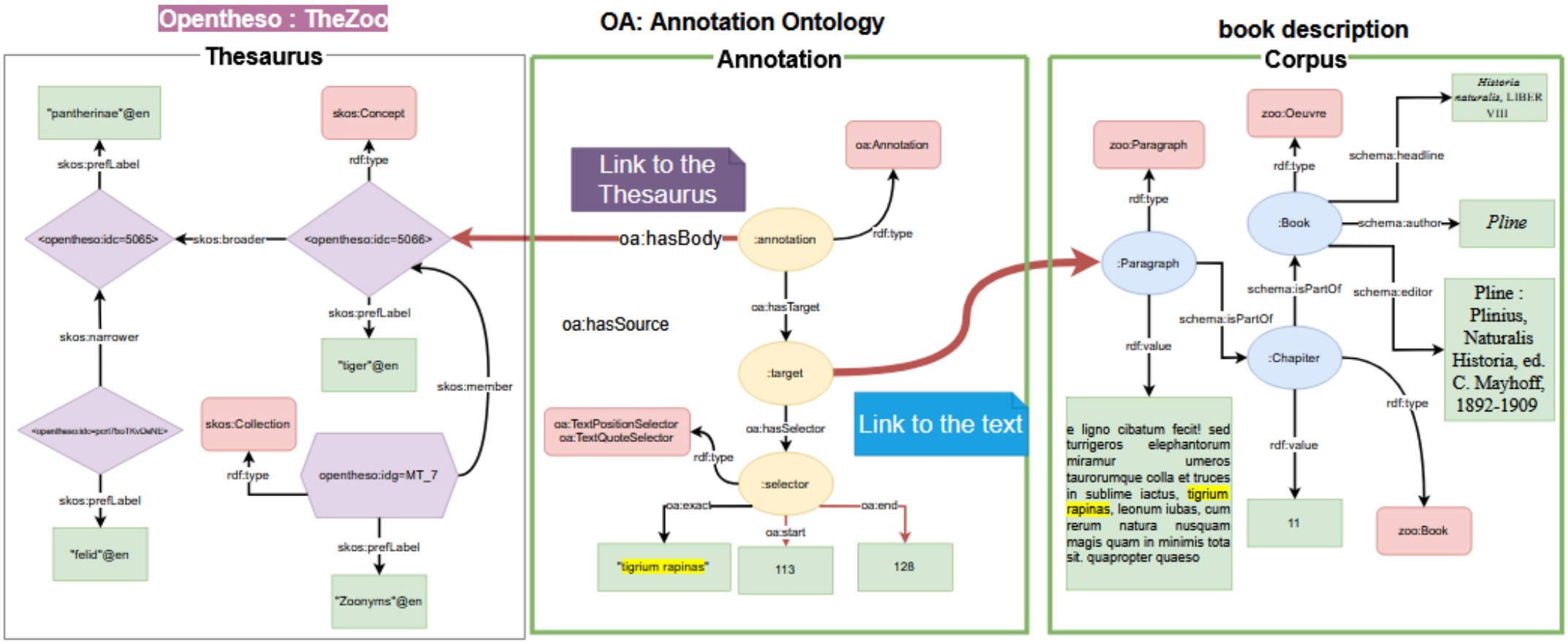
Thesaurus TheZoo



Manual annotation of the text



Construction of the Zoomathia knowledge graph



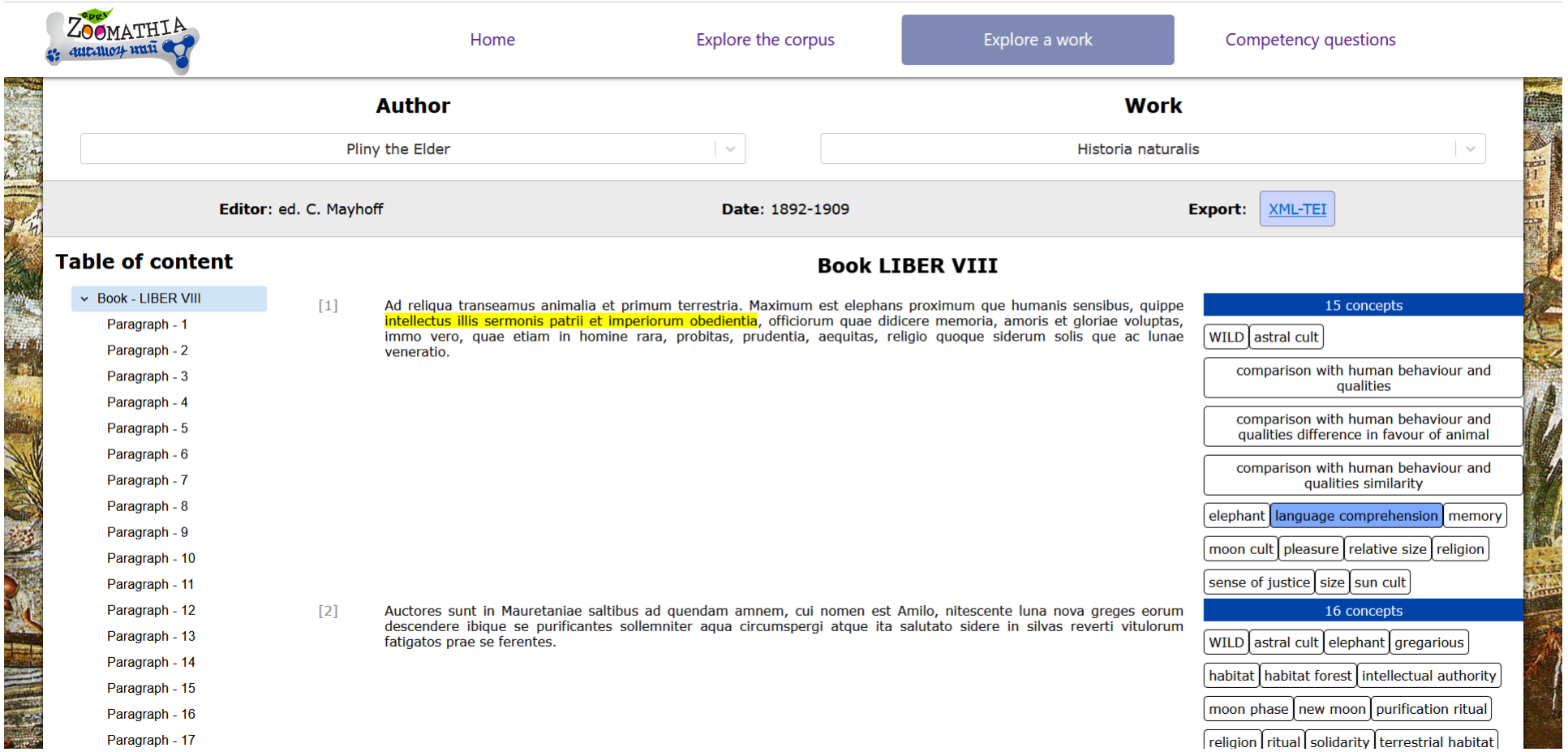
Automatic annotation of the Zoomathia corpus

- Classifier
- Named entity recognition on DBpedia and Wikidata

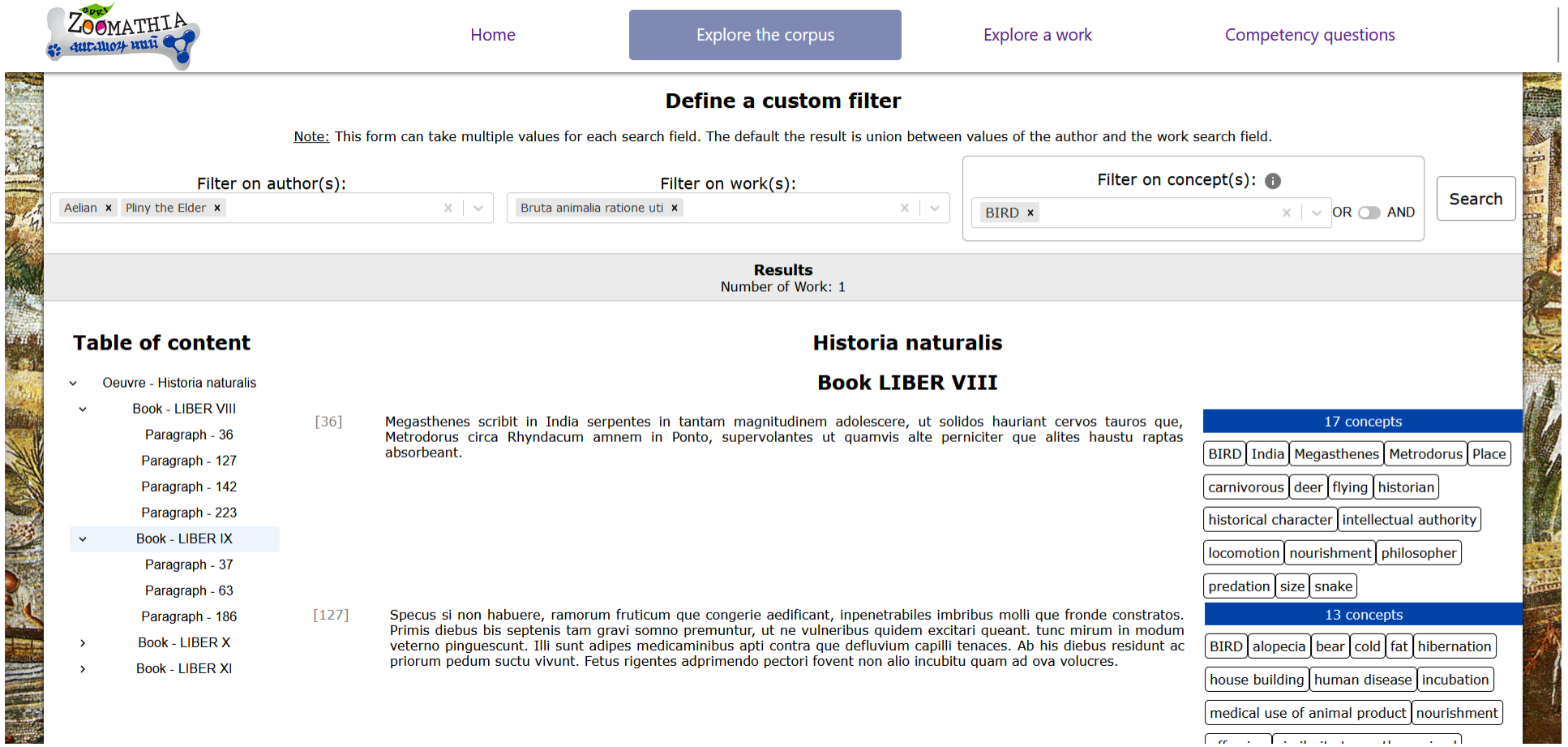
  DBpedia Spotlight

  spaCy fishing (Entity-Fishing)

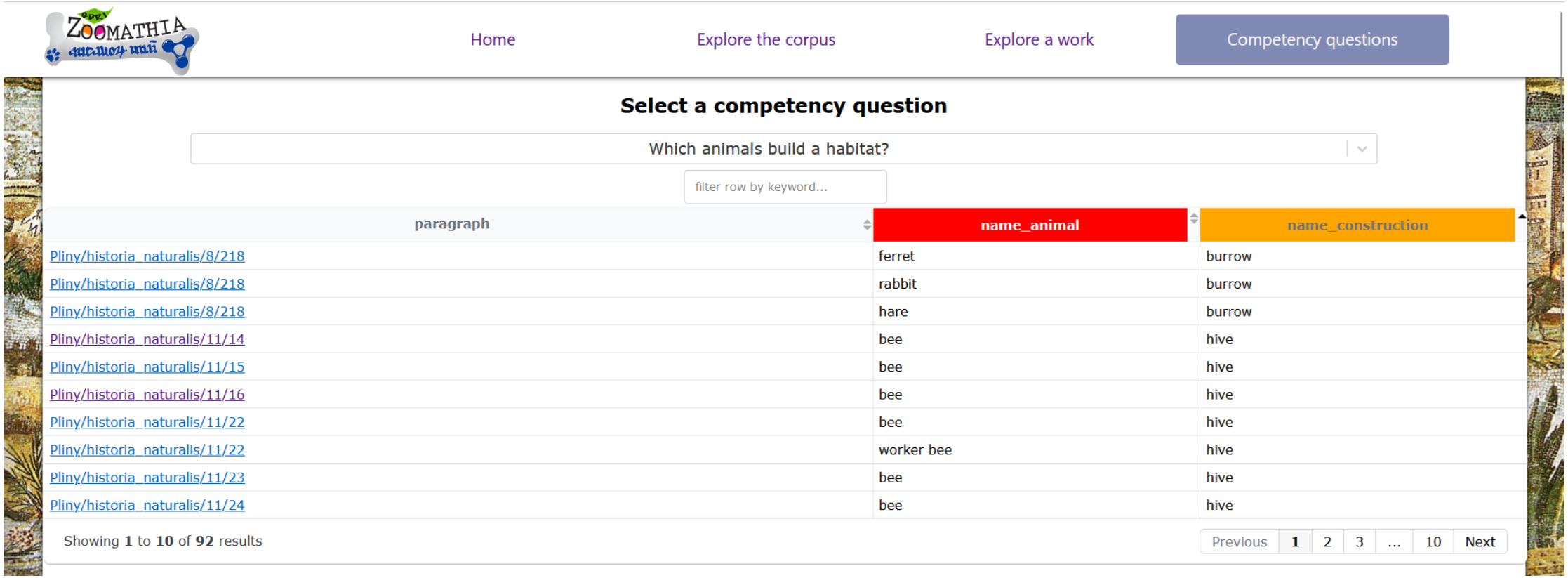
Exploration of the Zoomathia corpus



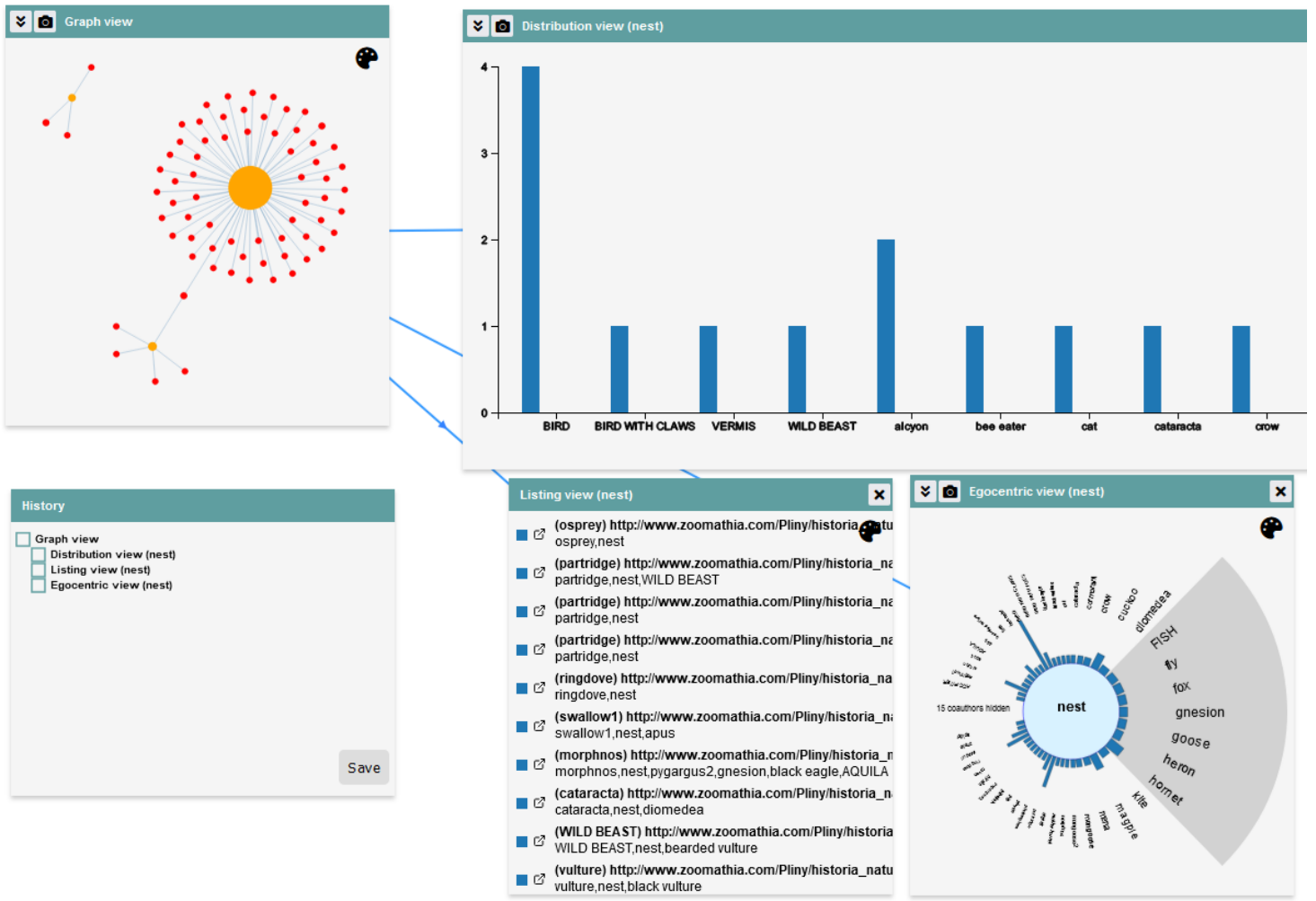
Exploration of the Zoomathia corpus

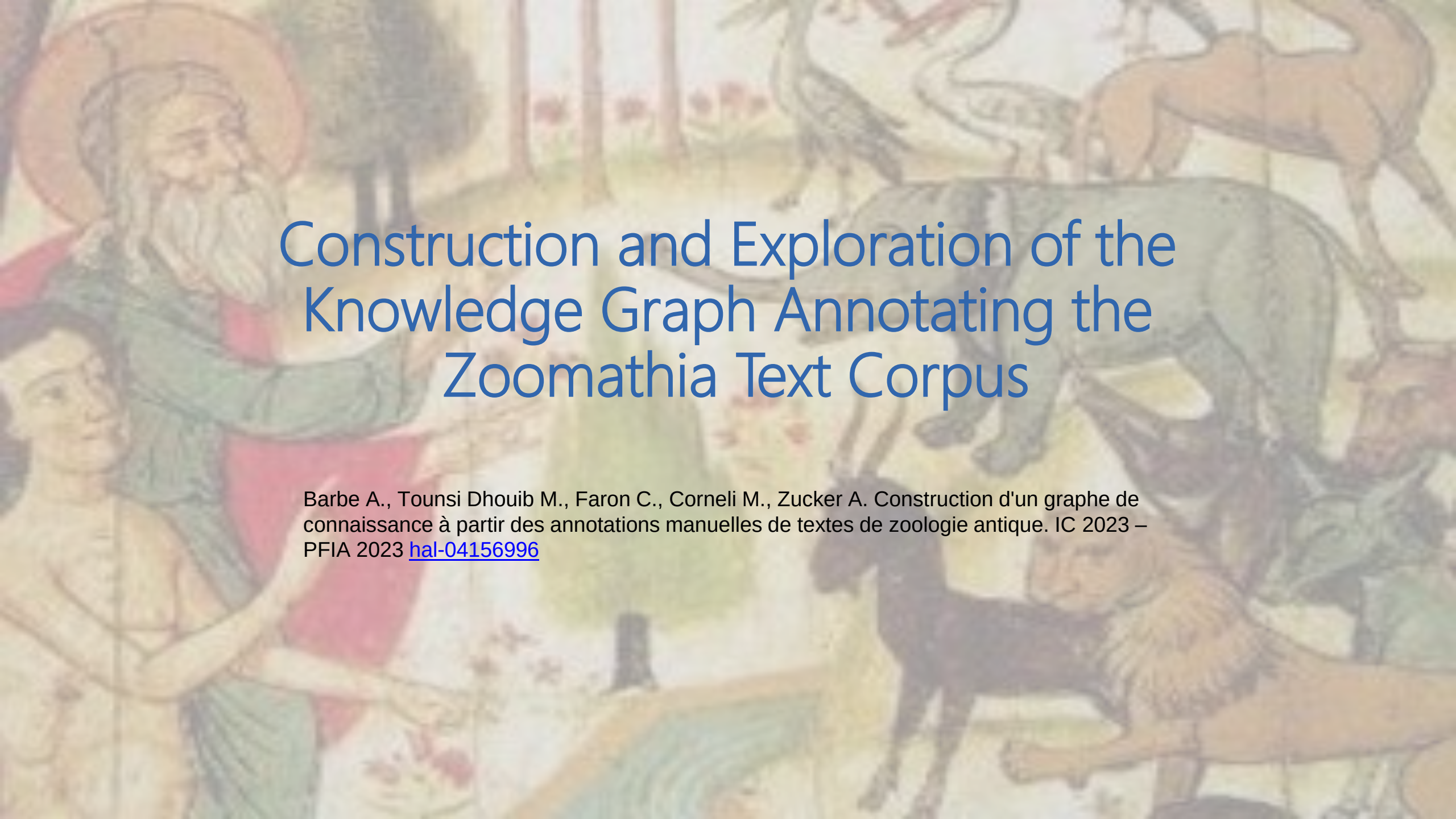


Exploration of the Zoomathia corpus



Exploration of the Zoomathia corpus





Construction and Exploration of the Knowledge Graph Annotating the Zoomathia Text Corpus

Barbe A., Tounsi Dhouib M., Faron C., Corneli M., Zucker A. Construction d'un graphe de connaissance à partir des annotations manuelles de textes de zoologie antique. IC 2023 – PFIA 2023 hal-04156996

Conclusion: a unified (reusable) Knowledge Graph based approach

- Semantic annotation of a text corpus
- Construction of a Knowledge Graph
- Competency questions implemented by queries on the KG
  - Concept search
  - Visualisation of enriched texts
  - Graph exploration (MGExplorer)
  - Temporal analysis (Muvin)

A Knowledge Graph based approach not limited to text annotation

- WeKG-MF: a Knowledge Graph of Observational Weather Data
- CoffeeWKG: a Weather Knowledge Graph for Coffee Regions in Colombia
- WheatObservationsKG: a Knowledge Graph of Crop Monitoring Data
- WoodKG: a Knowledge Graph of Wood and Charcoal Observations
- …

Federation of complementary Knowledge Graphs:

Scientific Literature KG + Observation data KG
e.g. scientific literature on Wheat + crop monitoring data + weather data