

Porteur de projet

Nom : ZUCKER

Prénom : Arnaud

Qualité : Professeur des Universités

Composante – Laboratoire : CEPAM, UMR 7264

Adresse :

Pôle Universitaire Saint-Jean d'Angély. SJA3

24, avenue des diables bleus

F-06357 Nice Cédex 4

Courriel : zucker@unice.fr

Téléphone : 04.98.88.15.14

Mobile : 06.80.30.59.54

Co-porteurs du projet

Nom : FARON-ZUCKER

Prénom : Catherine

Qualité : Maître de conférence

Composante – Laboratoire : I3S, UMR 7271

Courriel : faron@i3s.unice.fr

Nom : TETTAMANZI

Prénom : Andrea

Qualité : Professeur des Universités

Composante – Laboratoire : I3S, UMR 7271

Courriel : Andrea.TETTAMANZI@unice.fr

PRÉSENTATION DU PROJET

Libellé du projet :

Construction d'un thésaurus de zoologie antique et médiévale par fédération de données hétérogènes et annotation de textes anciens

Résumé du projet :

L'objectif du projet est la réalisation d'un thésaurus de zoologie antique et médiévale, élaboré à partir de la fédération de sources de données hétérogènes sur les connaissances zoologiques anciennes (antiquité/moyen-âge) et de l'annotation de textes latins, dans une perspective anthropologique et épistémologique, prenant en compte toutes les dimensions de ce savoir : taxinomie, anatomie, physiologie, éthologie, exploitation économique, thérapeutique.... Il s'inscrit dans un projet général porté par le CEPAM et soutenu par le CNRS à travers le GDRI *Zoomathia* (2014-2017) sur la transmission des connaissances zoologiques. Il se décline en deux volets et passe par le recrutement sur contrat post-doctoral de deux chercheurs internationaux niveau IR pour 12 mois, l'un en SHS et l'autre en STIC, appelés à travailler en synergie et co-encadrés par des binômes de chercheurs appartenant à deux UMR de l'UNS (notées A+ et A par l'AERES) : le CEPAM (INEE, INSHS), et l'I3S (INS2I). S'inscrivant dans le cadre des Humanités Numériques, ce projet permettra le renforcement de coopérations scientifiques interdisciplinaires STIC-SHS au sein de l'UNS, une plus grande visibilité internationale dans le domaine de l'histoire des sciences anciennes, et fournira un référentiel pionnier pour l'annotation de corpus scientifiques pré-modernes dans le but d'aider à l'étude de la transmission des connaissances zoologiques à travers ces textes.

Dans quel contexte votre projet s'inscrit-il ?

Les Humanités Numériques (*Digital Humanities*) ont bouleversé l'approche et l'exploitation des sources anciennes, et elles offrent aux chercheurs des possibilités nouvelles de traitement des données. Elles constituent un outil et un champ indispensables pour appréhender des savoirs complexes et diffus et des corpus textuels très étendus. C'est le cas de la littérature concernant les animaux, qui a une extension considérable dans la tradition gréco-latine aux époques antique et médiévale. Les savoirs zoologiques sont, en effet, souvent dispersés dans des textes qui ne sont pas identifiés comme zoologiques (géographiques, théologiques, historiques, encyclopédiques...) et une étude de la transmission des connaissances scientifiques (en l'occurrence zoologiques) sur un temps long exige une méthodologie renouvelée que permet désormais le développement du web sémantique. Le CEPAM (UMR 7264 de l'UNS) est porteur d'un Groupement de Recherche International (ZOOMATHIA), soutenu par le CNRS, consacré à l'étude de la transmission des savoirs dans l'Antiquité et au Moyen-Âge. Il fédère des laboratoires français de trois instituts du CNRS (INEE, INSHS, INS2I) et des universités européennes (Cornell (USA), Durham (GB), Trier (D), Siena (I) ...) et réunit des philologues, des archéologues, des historiens, des iconologues, et des spécialistes en zoologie ancienne, en ingénierie des connaissances et en web sémantique. Il a été l'occasion en particulier d'un rapprochement du CEPAM avec l'I3S (UMR 7271 de l'UNS). Le réseau de chercheurs constitué est un socle permettant le développement de projets spécifiques visant l'intégration de données hétérogènes (textuelles, iconographiques, archéologiques) dans le domaine de l'histoire de la zoologie. Certains réseaux existent déjà, mais limités au Moyen-Âge, qui se sont engagés dans l'inventaire des ressources archivistiques et

bibliographiques (en particulier *Animaliter. Les animaux dans la littérature du Moyen Age. Projet d'une encyclopédie interdisciplinaire*¹ ; *MAD. Medieval Animal Data-networks*²), mais ne proposent pas de traitement sémantique systématique des textes, dans une perspective historique et épistémologique. La numérisation des données textuelles, première étape accomplie de la révolution digitale dans le domaine des Sciences Humaines et de la philologie (la grande majorité des textes étant désormais disponible sous ce format) permet de travailler à la seconde étape, l'annotation et la structuration des données, grâce aux outils et à la méthodologie du web sémantique. En effet les recherches portant sur les sciences pré-modernes se heurtent à trois obstacles majeurs que ce projet peut aider à lever :

- Les approches ne croisent jamais systématiquement des données considérées comme relevant de champs disciplinaires et de formations académiques différents (philologie, archéologie, iconographie, biologie). Or l'intégration de données hétérogènes (en l'occurrence de nature textuelle, archéologique, biologique, iconographique) est un problème bien connu aussi dans les domaines des sciences fondamentales, notamment dans le domaine médical, où s'est développé la recherche "translationnelle", qui nécessite la médiation de sources de données variées. Les résultats de travaux en cours au laboratoire I3S sur la fédération de données scientifiques hétérogènes dans le domaine médical (projet Credible⁷ de l'appel à projet MASTODONS du CNRS) pourront être adaptés aux besoins du projet présenté de fédération de données zoologiques antiques et médiévales hétérogènes à l'aide d'une ontologie de référence.
- Les recherches actuelles dans le domaine sont restreintes aussi du fait qu'elles ne peuvent être conduites que par trois voies qui n'offrent qu'une approche (a) empirique, (b) partielle ou (c) conditionnée : la lecture individuelle des œuvres (a), la recherche automatique par termes ou lemmes *présents* dans le texte (b), la consultation de la littérature secondaire (c). Le balisage du corpus zoologique au moyen d'un thesaurus *ad hoc* (intégrant lemmes, identifiants zoologiques, catégories critiques, descripteurs historiques et génériques, etc.) permettra à terme de conduire de manière complète et extensive sur les sources des recherches multiples et de proposer une histoire de la zoologie ancienne.
- La science antique n'est pas prise en compte par les historiens de la zoologie, qui s'intéressent presque exclusivement à la science moderne (XVI-XXI), et leur approche n'est pas adaptée à la zoologie antique ou médiévale. Quant aux ontologies élaborées par les biologistes, elles sont généralement spécialisées (taxinomie⁸, anatomie⁹, biomédecine¹⁰, biologie moléculaire¹¹...) et non unifiées, et elles ne comportent en outre aucune dimension culturelle ou anthropologique. Or les études en ethnozoologie développées depuis la fin des années 60 (Berlin et al. 1968) montrent la nécessité d'envisager l'histoire de la zoologie en tenant compte des catégories de pensée et de discours des cultures étudiées. L'alignement de l'ontologie produite avec les existantes permettra une ré-articulation perdue entre la zoologie pré-moderne et la zoologie moderne.

Portée du projet, objectifs visés et retombées attendues?

Ce projet s'inscrit dans une logique de développement des Humanités Numériques et des relations et synergies entre SHS et STIC. Au niveau de la politique d'établissement de l'UNS il s'intègre à l'axe interdisciplinaire *Histoire des idées, des sciences et des arts*. Il sera l'occasion de compléter le réseau des disciplines impliquées, en visant une collaboration avec des membres du secteur de la biologie (IBV), sachant qu'un EC en sciences de la vie (MCF HDR) participe déjà aux travaux du CEPAM. Outre le rôle structurant pour

¹ <http://www.encyclopaedia-animalium.germanistik.uni-mainz.de/projet.html>

² <http://www.imareal.oeaw.ac.at/mad/>

⁷ <http://credible.i3s.unice.fr>

⁸ Voir Taxonomic rank vocabulary : http://phenoscape.org/wiki/Taxonomic_Rank_Vocabulary

⁹ Voir VSAO : <http://biportal.bioontology.org/ontologies/VSAO?p=classes>

¹⁰ Voir OBO Foundry : <http://www.obofoundry.org/>

¹¹ Voir Schulze-Kremer 1998 : <http://psb.stanford.edu/psb-online/proceedings/psb98/schulze-kremer.pdf>

la recherche à l'UNS, ce projet contribuera au développement et à la publication de données relatives à la zoologie antique et médiévale sur le web (LOD) et d'outils informatiques « intelligents » pour l'annotation et l'exploitation de la documentation du domaine.

Les deux chercheurs recrutés, dont l'un sera spécialiste de zoologie antique, et l'autre du web de données et de l'ingénierie des connaissances travailleront en étroite collaboration sur deux tâches complémentaires et simultanées :

(1) la construction d'une ontologie par l'annotation d'un corpus textuel

Le choix comme corpus initial d'un texte latin à caractère encyclopédique (Pline, *Historia naturalis*, livres 8-12) permettra la prise en compte de toutes les dimensions de ce savoir (taxinomie, anatomie, physiologie, éthologie, exploitation économique, thérapeutique...) et son utilisation pour un corpus élargi (Antiquité/Moyen Âge). L'ontologie produite intégrera à la fois le vocabulaire et les concepts originaux (dans le domaine de la classification, de la physiologie, de l'analyse des comportements, etc.) et ceux de la biologie contemporaine. Restituer la conception biologique ancienne manifestée dans l'ouvrage de Pline, à travers les descripteurs choisis, constitue un des défis de ce travail d'annotation. Ce thesaurus « expérimental » constituera le tout premier repère méthodique pour l'annotation des ressources littéraires en zoologie. Le travail consistera donc à conduire, à partir d'un corpus numérisé disponible, les deux phases fondamentales du « cycle de vie de l'information historique » défini par Boonstra et al. (2006) : enrichissement et édition (tagging), afin de permettre l'extraction de données et l'analyse textuelle du corpus annoté.

(2) la construction et l'alignement d'une ontologie issue de la fédération de données hétérogènes

Si la littérature antique à caractère zoologique est très riche, son exploitation doit être couplée à celle des autres données historiques. Un premier objectif de cette seconde tâche est l'extraction et la publication de vocabulaires dans les standards du web de données, à partir des ressources disponibles sur la zoologie antique et médiévale, dans le but de fédérer des données issues de sources de natures différentes (textes, images, objets archéologiques), dans des formats différents (bases de données relationnelles, bases de données XML). Pour cela, le chercheur reposera sur le standard R2RML qui permet de mettre en relation des données relationnelles et des données RDF, et sur une extension de R2RML sur la définition de laquelle travaille actuellement le laboratoire I3S dans le cadre du projet Credible pour associer des données XML à des données RDF. Un second objectif est d'aligner les schémas ou ontologies extraits afin de construire un thesaurus unifié de zoologie antique dans le standard SKOS du web de données, et de l'aligner avec les ontologies existantes en biologie moderne, en capturant les dimensions historiques et évolutives de ce savoir zoologique, afin de pouvoir interroger conjointement les données hétérogènes disponibles et les combiner dans un environnement unifié. Les questions de recherche posées sont donc ici les suivantes : « Comment automatiser autant que possible la construction d'un thesaurus à partir des schémas de bases de données relationnelles et des schémas XML des bases existantes hétérogènes ? », « Comment aligner les schémas produits entre eux et avec les ontologies de la biologie contemporaine, tout en explicitant et en représentant l'évolution conceptuelle du savoir zoologique ? ».

Ce thesaurus unifié permettra la valorisation du *corpus* des ressources et savoirs antiques et médiévaux à caractère zoologique. Une des caractéristiques des données zoologiques, dans les trois grands domaines visés par le projet (objets et biorestes, textes, images) est leur extrême fragmentation ; ainsi, dans le seul livre 8 de *l'Histoire Naturelle* de Pline, cet auteur signale explicitement 787 faits, histoires et observations : *res*, et

historiae et obseruationes). Un outil de traitement systématique de cette immense documentation sur des thématiques diverses (exploitation économique, domestication, communication intra- et inter-spécifiques, affectivité animale, comportements familiaux, zootechnies, etc.) permettra à terme des synthèses culturelles aujourd'hui pratiquement impossibles ou lacunaires sur le savoir zoologique et les relations hommes-animal, et révélera certainement, par le croisement des données, des problématiques non encore identifiées. Il proposera un modèle méthodologique applicable à d'autres domaines des sciences anciennes (astronomie, physique, médecine...), au-delà même de la culture gréco-latine visée par le projet (littérature arabe, persane, chinoise, indienne...). Il constituera ainsi, dans le champ des Humanités Numériques, un référentiel commun sur la thématique, et sa promotion et son évolution seront assurées grâce au réseau international des chercheurs du GDRI Zoomathia. Ce dernier, qui réunit 20 partenaires scientifiques, créé en 2014 et amené à répondre à des AAP européens en humanités numériques et SHS, pourra en effet sur cette base déployer un projet d'annotation intégral des ressources textuelles, iconographiques et matérielles relatives à la zoologie antique et médiévale, selon un mode collaboratif (plate-forme web) et à une échelle internationale, et offrir un modèle de recherche articulée Sciences humaines/Sciences du web/Sciences de la Vie. Les premiers résultats issus de ce travail de post-doctorat et de cette collaboration interdisciplinaire intra-UNS aideront également l'UNS à bien se positionner dans les réponses aux AAP en SHS.

Comment comptez-vous le mettre en œuvre et à quelle échéance ?

Le projet s'étend sur une année, avant une phase coopérative de validation et d'enrichissement du thésaurus par la communauté de spécialistes structurée par le GDRI.

En SHS : Le post-doctorant expert en philologie classique sera hébergé au CEPAM et co-encadré par deux experts en zoologie antique : Arnaud ZUCKER (CEPAM, UNS) et Catherine JACQUEMARD (CRAHAM, Caen, engagée dans le projet ICHTHYA de mise en ligne d'un corpus annoté de traités d'ichtyologie), et deux experts en ingénierie des connaissances et web de données : Andrea TETTAMANZI et Catherine FARON-ZUCKER (I3S, UNS), tous quatre partenaires du GDRI Zoomathia. Les missions du post-doctorant au CRAHAM ou de Catherine JACQUEMARD au CEPAM seront prises en charge sur le budget du GDRI. Andrea TETTAMANZI et Catherine FARON-ZUCKER fourniront au post-doctorant des outils et méthodes d'ingénierie des ontologies lui permettant de construire une ontologie de zoologie antique et médiévale. Il sera formé à l'utilisation d'un gestionnaire de thésaurus (OpenTheso¹²), conçu pour les ressources antiques, par un ingénieur du CEPAM, spécialisé dans la gestion de bases de données. Le projet ayant une portée internationale et le thésaurus étant multilingue, l'IR sera recruté par appel d'offres international et procèdera à l'annotation dans une langue européenne, en fonction de ses compétences linguistiques (français, anglais, allemand, italien).

En STIC : Le post-doctorant spécialiste du web de données sera hébergé à l'I3S et co-encadré par deux experts du domaine : Catherine FARON-ZUCKER (I3S, UNS), et Johan MONTAGNAT (I3S, CNRS), et deux experts en zoologie antique : Arnaud ZUCKER (CEPAM, UNS) et Isabelle DRAELANTS (IRHT, PARIS), tous quatre partenaires du GDRI Zoomathia. Isabelle DRAELANTS a notamment dirigé le projet ANR Sourcencyme, visant l'annotation de textes encyclopédiques médiévaux. Les missions du post-doctorant à l'IRHT ou d'Isabelle DRAELANTS à l'I3S pourront être prises en charge sur le budget du GDRI. Arnaud ZUCKER et Isabelle DRAELANTS apporteront leur expertise en zoologie antique pour expliciter les besoins des utilisateurs du thésaurus et valider celui-ci.

¹² <http://frantiq.mom.fr/thesaurus-pactols/logiciels-opentheso>

A travers leur participation aux activités du GDRI Zoomathia et leur co-encadrement interdisciplinaire STIC-SHS, les deux post-doctorants pourront se former efficacement au domaine des humanités numériques. Le programme de travail des deux chercheurs, appelés à interagir régulièrement par des réunions de travail se déclinera de la façon suivante :

Post-doctorant en SHS

- Formation personnelle à un gestionnaire de thésaurus par un IE du CEPAM, et à un outil de « tagging » intuitif fourni par les encadrants de l'I3S
- Annotation des livres 8-10 de Pline, *Historia Naturalis*
- Structuration du vocabulaire des annotations en une ontologie
- Documentation de l'ontologie (définitions, labels, ...)
- Alignement de l'ontologie produite avec les ontologies de la biologie moderne, liage au web de données
- Présentation de l'ontologie produite, de son expressivité et de son évolutivité
- Formalisation d'un guide méthodologique pour l'utilisation du thésaurus à l'annotation d'autres textes de zoologie antique ou médiévale

Post-doctorant en STIC

- Définition d'un ensemble de sources de données à traiter et analyse des schémas de ces données
- Extraction de schémas RDFS ou SKOS à partir de schémas de bases de données relationnelles existantes
- Extraction de schémas RDFS ou SKOS à partir de schémas XSD de bases de données XML
- Documentation des schémas extraits
- Alignement des schémas extraits entre eux et avec des vocabulaires existants sur le web de données
- Représentation de la dimension historique des concepts en zoologie : coexistence et mise en relation des concepts antiques et modernes en zoologie
- Fédération des sources de données hétérogènes à l'aide de l'ontologie produite

Budget global prévisionnel détaillé du projet et soutien sollicité auprès de la Fondation UNICE ?

	Rémunération forfaitaire brute annuelle	Coût mensuel avec charges et PPE et AT	Coût TTC	
1 salaire de post-doc	30.000 € (2.500 X 12)	47.221 € (3 935,14 x 12)		soutien sollicité auprès de la Fondation UNICE
1 salaire de post-doc	30.000 € (2.500 X 12)	47.221 € (3 935,14 x 12)		soutien sollicité auprès de la Fondation UNICE
2 postes de travail complets (ordinateur portable et écran)			4.000 € (2 x 2.000 €)	soutien sollicité auprès de la Fondation UNICE
4 missions (UNS/IRHT/CRAHAM)			2.000 € 4 x 500 €	Financement sur budget GDRI
BUDGET GLOBAL			100.442 €	

NB : Le budget global doit intégrer la participation en soutien à la recherche de trois personnels (IE), statutaires des laboratoires porteurs de la demande pour l'accompagnement des tâches, ainsi que l'implication des quatre encadrants de l'UNS.

Vous pouvez joindre à cette fiche de candidature tout document que vous estimerez utile ou nécessaire à l'appui de votre projet.

Annexe. Bibliographie

1. BARBARA, S. et TRINQUIER, J. (2012), *Ophiaca. Diffusion et réception des savoirs antiques sur les Ophidiens*. Paris, *Anthropozoologica*. 47/1.
2. BAUDUIN, P., JACQUEMARD, C. (2011), « Les pratiques de l'édition en ligne. Expériences et questionnement », in Genet Jean-Philippe et Zorzi Andrea, *Les historiens et l'informatique. Un métier à réinventer*, Rome, École Française de Rome (Collection de l'École Française de Rome, 444), p. 87-108.
3. BERLIN, B., BREEDLOVE, D., & RAVEN, P. (1968), « Covert Categories and Folk Taxonomies ». *American Anthropologist*, 70(2), p. 290-299.
4. BODSON, L. (1993), *L'histoire de la connaissance du comportement animal*. Colloques d'histoire des connaissances zoologiques, 13, Liège 1992, Université de Liège.
5. BODSON, L. (2010), "Les connaissances zoologiques de l'Antiquité grecque et romaine: aperçu de leurs spécificités fondamentales et de leur actualité." *Bull. De l'Association Guillaume Budé*, 1, p. 53-82.
6. BOONSTRA Onno, BREURE Leen et DOORN Peter (2006), « Past, present and future of historical information science », NIWI-KNAW (first edition 2004) : <http://www.ahc.ac.uk/docs/pastpresentfuture.pdf>
7. BOUFFARTIGUE, J. (2003), "Problématiques de l'animal dans l'Antiquité Grecque" *Lalies*, 23, p. 131-168.
8. EUZENAT, J., SHVAIKO, P. (2013), *Ontology Matching*, Springer Verlag.
9. GAIGNARD, A., MONTAGNAT, J., FARON-ZUCKER, C., CORBY, O., « Semantic Federation of Distributed Neurodata », in Proc. of the MICCAI Workshop on Data and Compute-Intensive Clinical and Translational Imaging Applications (DCICTIA-MICCAI 2012), pages 41-50, Nice, France, October 2012.
10. GANDON, F., FARON-ZUCKER, C., CORBY, O. (2012), *Le web sémantique - Comment lier les données et les schémas sur le web ?*, Dunod.
11. JACQUEMARD, C. (2012), « Éditer un texte et sa « bibliothèque » : les projets Ichtya et Bibliothèque virtuelle du Mont-Saint-Michel », Journée d'étude Ménéstrel et le CRAHAM, *Autour des études médiévales en ligne*, Université de Caen le 28 septembre 2012, <http://www.menestrel.fr/IMG/pdf/ich...>
12. KALOF, L. (2009), *A cultural history of animals in antiquity*, Berg Publishers.
13. MEROÑO-PEÑUELA, A., ASHKPOUR, A., VAN ERP, M., MANDEMAKERS, K., BREURE, L., SCHARNHORST, A., SCHLOBACH, S., VAN HARMELEN, F., *Semantic Technologies for Historical Research: A Survey*, submitted to Semantic Web Journal, <http://www.semantic-web-journal.net/system/files/swj588.pdf>
14. MICHEL, F., MONTAGNAT, J., FARON-ZUCKER, J., *A survey of RDB to RDF translation approaches and tools*, Research Report I3S/RR 2013-04-FR, 25 pages, Sophia Antipolis, France, November 2013
15. SCHULZE-KREMER 1998, « Ontologies for Molecular Biology », *Pac Symp Biocomput.* 1998, p. 695-706 : <http://psb.stanford.edu/psb-online/proceedings/psb98/schulze-kremer.pdf>
16. TANG, J., LEUNG, H., LUO, Q., CHEN, D., GONG, J: *Towards Ontology Learning from Folksonomies*. IJCAI 2009: 2089-2094 <http://keg.cs.tsinghua.edu.cn/jietang/publications/IJCAI09-Tang-et-al-Ontology-Learning.pdf>
17. ZUCKER, A. (2013). « Zoologie et philologie dans les grands traités ichtyologiques renaissants », *Kentron*, 29, 2013, p. 135-174.
18. ZUCKER, A. (2005), *Les classes zoologiques en Grèce ancienne*, Aix-en-Provence, Presses Universitaires de Provence.