



HAL
open science

Retour d'expérience VIVO Web

Windpouire Esther Dzale Yeumo

► **To cite this version:**

Windpouire Esther Dzale Yeumo. Retour d'expérience VIVO Web. Séminaire CNRS, Jan 2016, Paris, France. hal-02792937

HAL Id: hal-02792937

<https://hal.inrae.fr/hal-02792937>

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



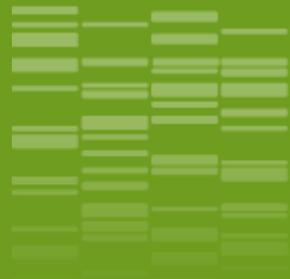
Retour d'expérience VIVO

Linked Open Publications (LOP)



SOMMAIRE

- ❖ **Présentation générale de VIVO**
- ❖ **Présentation générale de LOP**
- ❖ **Méthode et moyens mobilisés**
- ❖ **Résultats obtenus**
- ❖ **Difficultés rencontrées**
- ❖ **Conclusions**



_01

Présentation générale de VIVO



VIVO : historique¹

- ✓ 2003-2005 First realization for the life sciences at Cornell, as a relational database
- ✓ 2006-2008 Expansion to all disciplines at Cornell and conversion to Semantic Web
- ✓ 2009-2012 National Institutes of Health-sponsored VIVO: Enabling the National Networking of Scientists project transforms VIVO to a multi-institutional open source platform
- ✓ 2013-2014 VIVO incubator project with DuraSpace for open community development
 - <https://wiki.duraspace.org/display/VIVO/Introduction+to+VIVO+2014+Workshop>
- ✓ Une communauté anglophone active et réactive
 - Plusieurs listes de diffusion : <https://wiki.duraspace.org/display/VIVO/Email+Lists>



VIVO : définition et objectifs

Plateforme web sémantique et open-source de publication et de découverte des chercheurs et de leurs travaux de recherche

- ✓ Connecter : VIVO permet de lier des données provenant de différentes sources
- ✓ Partager : VIVO fournit un standard de représentation des produits de la recherche basé sur un vocabulaire commun, VIVO ISF
- ✓ Découvrir : VIVO permet de découvrir qui fait quoi avec qui dans le monde de la recherche

VIVO est une application J2EE, intégrant SolR et une base de données RDF, et utilisant un langage de template (Freemarker) pour la création des écrans

L'ontologie de VIVO

L'ontologie de VIVO permet de représenter

- ✓ Les acteurs de la recherche (personnes, organisations, groupes), les produits de la recherche, les activités (projets, thématiques), les moyens de la recherche (équipements, financements), les événements, etc.
- ✓ Les relations entre les différents concepts et objets ci-dessus cités
 - Typées de manière intelligible et
 - bidirectionnelles
- ✓ Des liens vers des identifiants globaux
 - ORCID, DOI, vocabulaires, autres identifiants pour les personnes, organisations, événements, etc.

L'intégration des données dans VIVO

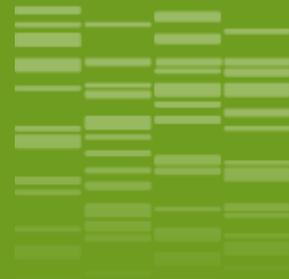
- ✓ VIVO Harvester : VIVO Harvester est une librairie d'outils pour lire et transformer des données à partir de sources externes. Voir <https://wiki.duraspace.org/display/VIVO/VIVO+Harvester>
- ✓ Karma Web pour l'intégration de données tabulaires. Voir <https://wiki.duraspace.org/display/VIVO/Using+the+Karma+data+integration+tool>
- ✓ Développements ad hoc

L'interface web de VIVO

- ✓ Un menu paramétrable basé sur les classes de l'ontologie de VIVO
- ✓ Des templates de page customisables, basées sur l'ontologie de VIVO
 - Une page de description pour chaque concept ou objet identifié par un URI
- ✓ Des fonctions de recherche et de navigation qui exploitent les relations entre les concepts, et les objets
- ✓ Des fonctions de représentation graphique
 - Map of science
 - Réseau de co-auteurs
 - Graphes temporels
- ✓ Une interface d'administrattion

Pour aller plus loin

- ✓ Portail officiel (liens vers les déploiements VIVO)
 - <http://vivoweb.org>
- ✓ WIKI et documentation
 - <https://wiki.duraspace.org/display/VIVO/VIVO>
- ✓ Introduction
 - <https://wiki.duraspace.org/display/VIVO/Introduction+to+VIVO+2014+Workshop>



_02

Présentation générale de LOP

Un projet expérimental avec des objectifs ambitieux

- ✓ Publier le contenu de Prodnra, l'archive institutionnelle de l'Inra en "Linked data" (ensemble de bonnes pratiques pour publier et connecter des données structurées sur le Web)
- ✓ Enrichir ce contenu avec d'une part, des liens vers des sources internes (référentiels des personnes, des structures, des activités) et externes diverses (sources bibliographiques, données de la recherche), d'autre part des annotations sémantiques basées sur des ressources telles que Agrovoc
- ✓ Valoriser ce contenu enrichi et mettre en évidence les apports du Web sémantique dans la recherche et l'analyse d'information, via une interface web
- ✓ Promouvoir et susciter la réutilisation de nos données
- ✓ Tester VIVO

Les principaux services envisagés

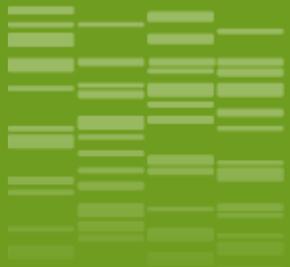
- ✓ Mettre à disposition des fonctions d'interrogation, de navigation et de contextualisation des données de LOP à travers une interface conviviale
 - Recherche par mots clés
 - Recherche étendue : les termes recherchés par l'utilisateur génèrent une équation sur les synonymes et les termes équivalents dans d'autres langues de ces termes. Cette équation étendue est affichée et modifiable par l'utilisateur qui peut supprimer certains termes complémentaires proposés.

- ✓ Offrir une fiche d'information sur les auteurs INRA contenant:
 - La liste des publications de l'auteur par ordre chronologique inverse sur le modèle d'un CV
 - Des facettes : thématiques, types de publications, coauteurs, années de publications, conférences, revues
 - Des réseaux d'auteurs, des nuages de mots, des articles citants

- ✓ Mettre à disposition un sparql endpoint exposant le contenu de l'entrepôt de données

Un peu de rêve...

- ✓ Répondre à des questions telles que « Combien de personnes à l'Inra travaillent sur le blé? Qui sont-elles? Quelles sont leurs spécialités/expertises? Sur quels projets travaillent-elles aujourd'hui? Qu'est-ce qu'elles publient? Comment les contacter? »
- ✓ Pour chaque espèce apparaissant dans les publications ou les thématiques, proposer une notice avec une illustration Dbpedia
- ✓ Une carte géographique de répartition des espèces à partir des données géographiques annotées dans la base ou à partir d'une carte préexistante extérieure



_03

Méthode et moyens mobilisés

Les chantiers du projet

Customisation de l'ontologie

Customisation de l'interface VIVO

Chargement des données dans VIVO

Rapprochement vers des référentiels externes

Transformation au format RDF

Récupération des données
sources

Annotation sémantique
des articles

Itérations successives



L'équipe

Conception de l'ontologie de LOP

- Professionnels IST avec une expertise en Web sémantique
- Experte Prodinra

Définition des mappings

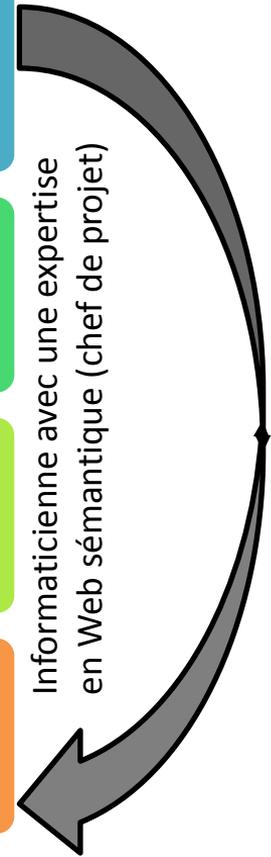
- Professionnels IST avec une expertise en Web sémantique
- Experte Prodinra

Scripts de récupération & transformation

- Prestataire externe

Customisation de l'interface

- Stagiaire informaticienne



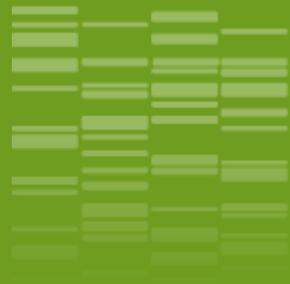
La conception de l'ontologie de LOP

- ✓ Basée sur l'ontologie de VIVO
- ✓ L'ontologie de VIVO réutilise plusieurs vocabulaires dont :
 - ✓ bibo: <<http://purl.org/ontology/bibo/>>
 - ✓ event: <<http://purl.org/NET/c4dm/event.owl#>>
 - ✓ foaf: <<http://xmlns.com/foaf/0.1/>>
 - ✓ skos: <<http://www.w3.org/2004/02/skos/core#>>
 - ✓ event : <<http://purl.org/NET/c4dm/event.owl#>>
- ✓ Autres ontologies réutilisées pour définir les concepts et propriétés de LOP
 - ✓ geopolitic fao : <<http://aims.fao.org/aos/geopolitical.owl>>
 - ✓ Dublin Core terms: <http://purl.org/dc/terms/>
 - ✓ rdf : <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>
 - ✓ rdfs : <<http://www.w3.org/2000/01/rdf-schema#>>



Transformation des données : travail itératif au long cours

- ✓ Recours à une prestation externe pour écrire les scripts de transformation
- ✓ Nous avons fourni au prestataire
 - Une description détaillée des différentes sources de données
 - L'ontologie de LOP
 - Le mapping entre les schémas des sources vers l'ontologie de LOP



_04

Résultats

Les outputs du projet

- ✓ L'ontologie de LOP
- ✓ Une base de données RDF contenant les données provenant des différentes sources internes (prodinra, référentiels des personnes, des activités, des structures) liées entre elles
- ✓ Alignement des mots clés Inra avec Agrovoc; le résultat en RDF vient enrichir la base de données RDF
- ✓ Un prototype VIVO restituant les données



Les exemples dans les captures d'écran qui suivent reflètent le contenu de notre base de données VIVO au moment où ces captures ont été faites.

Welcome to LOP

Le projet Linked Open ProdlInra (LOP ou LODP) vise à publier dans le Web des données une partie du contenu de ProdlInra, l'archive institutionnelle de l'Inra. Ce projet de publication s'appuie sur les principes des "Linked data" qui sont un ensemble de bonnes pratiques pour publier et connecter des données structurées sur le Web.

Browse or search information on people, departments, courses, grants, and grants.

Rechercher sur LOP

limit search -

Search

Log in

Email

Password

Log in



Search results for 'fatty acid'

[acide gras](#) | Concept

... Composé chimique ou facteur milieu Validated **acide gras fatty acid** 10000055390 The effect of temperature from flowering to maturity on seed composition of high ...

[acide gras volatil](#) | Concept

... Composé chimique ou facteur milieu agv Validated volatile **fatty acid acide gras volatil** 10000055430 Changes in gibbs free energy and stoichiometry of VFA ...

[acide gras mono insaturé](#) | Concept

... Composé chimique ou facteur milieu Validated **acide gras mono insaturé monounsaturated fatty acid** 10000055410 Concept ...

[acide gras essentiel](#) | Concept

[Not the results you expected?](#)

Display Only	
Research	(273)
Products	(203)
Organizations	(8)
Events	(5)

subject area of

[Trienoic fatty acids are required to maintain chloroplast function at low temperatures](#) → Article

[A QTL on pig chromosome 4 affects fatty acid metabolism: Evidence from an Iberian by Landrace intercross](#)

[Alimentation Humaine](#) → Département

[Biochemical characteristics of goat milk lipids and lipolytic system. A comparison with cow and human milk. Effects of lipid supplementation](#)

[Bioconversion of ricinoleic acid into gamma-decalactone : optimisation of the production using the yeasts *Sporidiobolus salmonicolor* and *Sporidiobolus ruinenii*](#)

... more

narrower concept

[acetic acid](#)

[butanoic acid](#)

[capric acid](#)

selected publications

abstract

[Fungal species resources: an integrated system for structural and functional genome annotation.](#) 81–81. 2012-01-01

[Analyse bioinformatique des données de métagénomique fongique.](#) 23–23. 2012-01-01

[GnpAnnot community annotation system: features, qualifiers, values](#) 2009-01-01

academic article

[Efficient comparison of sets of intervals with NC-lists.](#) *Bioinformatics*. 29:933–939. 2013-01-01

[GnpIS: an information system to integrate genetic and genomic data from plants and fungi.](#) *Database : the journal of biological databases and curation*. 2013:1–9. 2013-01-01

[The Capsella rubella genome and the genomic consequences of rapid mating system evolution.](#) *Nature genetics*. 45:831–835. 2013-01-01

[The wheat powdery mildew genome shows the unique evolution of an obligate biotroph.](#) *Nature genetics*. 45:1092–1098. 2013-01-01

[International Congress on Transposable Elements \(ICTE\) 2012 in Saint Malo and the sea of TE stories.](#) *Mobile DNA*. 3. 2012-01-01

more...

Profile

QUESNEVILLE, Hadi

[VIVO profile](#)

25 Publication(s)

32 Co-author(s)

2007 First Publication

2013 Last Publication

Note: This information is based solely on publications that have been loaded into the VIVO system. This may only be a small sample of the person's total work.

Log in to enter additional details about your publications on your profile page.

Interact

Hover over any name to see the number of joint publications and co-authors with QUESNEVILLE, Hadi.

Thresholding

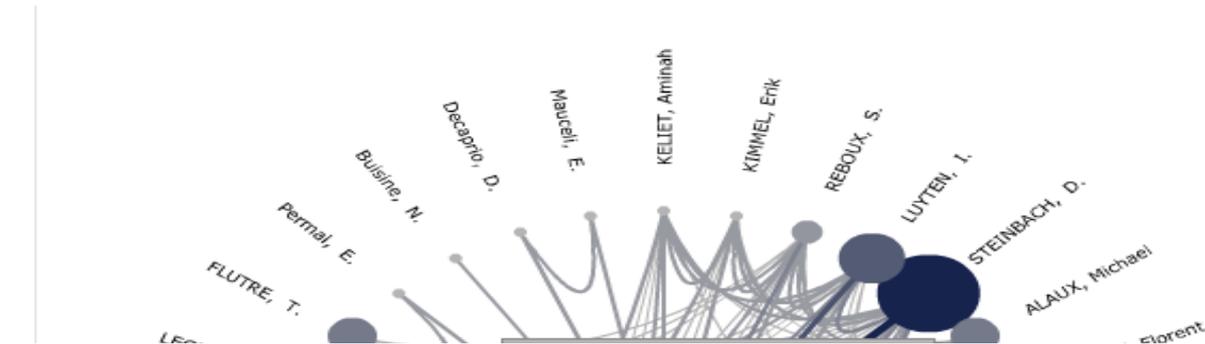
Only people that co-authored more than 1 paper(s) with QUESNEVILLE, Hadi are shown. 32 out of 442 co-author(s) are shown.

Leg

No. o



Sorted into communities: Co-authors are placed near one another's co-authors in the graph.



25 publications

from 2004 to 2013

[\(.CSV File\)](#)



442 co-authors

from 2004 to 2013

[\(.CSV File\)](#)

Tables

Publications per year [\(.CSV File\)](#)

[File](#)

Year	Publications
------	--------------

2007	4
------	---

2008	5
------	---

2009	1
------	---

Co-authors [\(.CSV File\)](#)

Author	Publications with QUESNEVILLE, Hadi
--------	-------------------------------------

STEINBACH, Delphine	7
---------------------	---

AMSELEM, Joelle	7
-----------------	---

Sciences Sociales, Agriculture et Alimentation, Espace et Environnement.

Explore activity (1,439 publications) across 554 scientific subdisciplines

Explore Sciences Sociales, Agriculture et Alimentation, Espace et Environnement. [i](#)

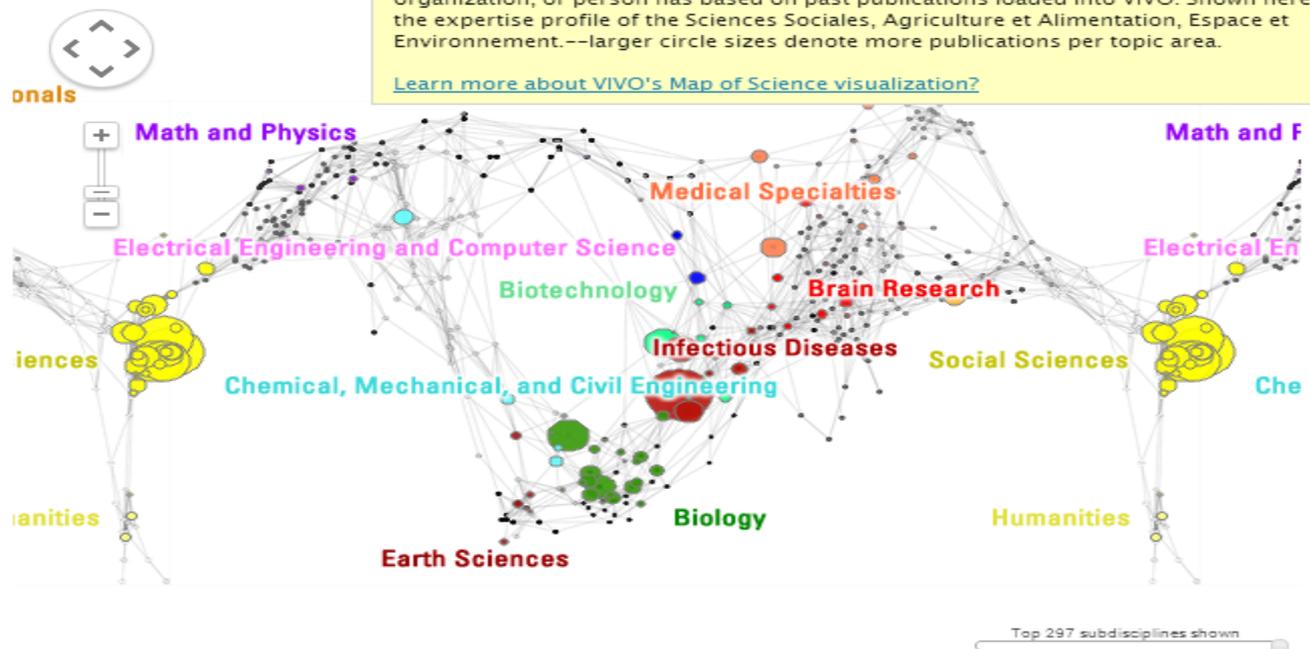
Compare organizations [i](#)

554 Subdisciplines | [13 Disciplines](#) [i](#)

Search: X [i](#)

1 - 13 of 554 * [First](#) [Prev](#) [Next](#) [Last](#) >

Subdisciplines	# of pubs.	% of activity
Animal Science	159.0	11.1
Agricultural Economics	146.0	10.1
Econometrics	122.0	8.5
Environmental Management	104.0	7.2
Economics	96.0	6.7
Weed Management	62.2	4.3
Food Engineering	48.8	3.4
Dairy Science	39.2	2.7
Plant Ecology	33.0	2.3
Poultry Science	30.1	2.1
Applied Economics	30.0	2.1
Sociology	30.0	2.1
Regional Studies	28.0	1.9



Google

© 2008 The Regents of the University of California and SciTech Strategies. Map updated by [SciTech Strategies](#), [OST](#), and [CNS](#) in 2011.

[Terms of Use](#)

mapped 21.44% of 6,713 publications [i](#)

[Save Unmapped Publications](#)

Unité Mixte de Recherche sur les Herbivores

Explore activity (245 publications) across 554 scientific subdisciplines 

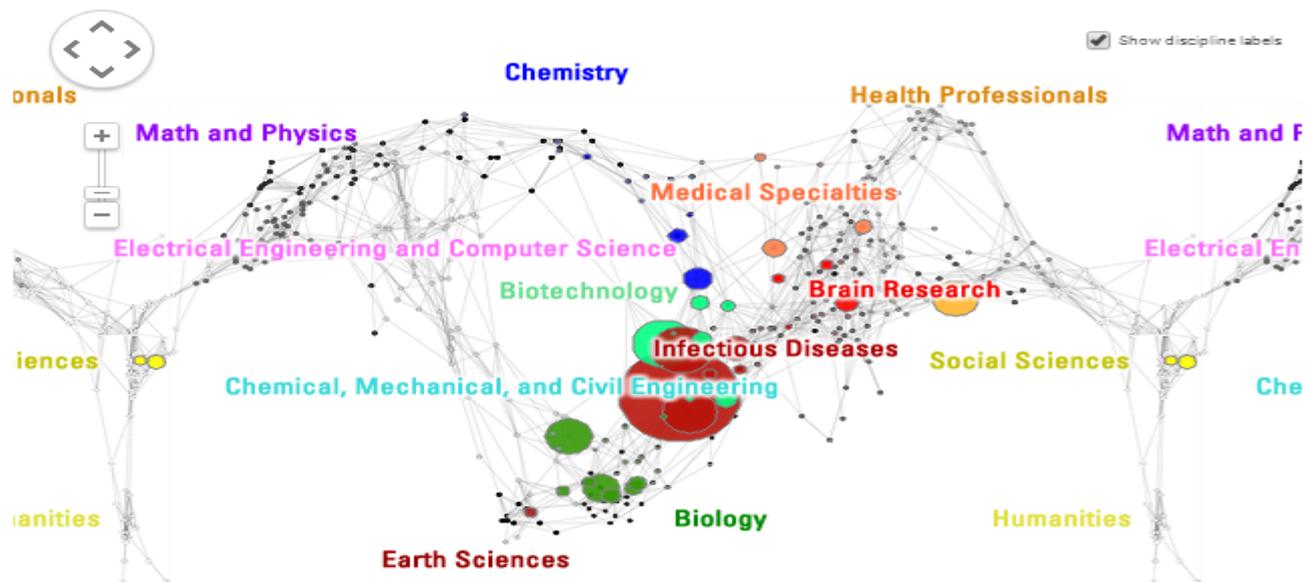
- Explore Unité Mixte de Recherche sur les Herbivores 
- Compare organizations 

554 Subdisciplines | [13 Disciplines](#) 

Search:  

1 - 13 of 564 * [First](#) [Prev](#) [Next](#) [Last](#) *

Subdisciplines	# of pubs.	% of activity
Animal Science	91.0	37.1
Food Engineering	23.5	9.6
Dairy Science	23.2	9.5
Poultry Science	18.1	7.4
Weed Management	13.1	5.4
Psychopharmacology	13.0	5.3
Plant Ecology	8.0	3.3
Reproduction Veterinary	6.0	2.4
Food Chemistry	5.2	2.1
Nutrition	3.2	1.3
Clinical Rehabilitation	3.1	1.3
Genomics & Nucleic Acids	3.0	1.2
Food Protection	2.2	0.9



Show discipline labels

Google

© 2008 The Regents of the University of California and SciTech Strategies.
Map updated by [SciTech Strategies](#), [OST](#), and [CNS](#) in 2011.

[Terms of Use](#)

mapped 33.93% of 722 publications 

[Save Unmapped Publications](#)

Sciences Sociales, Agriculture et Alimentation, Espace et Environnement.

How do you want to compare?

by Publications ▾ 

What do you want to compare?

Organizations | [People](#) 

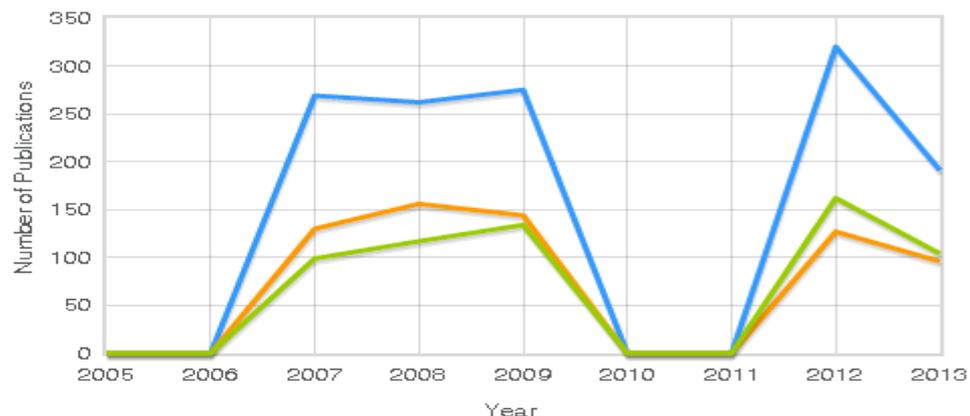
Search: X

Records 1 - 10 of 24

« [First](#) | [Prev](#) | [Next](#) | [Last](#) »

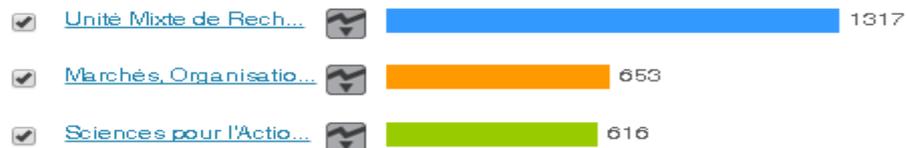
	Entity Name	PublicationEntity Type	Count
<input checked="" type="checkbox"/>	Unité Mixte de Recherche sur les Herbivores	Unité INRA, Unité de Recherche	1317
<input checked="" type="checkbox"/>	Marchés, Organisations, Institutions et Stratégies d'Acteurs	Unité INRA, Unité de Recherche	653
<input checked="" type="checkbox"/>	Sciences pour l'Action et le Développement : Activités, Produits, Territoires	Unité INRA, Unité de Recherche	616
<input type="checkbox"/>	Structures et Marchés Agricoles, Ressources et	Unité INRA, Unité de Recherche	442

Comparing Publications of Organizations & People in Sciences Sociales, Agriculture et Alimentation, Espace et Environnement.

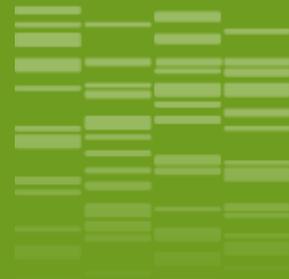


Total Number of Publications 

You have selected 3 of a maximum 10 organizations & people. [Clear](#)



Legend



_06

Limites et difficultés rencontrées

Difficultés liées à VIVO

- ✓ Problèmes de performance
 - Temps de chargement et d'indexation des données trop longs
 - Délai d'affichage des pages pour de gros volumes de données
 - Délai de génération des représentations graphiques

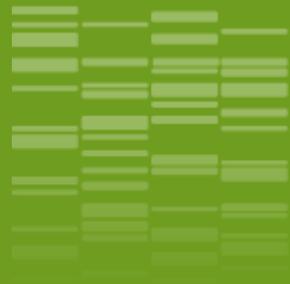
- ✓ Créer /modifier des templates pour inclure des résultats de requêtes sparql à la volée ralentit considérablement le fonctionnement de l'application

- ✓ Bien comprendre le lien entre le modèle de VIVO et l'interface : quelle propriété est utilisée où et pourquoi ?
 - Exemple : On n'a pas les mêmes informations pour toutes les personnes Inra co-auteurs d'une publication donnée. Pour un des auteurs, on a bien une courbe des publications sur 10 ans, un réseau de co-auteurs et une map of science, pas pour le second.
 - Après vérification, le problème provient du fait que le second auteur est uniquement `inraStaff` et pas `foaf:Person`. VIVO se base sur la classe `foaf:Person` pour générer le réseau de co-auteurs et la map of science tandis que `inraStaff` est une classe spécifique à l'ontologie de LOP que VIVO ne connaît pas



Quelques difficultés liées aux données sources

- ✓ Comment traiter les co-auteurs non INRA ?
 - Identification?
 - Désambiguïsation/dédoublonnage?
- ✓ Auteurs INRA absents du référentiel des personnes (auteurs ayant quitté l'INRA)
- ✓ Structures INRA présentes dans les articles mais absentes du référentiel des structures (anciennes structures)
- ✓ Publications d'une structure INRA : cas des auteurs INRA qui ont changé de structure et qui ont publié dans leur ancienne structure
- ✓ Hétérogénéité des niveaux de description des données pouvant impacter la pertinence des facettes
- ✓ Actualité et exhaustivité de certaines données : activités, revues
- ✓ Volume des données



_06

Conclusions



Bilan 1/3

- ✓ Publier le contenu de ProdInra, l'archive institutionnelle de l'Inra en "Linked data"
- ✓ Enrichir ce contenu avec d'une part, des liens vers des sources internes (référentiels des personnes, des structures, des activités) ~~et des sources externes diverses (sources bibliographiques, données de la recherche)~~, d'autre part des annotations sémantiques basées sur des ressources telles que Agrovoc
- ✗ Valoriser ce contenu enrichi et mettre en évidence les apports du Web sémantique dans la recherche et l'analyse d'information, via une interface web
- ✗ Promouvoir et susciter la réutilisation de nos données.
- ✓ Tester l'application VIVO

Bilan 2/3

- ✓ Mettre à disposition des fonctions d'interrogation, de navigation et de contextualisation des données de LOP à travers une interface conviviale
 - Recherche par mots clés
 - Recherche étendue : les termes recherchés par l'utilisateur génèrent une équation sur les synonymes et les termes équivalents dans d'autres langues de ces termes.
~~Cette équation étendue est affichée et modifiable par l'utilisateur qui peut supprimer certains termes complémentaires proposés.~~

- ✓ Offrir une fiche d'information sur les auteurs INRA contenant:
 - La liste des publications de l'auteur ~~par ordre chronologique inverse sur le modèle d'un CV~~
 - ~~Des facettes : thématiques, types de publications, coauteurs, années de publications, conférences, revues~~
 - Des réseaux d'auteurs, ~~des nuages de mots, des articles citants~~

- ✗ Mettre à disposition un sparql endpoint exposant le contenu de l'entrepôt de données

Bilan 3/3

- ✘ Répondre à des questions telles que « Combien de personnes à l'Inra travaillent sur le blé? Qui sont-elles? Quelles sont leurs spécialités/expertises? Sur quels projets travaillent-ils aujourd'hui? Qu'est-ce qu'elles publient? Comment les contacter? »
- ~~✓ Pour chaque espèce apparaissant dans les publications ou les thématiques, proposer une notice avec une illustration Dbpedia~~
- ~~✓ Une carte géographique de répartition des espèces à partir des données géographiques annotées dans la base ou à partir d'une carte préexistante extérieure~~

Leçons tirées de l'expérimentation

1. VIVO : une application pour lier, partager et découvrir les données SUR la recherche et non pas les données de la recherche
2. Se poser les bonnes questions dès le début
 - À quelles questions souhaite-t-on répondre?
 - Comment/par qui les données seront mises à jour?
 - En tant qu'outil de découverte des chercheurs et de leurs activités/expertises, quelle antériorité des données (publications, projets) est pertinente?
3. Si les fonctionnalités attendues sont assez éloignées de celles que propose VIVO de base, il peut valoir le coup de considérer une autre approche telle que :
 - Réutiliser l'ontologie de VIVO pour la représentation des données et développer l'interface Web

VIVO 1.8.1 Improved Performance and New Visualizations

- ✓ Performance improvements
 - Users should see up to 75% reduction in time to display profiles compared to VIVO 1.8, and a 30% reduction compared to VIVO 1.7. These findings have been observed on both small data sets (200 people, 3,500 articles), and large data sets (4,500 people, 40,000 articles). Most profiles - even large/complex profiles - display within approximately two seconds.
- ✓ Visualizations
 - The ability to examine the VIVO network from many points-of-view is at the heart of connecting researchers, ideas and resources. VIVO 1.8.1 offers users new and improved visualizations that make it easier to see formative research as it emerges at individual, institutional and topic levels.
- ✓ New Javascript based versions of Co-Author and Co-Investigator networks are available. These new visualizations do not require Flash and display on mobile devices such as phones and tablets.
- ✓ All visualizations have significant performance improvements. Maps of Science, Temporal Graphs and Co-author and Co-Investigator networks now all complete in just a few seconds. The very largest Maps of Science may require up to two minutes to complete.
- ✓ Additional Improvements
 - New AltMetric badges are enabled on publications by default, providing direct link to AltMetric information regarding the publication.
 - Additional improvements reduce time and resource usage for indexing and inferencing. More than a dozen have been made to improve user experience.