



Projet ISSA Lauréat de l'appel à projet 2019-2020  
Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte

# Rapport scientifique du Projet ISSA

## Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte

---

### Table des matières

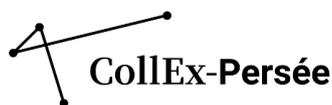
|   |    |
|---|----|
| Présentation synthétique du projet.....           | 1  |
| Description générale du projet.....               | 2  |
| Contexte et motivations.....                      | 2  |
| Objectifs du projet.....                          | 3  |
| Cas d'usage : Agritrop.....                       | 3  |
| Environnement institutionnel et partenariats..... | 4  |
| Résultats scientifiques.....                      | 4  |
| Le pipeline ISSA.....                             | 4  |
| Services de visualisation.....                    | 6  |
| Diffusion et valorisation des résultats.....      | 8  |
| Dissémination.....                                | 8  |
| Communication.....                                | 9  |
| Code source et documentation.....                 | 9  |
| Publications en lien avec le projet.....          | 11 |

### Présentation synthétique du projet

#### Résumé

Les services d'information scientifique et technique ont toujours été confrontés aux questions de gestion documentaire. Leur mission leur confère une responsabilité dans la conservation, la mise à disposition et la valorisation de ressources culturelles ou issues de la recherche. Dans cette optique, ce projet s'intéresse à l'indexation sémantique des publications dans une archive ouverte et à l'exploitation de cette indexation pour développer des services aux chercheurs et aux documentalistes dans une perspective de science ouverte. La méthodologie mise en œuvre, adossée aux principes FAIR, se veut résolument générique : elle est exploitable par toute communauté adoptant ces mêmes principes au travers de métadonnées interopérables, en particulier dans le cadre de systèmes de gestion de ressources documentaires.

L'objectif du projet est triple :



## Projet ISSA Lauréat de l'appel à projet 2019-2020 Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte

- (1) Fournir un **pipeline générique, réutilisable et extensible pour l'analyse des documents d'une archive scientifique ouverte** ;
- (2) Traduire les résultats en un **index sémantique sous la forme d'un graphe de connaissances RDF** ;
- (3) Développer des **services innovants de recherche et de visualisation** exploitant cet index.

Les traitements font appel à diverses techniques d'intelligence artificielle : TALN, ingénierie des connaissances, web sémantique. Les métadonnées et le texte intégral des publications sont traités afin d'en extraire des descripteurs thématiques<sup>1</sup> et des entités nommées<sup>2</sup>. Pour exploiter au mieux la puissance du web sémantique, les descripteurs thématiques et les entités nommées sont liés à des référentiels sémantiques (bases de connaissance, ontologies, thésaurus) tels que [Wikidata](#), [DBpedia](#) et [GeoNames](#). Le graphe de connaissances résultant sert de clé de voûte au développement de services de recherche et de visualisation. Agritrop, l'archive ouverte des publications du Cirad, a servi de cas d'usage et de preuve de concept tout au long du projet.

### Coordination scientifique

- **Anne Toulet**, Cirad (Délégation à l'information scientifique et technique), Ingénieure de recherche, Ingénierie des connaissances et web sémantique.
- **Franck Michel**, Université Côte d'Azur, CNRS, Inria (Wimmics), Ingénieur de recherche CNRS, Ingénierie des connaissances et web sémantique.
- **Andon Tchechmedjiev**, IMT MINES ALÈS (LGI2P), Enseignant-chercheur, Ingénierie des connaissances et Traitement Automatique des Langues.

**Dates du projet** : du 1<sup>er</sup> octobre 2020 au 31 décembre 2022

**Subvention CollEx-Persée** : 94 958 €

**Recrutements et stages spécifiques au projet** : 1 Ingénieur Inria (CDD 12 mois), 3 stages de Master 2, 1 stage de licence professionnelle « Métiers de l'information, veille et gestion des ressources numériques ».

## Description générale du projet

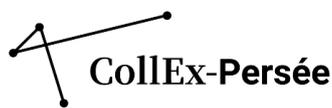
### Contexte et motivations

Ces dernières années, l'accélération du rythme des publications et le « tout numérique » ont radicalement transformé la façon d'interagir avec la littérature scientifique, et les utilisateurs ont dû adapter leurs pratiques en particulier en termes de recherche bibliographique. Cependant, les services de recherche classiques permettant de fouiller la masse des publications produites ne parviennent souvent pas à saisir la richesse des associations sémantiques entre les articles. Certaines recherches complexes trouvent difficilement des réponses et les résultats parfois peu pertinents obligent l'utilisateur à un filtrage manuel fastidieux. Devant ces difficultés, il est important de proposer aux chercheurs et aux professionnels de l'information des outils permettant de s'y retrouver.

---

<sup>1</sup> Les descripteurs thématiques sont des mots-clés liés à des vocabulaires de référence, des thésaurus ou des ontologies, qui caractérisent un article dans son ensemble.

<sup>2</sup> Mention de concepts identifiés dans un texte



## Projet ISSA Lauréat de l'appel à projet 2019-2020 Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte

Dans ce contexte, les bases de données bibliographiques, les moteurs de recherche et les archives ouvertes occupent une place de premier plan. Avec l'avènement du Web de données, les établissements de recherche et les établissements de mission culturelle se sont tournés depuis une dizaine d'années vers des solutions de gestion de leurs ressources s'appuyant sur les technologies du Web sémantique pour structurer, publier sur le web, relier et rendre exploitables par l'homme et les machines, des informations issues de différentes sources. Cette dynamique pour un accès et une interopérabilité accrues répond aux objectifs du plan national pour la science ouverte<sup>3</sup>. Dans cet écosystème, les ressources sémantiques (ontologies, vocabulaires contrôlés, thésaurus, taxonomies, listes d'autorité, etc.)<sup>4</sup> occupent une place de premier plan et leur mobilisation est un élément clé pour la gestion et l'ouverture des données. Elles permettent leur structuration, leur accessibilité, leur partage, leur interopérabilité et leur réutilisation selon les principes FAIR (Findable, Accessible, Interoperable and Reusable). Elles constituent une réponse aux enjeux plus larges de la gestion des connaissances. Par ailleurs, l'exploitation de ces ressources sémantiques est largement encouragée par les politiques institutionnelles dans une visée d'ouverture des données. Elle est au centre de nombreux projets, tant au niveau national qu'international, dans lesquels nos institutions sont majoritairement impliquées. L'objectif est de « comprendre » les données afin non seulement d'améliorer la pertinence des résultats de recherches, mais également d'offrir des services plus riches et contextualisés.

### Objectifs du projet

S'inscrivant dans une démarche de science ouverte, **le présent projet** poursuit plusieurs objectifs : (1) fournir un **pipeline générique, réutilisable et extensible permettant l'indexation automatique des publications d'une archive ouverte** avec des mots-clés thématiques et géographiques issus des **ressources sémantiques adaptées à chaque domaine et communauté scientifique** ; (2) traduire le résultat en un index sémantique représenté et publié sous la forme d'un graphe de connaissances RDF ; (3) démontrer l'intérêt d'une telle démarche en développant des **services capables d'exploiter cette indexation sémantique** pour produire des outils innovants de recherche et de visualisation à destination des utilisateurs.

### Cas d'usage : Agritrop

L'**archive ouverte des publications du Cirad, Agritrop** a servi de cas d'usage et de preuve de concept tout au long du projet ISSA. Créée en 1986, la base des publications du Cirad a évolué en 2015 vers une archive ouverte institutionnelle. En 2020, avec plus de 106 000 références, 45 500 documents en texte intégral dont 27 000 accessibles à tous sur le Web, Agritrop offrait une large visibilité aux publications du Cirad.

Actuellement dans Agritrop, le catalogage et l'indexation par mots-clés thématiques et géographiques sont effectués manuellement par des documentalistes. L'indexation s'appuie sur le **thésaurus Agrovoc**, vocabulaire multilingue développé et maintenu par la **FAO** qui fait référence au niveau mondial dans le domaine de l'agronomie. Cette indexation sémantique, adossée à un vocabulaire standard très largement adopté, permet l'exploitation et l'affichage de termes en plusieurs langues (Agrovoc propose jusqu'à 29 traductions par concept). Les objectifs du projet ISSA (automatisation de l'indexation, ouverture vers d'autres

<sup>3</sup> <https://www.ouvrirlascience.fr/>

<sup>4</sup> Nous désignons ces différents systèmes d'organisation des connaissances sous le nom de *référentiels ou ressources sémantiques*.



## Projet ISSA Lauréat de l'appel à projet 2019-2020 Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte

référentiels terminologiques, et développement de services basés sur l'indexation) présentent donc un intérêt majeur tant pour les chercheurs que pour les documentalistes.

### Environnement institutionnel et partenariats

Le projet ISSA associe trois partenaires institutionnels complémentaires : la [Délégation à l'information scientifique et technique du Cirad](#) (Dist - Cirad), l'équipe de recherche [Wimmics](#) du centre Inria Sophia Antipolis Méditerranée et le centre de recherche et d'enseignement [LGI2P](#) de l'[IMT Mines Alès](#).

Le [Cirad](#) est un organisme français de recherche agronomique et de coopération internationale pour le développement durable des régions tropicales et méditerranéennes. Depuis 2006, avec la signature de la déclaration de Berlin sur le libre accès aux connaissances, le Cirad est engagé dans une démarche de science ouverte. Il a inscrit ce principe dans sa stratégie institutionnelle. Ses équipes de recherche sont encouragées à diffuser aussi ouvertement que possible leurs résultats scientifiques. La délégation à l'information scientifique et technique (Dist) les accompagne tout en concevant, administrant et faisant évoluer les systèmes d'information scientifiques et leurs services associés. Dans cet objectif, l'adoption de standards et des technologies du Web sémantique selon les principes FAIR (Findable, Accessible, Reusable, Interoperable) est cruciale.

[Wimmics](#) est une équipe-projet commune à [Inria](#) (Sophia Antipolis Méditerranée) et au laboratoire I3S (UMR 7271 Université Nice Côte d'Azur, CNRS). Ses travaux visent à établir des ponts entre deux formes de sémantique qui coexistent sur le Web : d'un côté la sémantique formelle qui inclut la modélisation des connaissances et le raisonnement formel, d'un autre côté la sémantique sociale associée à l'information produite par les humains.

Le [LGI2P](#) de l'[IMT Mines Alès](#) axe une partie de ses recherches sur l'utilisation de modélisations sémantiques et d'ontologies dans la recherche d'information, le filtrage, la visualisation ou encore la recommandation. En parallèle de ces thématiques, le LGI2P a acquis une compétence forte en Traitement Automatique du Langage Naturel (TAL), à la fois en recherche (académique et partenariale) et en enseignement dans son département 2IA/IA-Sciences des Données.

Ainsi, le projet s'appuie d'une part sur l'expérience de la Délégation à l'information scientifique et technique du Cirad en termes de gestion de ressources documentaires, d'autre part sur l'expertise des scientifiques de l'équipe Wimmics et du laboratoire LGI2P.

### Résultats scientifiques

#### Le pipeline ISSA

Le projet ISSA a permis le développement d'un pipeline pour l'analyse des documents d'une archive scientifique ouverte. Ce pipeline est conçu pour être **générique** (non lié à un domaine ou une communauté), **réutilisable** (on peut le connecter à une archive autre que celle qui a servi de cas d'usage) et **extensible** (on peut y ajouter de nouvelles étapes de traitement).

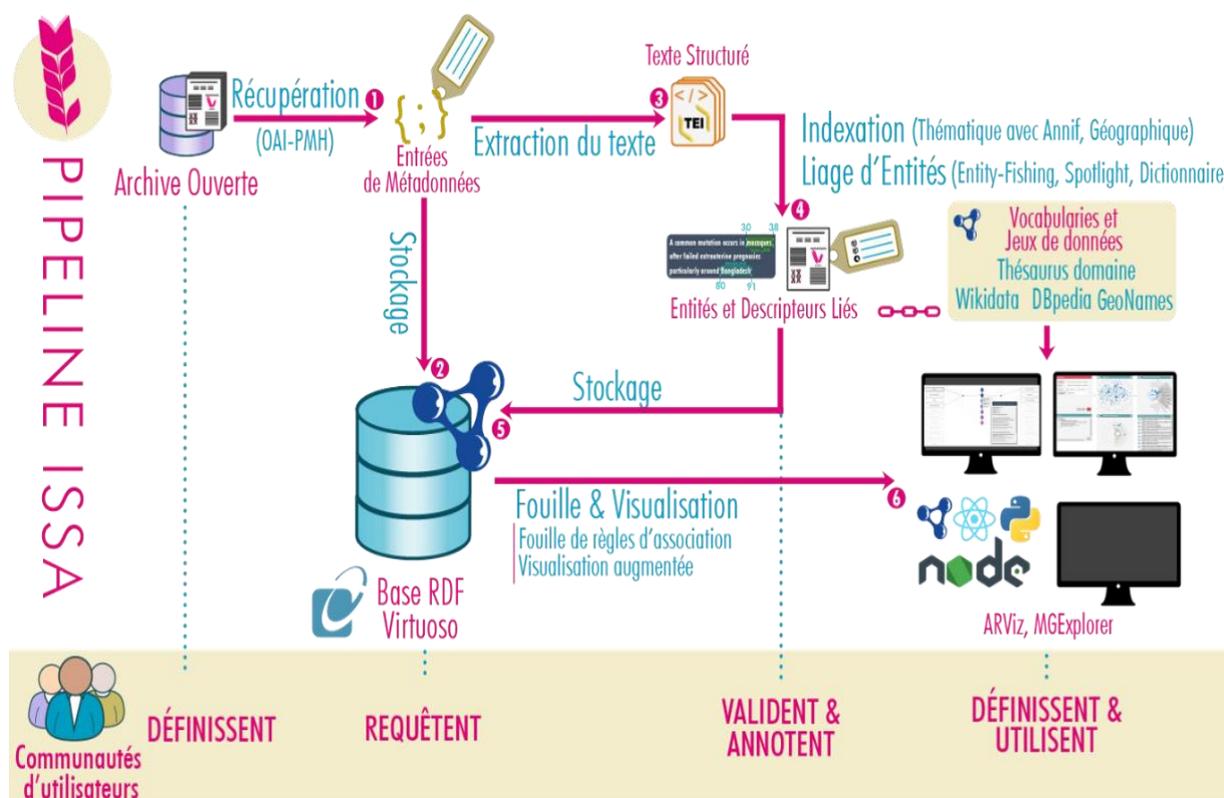
Les métadonnées et le texte intégral des publications sont traités afin d'en extraire des **descripteurs thématiques** (des mots-clés liés à des ressources sémantiques, qui caractérisent un article dans son ensemble), et des **entités nommées** (EN, des mentions de texte liées aux concepts des ressources

**Projet ISSA Lauréat de l'appel à projet 2019-2020**  
**Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte**

sémantiques). Les descripteurs thématiques et les entités nommées sont liés à trois référentiels sémantiques généralistes, applicables à toute archive quel que soit le domaine de recherche considéré : Wikidata, DBpedia et GeoNames. D'autres ressources sémantiques spécifiques au domaine peuvent être intégrées à chaque étape.

Le pipeline intègre et exploite des outils existants suffisamment matures pour une utilisation en production. La Figure 1 décrit chaque étape en détails :

- (1) Les métadonnées sont extraites de l'archive ouverte. Le pipeline est livré avec un connecteur compatible avec le protocole OAI-PMH implémenté par de nombreuses archives ouvertes.
- (2) Ces métadonnées sont traduites en RDF et stockées dans une base de données RDF Virtuoso.
- (3) Le texte intégral est extrait des PDFs à l'aide de l'outil *Grobid*.
- (4) Pour chaque article, les descripteurs thématiques sont extraits à l'aide du système de classification *Annif* développé par la Bibliothèque nationale de Finlande. Les entités nommées sont extraites du texte à l'aide



**Figure 1. Étapes du pipeline de traitement d'une archive ouverte**

de trois outils : *DBpedia Spotlight* annote les textes en huit langues différentes avec des entités DBpedia ; *Entity-fishing* identifie et désambiguïse les EN avec Wikidata ; enfin, *pyclinrec* implémente une annotation par projection sur dictionnaire, il est utilisé pour l'extraction et le liage d'EN avec un vocabulaire spécifique au domaine. Dans notre cas d'usage, il s'agit du thésaurus Agrovoc. Une étape

## Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte

supplémentaire de post-traitement identifie spécifiquement les entités géographiques et leurs correspondances dans GeoNames.

- (5) Les descripteurs et les EN sont traduits en un ensemble unifié de données RDF stockées dans le serveur Virtuoso avec les enregistrements de métadonnées.
- (6) Une fois publié, le graphe de connaissances peut être exploité pour proposer des services de visualisation et d'exploration.

La transformation en RDF est faite à l'aide du logiciel *Morph-xR2RML* et s'appuie sur des vocabulaires couramment utilisés : Dublin Core Metadata, the FRBR-aligned Bibliographic Ontology (FaBiO), the Bibliographic Ontology (BibO), FOAF, Schema.org, the Web Annotation Vocabulary et PROV-O. Une description complète de la représentation RDF ainsi que des exemples sont fournis dans le [dépôt GitHub du pipeline](#).

L'ensemble du pipeline ISSA a été déployé sur l'archive institutionnelle du Cirad, Agritrop. Dans ce contexte, le thésaurus multilingue Agrovoc a été utilisé comme vocabulaire de référence spécifique au domaine.

### Services de visualisation

Afin de démontrer l'intérêt d'un tel index sémantique des publications d'une archive ouverte (objectif 3), nous avons développé plusieurs outils de visualisation à destination des utilisateurs. Bien qu'encore à l'état de prototypes, ceux-ci illustrent le type de question de compétence qu'il est possible de résoudre en s'appuyant sur les informations extraites des articles et les liens sémantiques établis entre eux.

- 1) L'outil de **visualisation des notices bibliographiques**, illustré dans la Figure 2 répond au besoin le plus classique d'une archive : accéder aux métadonnées des documents. Le prototype propose une vue

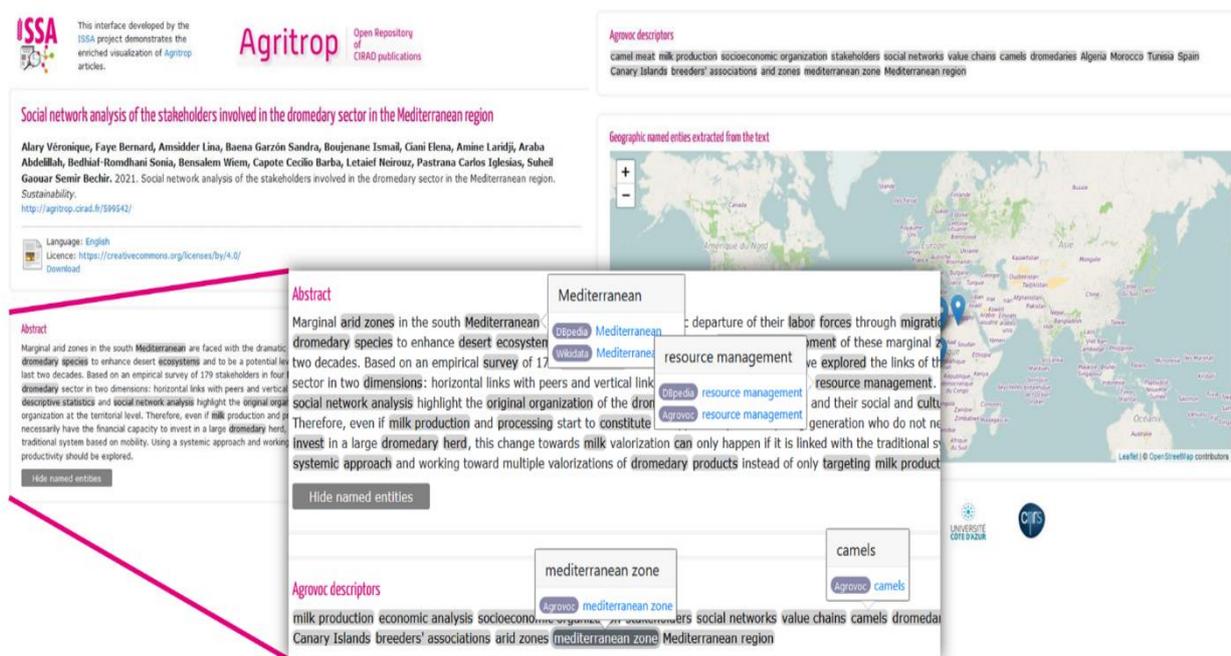


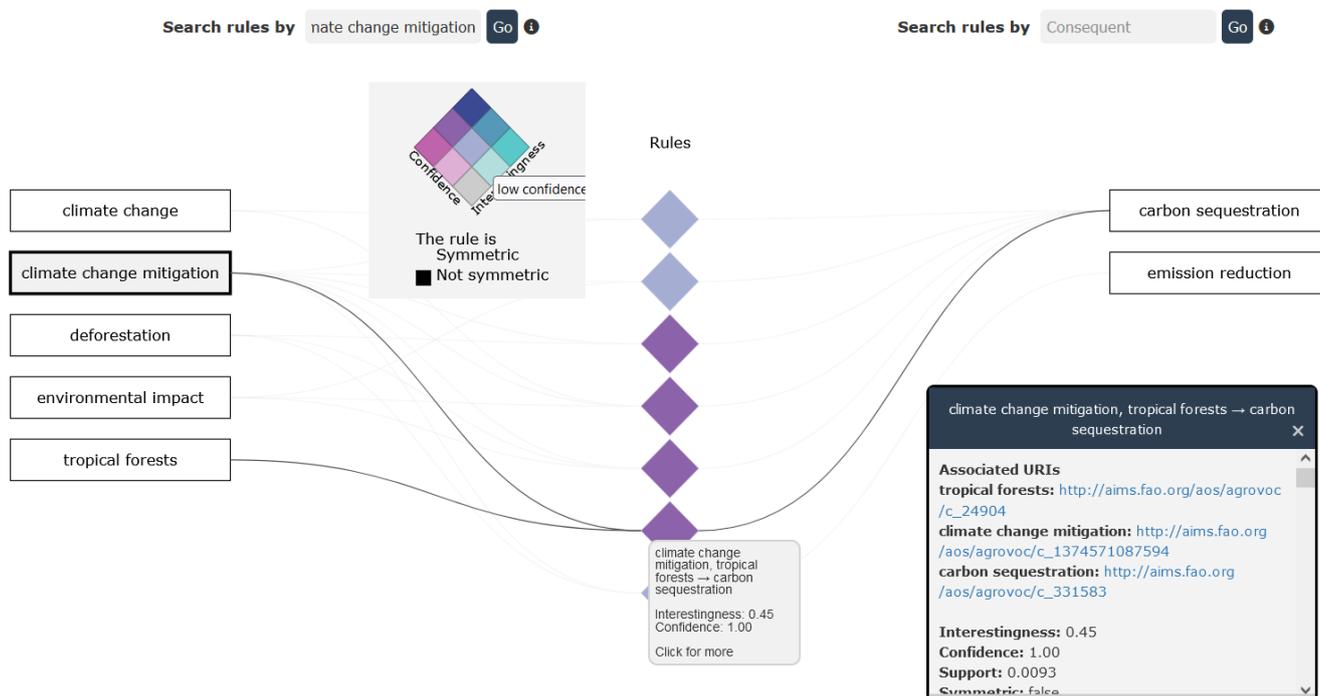
Figure 2. Visualisation enrichie d'une notice bibliographique

enrichie de ces notices en intégrant non seulement les métadonnées mais aussi le résumé de l'article dont les entités nommées sont surlignées et liées aux bases de connaissances (Wikidata, DBpedia,

**Projet ISSA Lauréat de l'appel à projet 2019-2020**  
**Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte**

GeoNames et Agrovoc dans notre cas d'usage), les descripteurs thématiques extraits automatiquement, ainsi qu'une visualisation cartographique des lieux mentionnés dans l'article.

- 2) **ARViz** est un outil générique conçu pour l'exploration de règles d'association. Une règle d'association "A → B" se comprend de la façon suivante : si un document possède le descripteur thématique A (l'antécédent), alors il possède le descripteur thématique B (le conséquent) avec une certaine probabilité. Dans la Figure 3, les concepts antécédents et conséquents sont représentés à gauche et à droite de la figure respectivement, tandis que des losanges représentent les règles d'association. Leur couleur indique l'intérêt et la confiance des règles mentionnées. Dans cet exemple, la règle survolée indique que les antécédents *climate change mitigation* et *tropical forests* ont comme conséquent *carbon sequestration* avec une confiance maximale.



**Figure 3. ARViz, visualisation et exploration de règles d'association entre descripteurs thématiques**

- 3) Enfin, l'outil **LDViz** permet de naviguer de façon exploratoire dans le graphe de connaissances et de répondre à des questions de compétence complexes. Dans l'exemple qui suit, nous nous intéressons à l'initiative « One Health » qui vise à unifier les thèmes de santé publique, animale et environnementale pour mieux comprendre le développement des pandémies et la propagation des maladies émergentes.

## Projet ISSA Lauréat de l'appel à projet 2019-2020 Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte

La Figure 4 montre comment aider les utilisateurs à rechercher des articles mentionnant le concept de santé ou l'un de ses sous-concepts (a et b), à découvrir qu'il est souvent co-mentionné avec le changement climatique (c), et à obtenir la liste des publications connexes (d) et leur répartition dans le temps (e).



Figure 4. MGExplorer, navigation exploratoire dans le graphe de connaissances

## Diffusion et valorisation des résultats

### Dissémination

Un effort important du projet a porté sur l'aspect **pédagogie/retour vers les communautés** pour une **montée en compétence des jeunes** et le transfert de la méthodologie vers d'autres établissements.

Au cours du projet, 3 stages ont été proposés à des étudiants de Master 2 en informatique et un stage à un étudiant en L3 licence professionnelle « Métiers de l'information, veille et gestion des ressources numériques ».

Deux évènements ont été organisés :

- Un **atelier**, à destination de professionnels de l'information, s'est déroulé à la Bnu de Strasbourg le 9 juin 2022. Cette journée a permis de présenter les objectifs et les résultats obtenus.
- Un **hackathon** (réunion de développeurs en informatique) s'est déroulé les 28 et 29 octobre à Sophia-Antipolis avec les étudiants de l'Université de Nice.



## Projet ISSA Lauréat de l'appel à projet 2019-2020 Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte

### Communication

Différents canaux et différentes formes de communication ont été utilisés pour diffuser les informations concernant le projet :

- La [page projet CollEx-Persée](#) qui propose une fiche synthétique du projet ainsi que des pointeurs vers des informations en lien avec le projet
- Le [site Web du projet ISSA](#) ; on y trouve les informations importantes, les annonces d'évènements, les pointeurs vers des informations en lien, ...
- Un billet publié le 3 juin 2021 « [Paroles de chercheurs](#) », qui a pour ambition de recueillir, sous la forme d'un article-témoignage, le point de vue d'acteurs importants de l'écosystème CollEx-Persée, ici les chercheurs.
- Un « #FocusProjet » publié le 27 janvier 2022 : « [ISSA, l'intelligence artificielle au service de la recherche bibliographique](#) ». Il s'agit d'un article court qui donne un éclairage sur la vie du projet à un instant donné.
- Le fil Twitter du projet : @ProjetISSA

### Code source et documentation

#### – Code et composants logiciels

Le code du pipeline ISSA et des composants logiciels sont identifiés par un DOI et disponibles sur les dépôts GitHub du projet sous licence **Apache 2.0** (open-source et libre). Le Tableau 1 précise ces éléments.

| Projet                                     | DOI  | Dépôt Github   |
|--|--|--|
| ISSA pipeline                              | 10.5281/zenodo.6513983                           | <a href="https://github.com/issa-project/issa-pipeline">https://github.com/issa-project/issa-pipeline</a>  |
| Arviz et fouille des règles d'association  | 10.5281/zenodo.6511786<br>10.5281/zenodo.6511146 | <a href="https://github.com/Wimmics/arviz">https://github.com/Wimmics/arviz</a><br><a href="https://github.com/Wimmics/association-rules-mining">https://github.com/Wimmics/association-rules-mining</a>                   |
| MGExplorer                                 | 10.5281/zenodo.6511782                           | <a href="https://github.com/Wimmics/ldviz">https://github.com/Wimmics/ldviz</a>  |
| Visualisation des notices bibliographiques | 10.5281/zenodo.6510031<br>10.5281/zenodo.6510029 | <a href="https://github.com/issa-project/web-visualization">https://github.com/issa-project/web-visualization</a><br><a href="https://github.com/issa-project/web-backend">https://github.com/issa-project/web-backend</a> |

Tableau 1. Code et composants logiciels



**Projet ISSA Lauréat de l'appel à projet 2019-2020**  
**Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte**

– **Dataset ISSA Agritrop**

Le jeu de données produit dans le cadre du cas d'usage Agritrop est disponible sous forme d'un dump RDF téléchargeable et via un point d'accès SPARQL public. Il est identifié par un DOI et placé sous licence *Open Data Commons*.

Ces informations sont résumées dans le Tableau 2.

|  |  |
|--|--|
| DOI du jeu de données ISSA Agritrop                            | 10.5281/zenodo.6505847   |
| Licence  | Open Data Commons 1.0  |
| Nombre d'articles traités (texte intégral)<br>Nombre d'auteurs | 12,000+<br>35,000+   |
| RDF dump téléchargeable  | <a href="https://doi.org/10.5281/zenodo.6505847">https://doi.org/10.5281/zenodo.6505847</a>                                      |
| Point d'accès SPARQL   | <a href="http://data-issa.cirad.fr/sparql">http://data-issa.cirad.fr/sparql</a>  |
| Entités nommées extraites                                      | Entités nommées : 9.69M (Wikidata : 1.65M, DBpedia : 1.56M, Agrovoc : 6.33M, GeoNames : 180K)<br>Descripteurs thématiques : 350K |
| Nombre de triplets RDF   | 64.8M  |

**Tableau 2. Éléments d'information sur le jeu de données produit**

– **Documentation**

La documentation publiée dans le cadre du projet ISSA est en accès ouvert sous licence *Creative Commons Attribution 4.0 International*, à l'exception des notebooks Jupyter (dernière ligne du tableau ci-dessous) sous licence Apache 2.0 comme le reste du code.

Ces informations sont rassemblées dans le Tableau 3 ci-dessous :

| Document  | DOI / Lien de téléchargement  |
|---|---|
| État de l'art des référentiels géographiques  | <a href="https://doi.org/10.5281/zenodo.7458064">https://doi.org/10.5281/zenodo.7458064</a>   |
| Liste de référentiels sémantiques pouvant servir à l'indexation d'un corpus de publications scientifiques | <a href="https://doi.org/10.5281/zenodo.7458319">https://doi.org/10.5281/zenodo.7458319</a>   |
| Analyse statistique des descripteurs thématiques prédits et des entités nommées (notebooks Jupyter)       | <a href="https://github.com/issa-project/issa-pipeline/tree/main/notebooks">https://github.com/issa-project/issa-pipeline/tree/main/notebooks</a> |

**Tableau 3. Documentation disponible**



## Projet ISSA Lauréat de l'appel à projet 2019-2020 Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte

### Publications en lien avec le projet

Les différents articles mentionnés ci-dessous sont en accès ouvert sous licence [Creative Commons Attribution 4.0 International](#) (conférences internationales [ISWC 2022](#) et [TDWG 2022](#)) ou sous licence [CC BY-ND 2.0](#) (article publié dans le numéro 107 de la revue « [Arabesques](#) »). Un article est à paraître et sera également publié en accès libre ([Conférence EGC 2023](#)).

Liste détaillée des publications :

- Anne Toulet, Franck Michel, Anna Bobasheva, Aline Menin, Sébastien Dupré, Marie-Claude Deboin, Marco Winckler, Andon Tchechmedjiev. *ISSA: Generic Pipeline, Knowledge Model and Visualization tools to Help Scientists Search and Make Sense of a Scientific Archive*. ISWC 2022 - 21st International Semantic Web Conference, Oct 2022, Hangzhou, China, [https://doi.org/10.1007/978-3-031-19433-7\\_38](https://doi.org/10.1007/978-3-031-19433-7_38).
- Franck Michel, Anne Toulet, Anna Bobasheva, Marie-Claude Deboin, Sébastien Dupré, Aline Menin, Marco Winckler, Andon Tchechmedjiev. *Semantic Indexing of Open Scientific Literature to Help Users Discover and Navigate through Publications Networks*. TDWG 2022 - The Biodiversity Information Standards annual conference, Oct 2022, Sofia, Bulgaria, <https://doi.org/10.3897/biss.6.93640>
- Anne Toulet, Franck Michel et Andon Tchechmedjiev. *Le projet ISSA : l'intelligence artificielle au service de la recherche bibliographique*. Arabesques 107, p. 6-7, 2022. <https://dx.doi.org/10.35562/arabesques.3046>
- Anne Toulet, Franck Michel, Anna Bobasheva, Aline Menin, Sébastien Dupré, Marie-Claude Deboin, Marco Winckler, Andon Tchechmedjiev. *ISSA, un graphe de connaissances au service de la recherche bibliographique*. EGC 2023 - 23ème conférence francophone sur l'extraction et la gestion des connaissances. 16-20 janvier 2023, Lyon, France (*à paraître*).