



Le projet européen H2020 OpenMintED a débuté en juin 2015 avec l'objectif de mettre en place une e-infrastructure, encourageant et facilitant l'utilisation des technologies de fouille de textes (*text and data mining* : TDM).

Le lancement officiel de cette plateforme fin mai 2018 à Bruxelles

OMTDay - Journée OpenMinTed

4 juin 2018



L'objectif de l'évènement « OMTDay » est de vous présenter les résultats du projet et les solutions qu'il apporte aux grandes questions liées au *text-mining*, telles que l'accès à la littérature scientifique, l'extraction d'information fine, la prise en main des outils TDM et leur interopérabilité avec d'autres applications d'analyse de l'information

Programme de la journée

Programme de la matinée

09 h 30 – 09 h 45	Accueil et café
09 h 45 – 10 h 00	Introduction de la journée Hervé Monod (dépt MIA, Inra)
10 h 00 – 10 h 10	Présentation du programme et des objectifs de la journée Sophie Aubin (DIST, Inra)
10 h 10 – 10 h 30	Introduction au projet OpenMinTeD Claire Nédellec (MaIAGE, Inra)
10 h 30 – 11 h 30	Présentation de cas d'usage <ul style="list-style-type: none"> <input type="checkbox"/> Le text-mining pour la microbiologie des aliments Hélène Falentin (STLO, Inra) <input type="checkbox"/> Le text-mining pour les phénotypes du blé Michaël Alaux (URGI, Inra)
PAUSE	
11 h 40 – 12 h 10	Principes et outils du text-mining Arnaud Ferré (MaIAGE Inra) 
12 h 10 – 12 h 30	Répondre à des besoins diversifiés Sophie Aubin (DIST, Inra)
Déjeuner au restaurant du personnel	

Programme de l'après-midi

14 h 00 – 14 h 20	Une vision Open Science Odile Hologne (DIST, Inra)
14 h 20 – 14 h 50	Exploiter la plateforme logicielle OpenMinTeD Robert Bossy (MaIAGE, Inra)
14 h 50 – 15 h 20	Des acteurs et des ressources Claire Nédellec (MaIAGE, Inra)
15 h 20 – 16 h 20	Table ronde : contours d'une e-infrastructure de TDM française <ul style="list-style-type: none"> <input type="checkbox"/> Béatrice Daille (LINA, Université de Nantes) <input type="checkbox"/> Juliette Dibie (MIA-Paris, Inra) <input type="checkbox"/> Laurence El Khoury (DIST, CNRS) <input type="checkbox"/> Valentin Loux (MaIAGE, Inra) <input type="checkbox"/> Laurent Schmitt (INIST, CNRS) <input type="checkbox"/> Pierre Zweigenbaum (LIMSI, CNRS)
16 h 20 – 16 h 30	Clôture de la journée

Les présentations seront disponibles sur le site de la journée
<https://journées.inra.fr/openminted/>

Plus de 60 millions d'articles, 2.5 par an
160 millions de documents scientifiques
indexés

*Orduña-Malea
et al. , 2014*

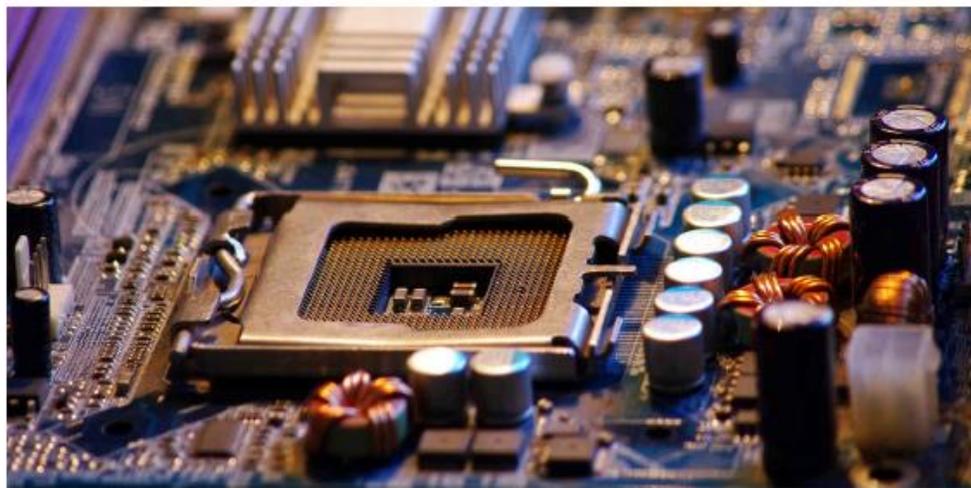
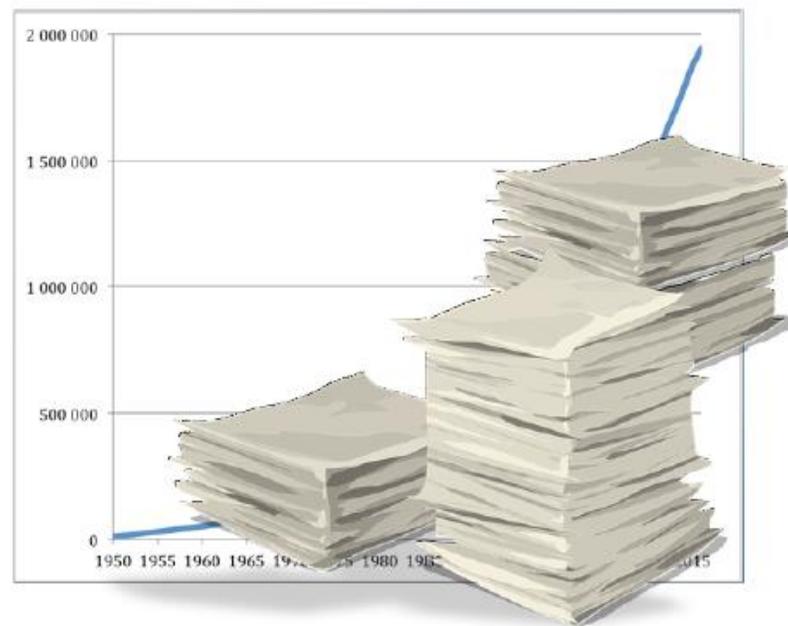
50% des articles ne sont jamais lus
90% des articles ne sont pas cités

80% des articles cités ne sont pas lus

*The STM
report, 2015*

*Lokman I. Meho, the
rise and rise of citation
analysis, 2007.*

*Simkin &
Roychowdhury. Read
before you cite!*



Au cœur du TDM

Donner du sens aux données textuelles

Transformer une donnée non structurée
en donnée structurée, manipulable par
un ordinateur

Intégrer le TDM scientifique
au cœur de l'activité du chercheur non
spécialiste

Accéder

Analyser les sources textuelles, les organiser, les classer, les indexer

... et comprendre

extraire les entités et les relations entre elles, sélectionner les faits saillants
conceptualiser les informations

... et prédire

modélisation à partir des données structurées

Textes d'articles

Bifidobacterium longum subsp. longum is found in newborn infant as a normal component of gut flora.

Extraction
d'information

Bacteria	<i>Bifidobacterium longum</i>	
	<i>subsp. longum</i>	[taxid: <u>1679</u>]
hosted by	<i>newborn infant</i>	[<u>baby</u>]
lives_in	<i>gut</i>	[<u>intestine</u>]

Représentation structurée de l'information

OpenMinTeD : quelle proposition pour des services de TDM ?



Concevoir le support d'une activité d'ingénierie et de service

Nourrie par la recherche en TDM

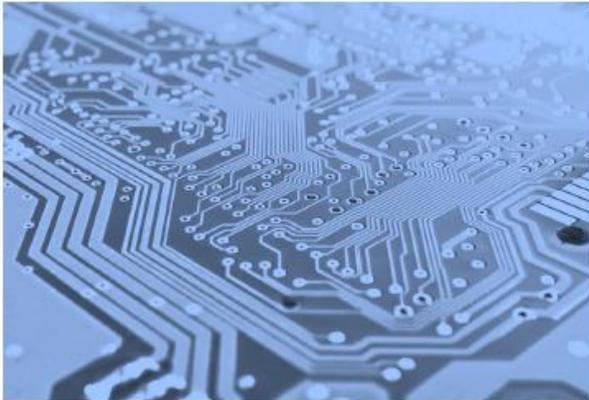
Guidée par les besoins

Besoins

Capitalisation des solutions TDM

- Pour une utilisation par des non-spécialistes du TDM
- Moyen de calcul pour des traitements à grande échelle

Besoin grandissants de traitements TDM spécifiques à haute valeur ajoutée



Réponse sous la forme de plateforme offrant des moyens

- de facilitation de la conception et de la maintenance de services TDM
- de connection aux bibliothèques numériques et aux ressources sémantiques pour une offre adaptée
- d'intégration dans des applications clientes intégrant textes et données

Rôle positif des microorganismes dans les aliments

Fermentation

Cas d'usage #1

- Phénotypes**
- Protéolyse
 - Utilisation des sucres
 - Lipolyse
 - Arômes
 - Pigments
 - Vitamines
 - Biopréservation
 - Prévention de l'altération



Les descriptions de phénotypes sont dispersées sous forme de texte, dans les articles scientifiques, les catalogues de collections microbiennes, les bases de données génétiques

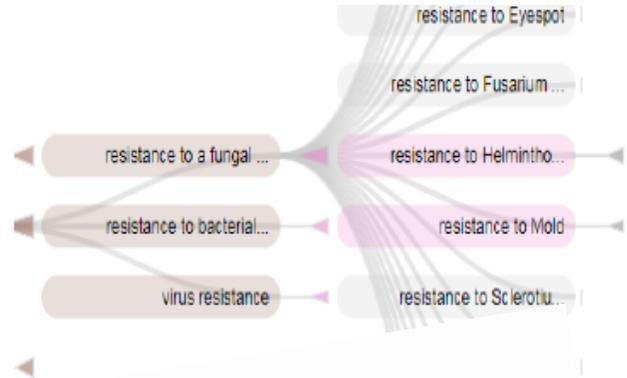


Objectif : rassembler et organiser les informations dans une base de données relationnelle

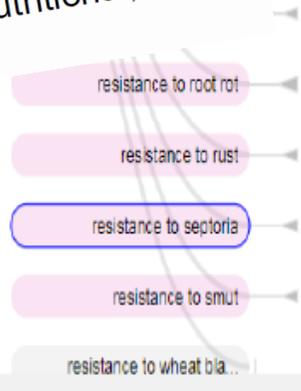
Alignment example

Cas
d'usage
#2

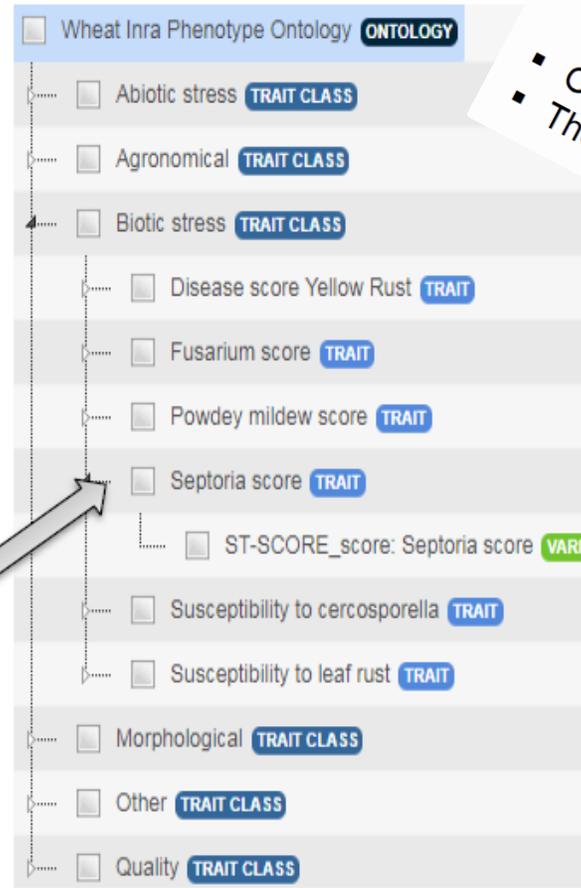
WTO



WTO:
▪ Wheat Trait Ontology
▪ General traits found in the literature:
traits of resistance, development, nutritional, baking
quality, etc.



WIPO



WIPO:
▪ The Wheat INRA Phenotypes Ontology is developed
in the frame of the BreedWheat project.
▪ Observations variables
▪ The WIPO v1.3 contains 262 variables.

Q1: Quels sont les bénéfices que vous espérez tirer d'une e-infrastructure comme OpenMinTeD telle qu'elle est aujourd'hui ou à l'avenir ?

Q2 : Quel rôle seriez-vous prêt(e) à jouer et **comment le paysage actuel devrait évoluer** pour que vous puissiez jouer ce rôle pleinement ?

Q3 : **Quelles seraient les étapes de la mise en œuvre d'une telle e-infrastructure TDM au niveau français**, des points de vue technologiques, organisationnels, de la gouvernance, etc.

