



Le TDM par l'exemple : *des microbes dans mon fromage !*



Claire Nédellec

La loi numérique, et après ?

Colloque de Meudon, 10 nov. 2016

OntoBiotope, une application TDM pour la microbiologie

Mieux connaître les microorganismes et l'adaptation à leur milieu : où, pourquoi, comment ?
Un enjeu pour la recherche et l'industrie : santé, agro-alimentaire, environnement.

Des développements majeurs des technologies moléculaires et des sciences de l'information

- ➔ Déluge de données expérimentales, modélisation des mécanismes.
- ➔ Déluge de documents en texte libre, articles, brevets, bases de données.

Des infrastructures bioinformatiques

pour l'analyse intégrée de l'information, multi-sources, multi-espèces, multi-échelle.

OntoBiotope

Extraire des textes toute l'information sur les habitats des microorganismes et la **normaliser** pour l'analyser avec l'information expérimentale.



© Myriam Louvet-Paoli 2015

OntoBiotope Database

Welcome to the Ontobiotope database. You can browse through bacteria and their habitats found in over 700,000 PubMed abstracts.

You can start exploring the data either by bacteria taxonomy, or by the bacteria habitat ontology.

Psychrobacter aquimaris

Search habitat

Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Moraxellaceae

Psychrobacter Psychrobacter aquimaris

bacteria habitat

Psychrobacter aquimaris

marine environment

Toute l'information de PubMed,
analysée et accessible en ligne

- 1,16 millions de documents
- 3,63 millions de relations
bactérie - habitat, extraites et
normalisées

Bacteria		
facet value	freq.	doc.
Psychrobacter aqu	16	11
Pseudorhodobacte	8	4
Photobacterium pi	8	4
Psychrobacter	14	3

Psychrobacter aquimaris sp. nov. and **Psychrobacter namhaensis** sp. nov., isolated from **sea water of the South Sea** in Korea.

2005 *International journal of systematic and evolutionary microbiology*

Abstract Two Gram-negative, non-motile, non-spore-forming, slightly halophilic bacteria isolated from **sea water of the South Sea** in Korea, and were characterized taxonomically by means of 16S rDNA sequencing.

Psychrobacter aquimaris
(taxon) (10)

Psychrobacter aquimaris

Jamais identifiée dans des aliments

Très présente dans les milieux marins

Explication : elle est venue avec le sel ajouté

également trouvé ... Thermus thermophilus

vit dans les milieux naturels chauds (*hot spring, geothermal*, ...)

Explication : provient des échangeurs thermiques à 60°C.

OntoBiotope : une application TDM pour les microbiologistes

- **Extrait** précisément les informations sur les habitats des microorganismes et les lie entre elles.
- Associe l'information à des **catégories hiérarchiques**, taxon, habitat, phénotype, ... pour la "standardiser" à différents niveaux de généralité.
- **Intègre et analyse** les informations sur la plateforme bioinformatique IFB-Migale pour détecter des anomalies et produire automatiquement des hypothèses.





Intérêts et limites d'OntoBiotope

Comment

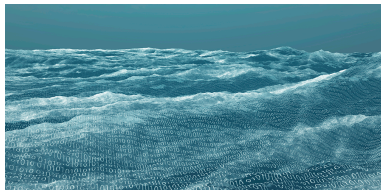


Repose sur un processus automatique de TDM d'une grande complexité, basé sur des outils innovants d'intelligence artificielle

- Prédiction par apprentissage automatique
- Analyse sémantique profonde par traitement automatique de la langue

En cours

openMIN7ED



Use Case du projet européen H2020 OpenMinTeD, infrastructure européenne de TDM

Enjeu majeur pour l'Europe : mettre à disposition de tous les chercheurs, une plateforme libre de TDM, utilisable sans compétence en TDM.

Plus généralement

Leadership français en TDM pour les SdV, mais frein majeur : accès au texte complet



Source bibliographique d'OntoBiotope

Importance de l'exhaustivité

Il faut **toute l'information** sur les milieux de vie des microorganismes pour des analyses transversales (phylogénétiques, écosystèmes, métagénomiques)

Textes complets : 6 fois plus d'habitats bactériens dans les articles que dans les références.

Dans IJSEM (*International Journal of systematic and evolutionnary microbiology*)

32 par article mais seulement 5,3 par résumé.

Collection : **1 156 824** références d'articles (PubMed)

Composition

- **8 670** journaux
- Issus de **30 plateformes**, principalement SpringerLink, Wiley Online Library, HighWire Press et ScienceDirect (10 couvrent la plus grande part)
- **Les abonnements INRA et OA** permettent l'accès à **48% des articles** utiles (551 817 articles), soient 1 098 journaux (**1/8 des journaux**).

Des limites juridiques et techniques au TDM

Juridique

Licences inutilisables pour le TDM, parce que non analysables automatiquement et elles dépendent de l'année, de la rubrique et du souscripteur.



Pas de partage des collections.

Vers une **restriction à OA** ?



Ex. l'infrastructure



Absurde du point de vue de l'utilisateur du TDM !

A noter : seuls 5% des journaux d'OntoBiotope ont une **clause TDM** connue (13% des articles)

Technique

- **Pas de centralisation** : beaucoup de sociétés savantes importantes ont leur propre plateforme.
Annual Reviews, Endocrine Society, The Microbiology Society
- Sur certaines plateformes (*HighWire Press*), chaque revue a son serveur
Oxford Journal = 59 revues = 165 694 articles d'OntoBiotope
- Utilisation du **DOI et de CrossRef** : beaucoup d'erreurs, page d'atterrissage souvent incorrecte
- Impossibilité technique de **sauvegarder automatiquement** (Ex : Science Direct. 15% des articles)



