



Persée et le web de données



data.persee.fr

Viviane Boulétreau

SemWeb.Pro 2016



MINISTÈRE
DE L'ÉDUCATION NATIONALE,
DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE



- Pourquoi un triplestore ?
- Comment est-il construit ?
- Des données aux données liées...
- Exhaustivité vs confiance

- À propos de Persée
 - Bibliothèque numérique
 - Accès libre et gratuit
 - Documents scientifiques
 - Collections exhaustives
 - Public scientifique

- Bibliothèque → Corpus de recherche
 - Croisement de requêtes
 - Extraction d'information
 - « Outillage »

- Spécificité des données
 - Caractère historique
 - Production

➤ Quelles données ?

➤ Bibliothèque

→ Collection / Organisation

➤ Contexte de publication

→ Historique

→ Scientifique

➤ Contexte de production

→ Historique

→ Scientifique

➤ Description bibliographique « standard »

➤ Quels modèles ?

- RDA : œuvre / expression / manifestation...
- FOAF+RIO : personnes et liens
- BIBO : nature des documents
- CITO : liens entre documents
- DCMI : description classique
- SKOS : concepts
- PERSEE : typologie fine des documents

- Exploitabilité du TS
 - À quel besoin doit-on répondre ?
 - Quels sont les éléments pertinents
 - Comment y accéder
 - Comment les exploiter

- Idéation / Évaluation

- Schéma de données simplifié
 - Allègement (RDA)
 - Ontologie Persée

- Requêtes
 - Sparklis (Sébastien Ferré, LIS, IRISA)
 - Interrogation en langage naturel
 - Réorganisation des classes
 - Tutoriels

- Visualisation
 - RelFinder, LodLive, Svcizler... (*en cours*)

YASGUI view
Tutoriels
Modèles
Exemples

Your query and its **current focus** permalink

give me every **person**
 that has a **name**
 and that is the **focus** of a **document**
 that is a **bibliographic note**
 and that has a **bibliographic citation**
 and that is the **aut** of a **document** whose **aut** is
 a **person**
 and a **person** that has a **name**
 and for **each focus x**
 give me the list of **aut's name**

Sparklis suggestions to refine your query

The current focus is **the focus** (click on different parts of the query to change it)

matches all of

and that has a **death date** 7

and that has a **family name** 7

and that is the **focus** of ... 7

and that has a **name** 7

and that has a **given name** 7

and that has a **label** 7

and that has a **page** 7

22 concepts

matches all of OK

Person 7

Person 7

Person 7

Person 7

Person 7

Person 7

7 entities

matches all of

for each **1st document's aut**

for each **1st name**

for each **2nd document**

for each **bibliographic note**

give me a **sample of 1st name**

give me a **sample of bibliographic note**

give me a **sample of bibliographic citation**

29 modifiers

Results of your query

7 results Show 10 results

focus	list of name
1 Person	A. Kojevnikoff / Adolphe Lods / Charles Virolleaud / Claude Schaeffer-Forrer / Edmond Pottier / Georges Dumézil / Henri Jeanmaire / Juliette Renaud / M. O. Janse / Olov Janse / P. Kraus / Pierre Montet / René Dussaud / S. Elisséev / Victor Goloubew / Victor Leroquais / Édouard Dhorme
2 Person	Bruno Bréart / Claude Violet / Ernest Will / François Vasselle / Jean Louis Cadoux / Jean-Claude Blanchet / O. Chadwick / Pierre Bouchain / Roger Agache
3 Person	André Parrot / Andrée Herdner / Bernard Bavant / Daniel Schlumberger / Dominique Orssaud / Ernest Will / François Vasselle / Gabriel Saadé / Georges Tate / Henri Seyrig / Jacqueline Gachet / Jean PERROT / Jean Starcky / Jean-Luc Biscop / Jean-Pierre Sodini / Marguerite Yon-Calvet / Pierre Demargne / René Dussaud / Roger Agache / Roger Hanoune / Roland Martin R / Swantje Bavant

- Gestion des autorités personne
 - Alignement avec IdRef
 - Traitement manuel
 - Intégré au processus de production (mars 2016)
 - Traitement automatique
 - Réalisé par l'ABES, Qualinca
 - Taux actuel : 19%
 - Autres référentiels
 - BNF, VIAF, ISNI, Dbpedia, RePEc,...

- **Corpus ATHAR**
 - InVisu (INHA) / Persée, BSN5
 - Module d'alignement intégré à la production
 - Etablissement du référentiel (SKOS)

- Diffusion :
 - Cayro Gazeteer
 - Collection en ligne

- Référentiel en biologie
 - Alignement initial : GBIF
 - Forme normalisé (actuelle) du nom d'espèce
 - Autres alignements déduits
 - EOL, BHL, wikiSpecies, Dbpedia

- Diffusion (dvpt en cours)
 - Intégration au TS
 - Pages composites sur le portail
 - Interopérabilité

- Alignements massifs
 - Risque de bruit

- Alignements contrôlés
 - Qualité des données
 - Confiance dans la source

- Ouverture : début 2017
 - Nouvelle phase d'évaluation
 - Tests & retours bienvenus

- Interactions
 - Appariement des données
 - Nouveaux partenariats

- <http://data.persee.fr>
- data@persee.fr