

# openMIN7ED

Open Mining Infrastructure for Text & Data



Claire Nédellec, Inra

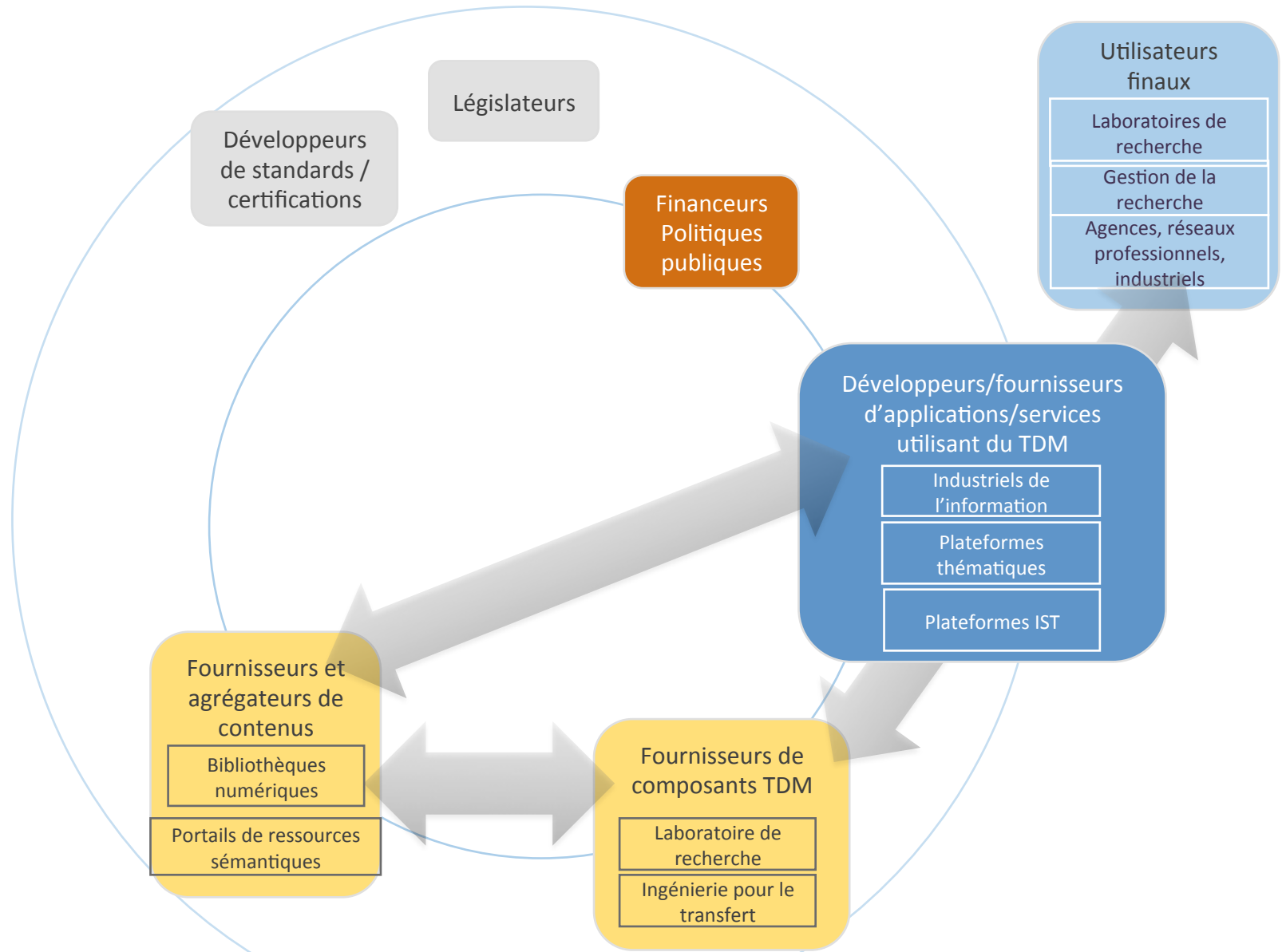
## Fouille de Texte et Science Ouverte

Ouverture des données  
ouverture des logiciels  
ouverture des services

JNSO 2018  
MESRI, Paris  
6 décembre 2018



# Aujourd'hui



# Demain

- Quelles conditions de production de **services TDM**

- A haute valeur ajoutée basés sur **l'analyse sémantique**

A destination des **chercheurs**

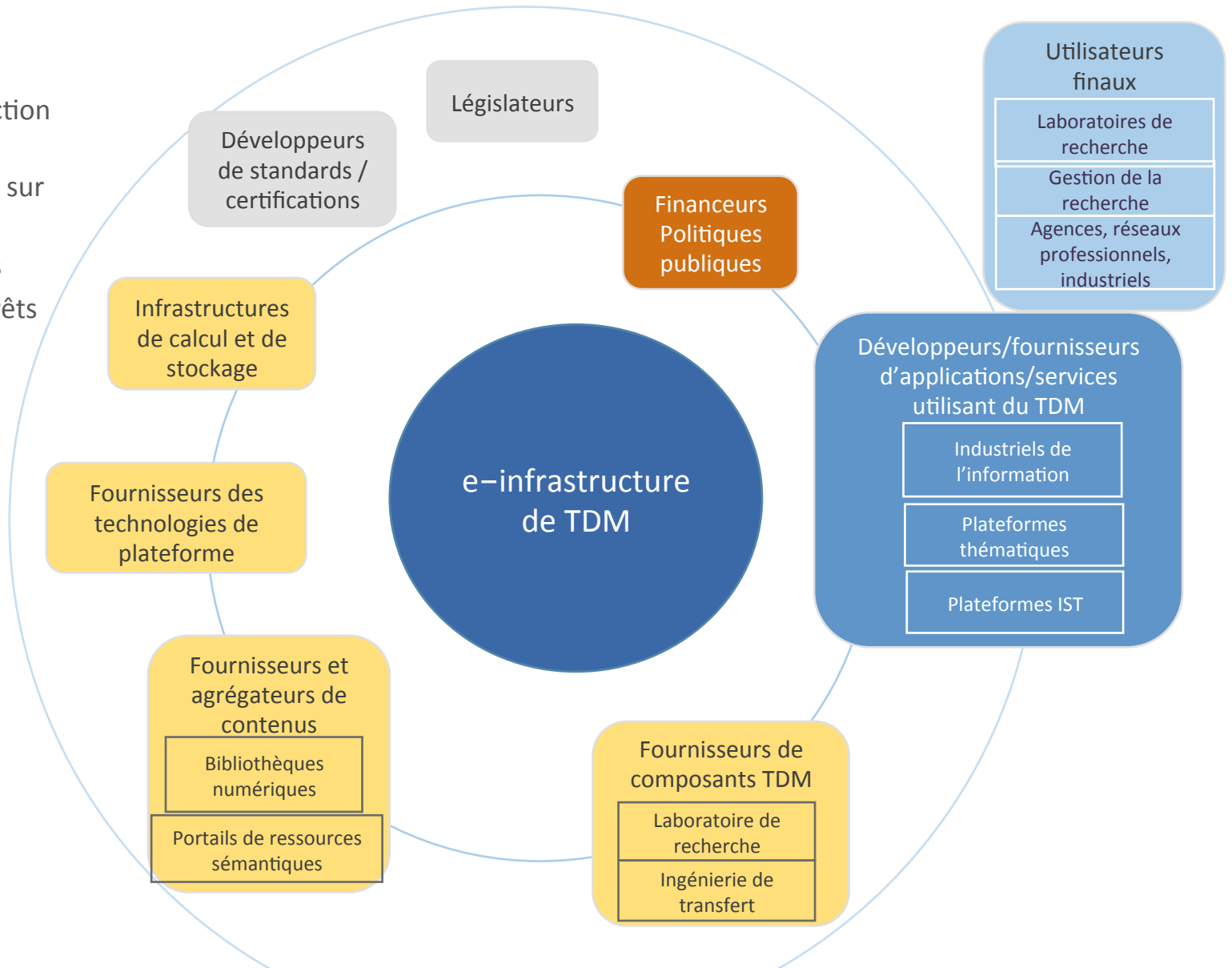
Mettant en **synergie** les intérêts et les complémentarités

## Objectifs des projets

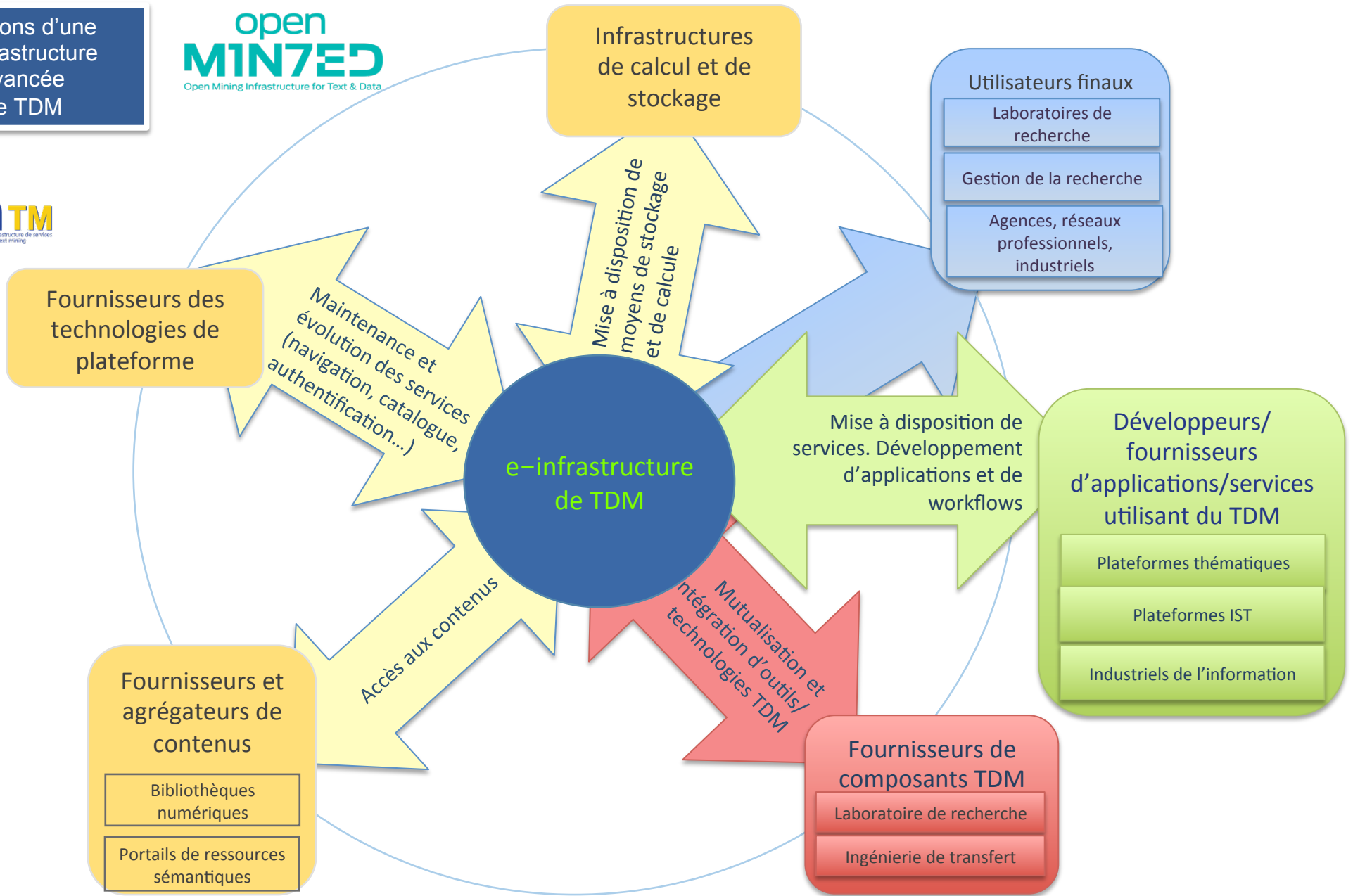


- mutualiser
- rationaliser
- reproduire
- partager
- accompagner

- Proposer des technologies  
Et des organisations



Missions d'une e-infrastructure avancée de TDM



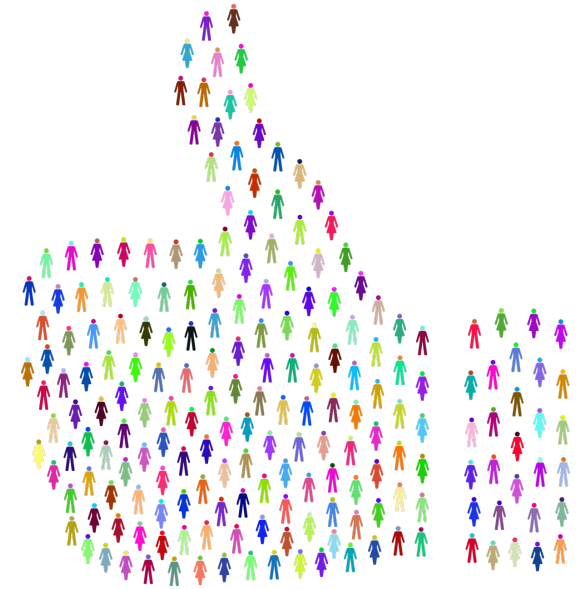


# Missions d'un dispositif, d'une e-infrastructure avancée de TDM

Faciliter l'appropriation et l'utilisation des technologies de TDM par les chercheurs et les acteurs de l'appui à la recherche

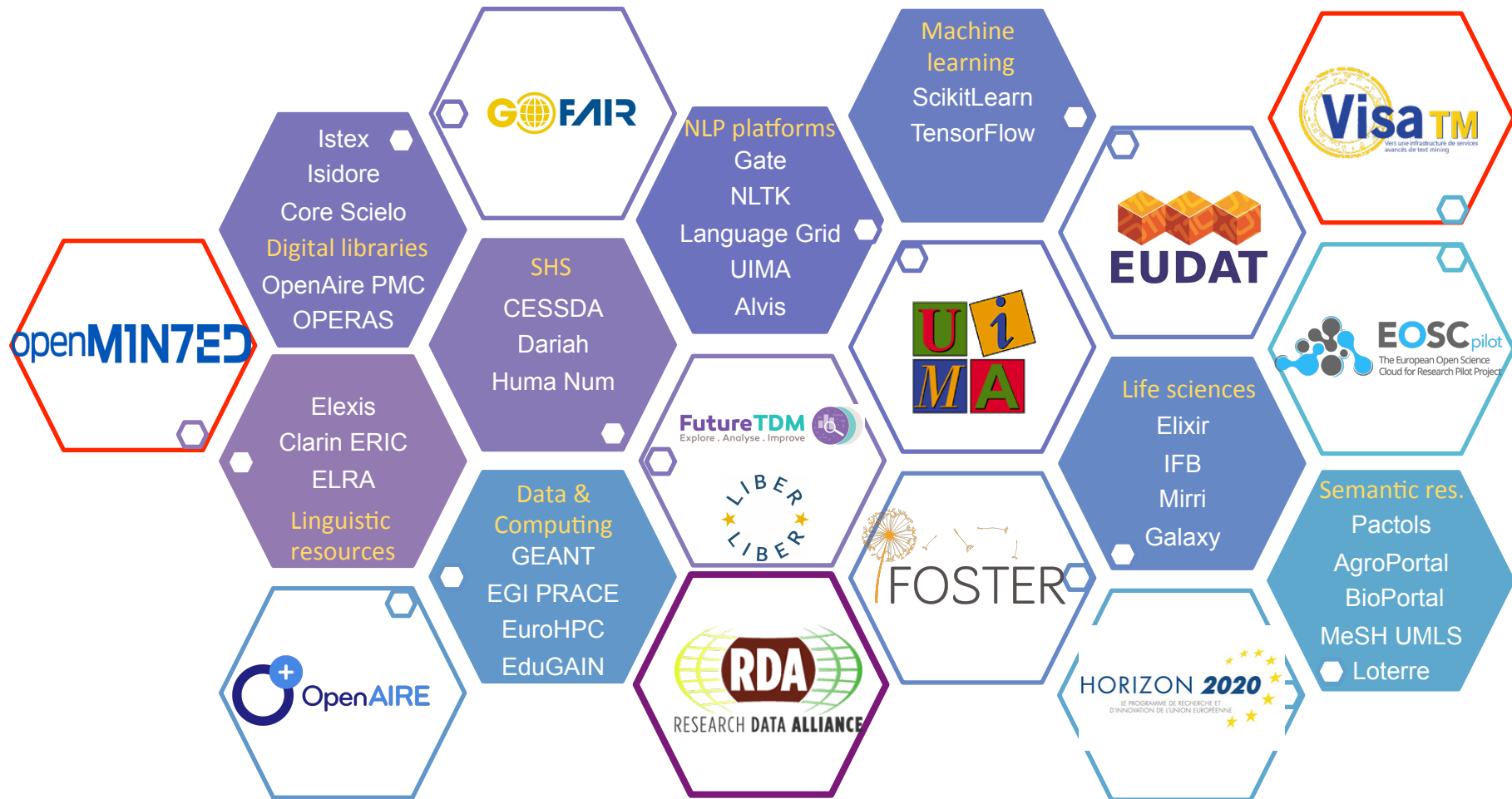
## Expertise, accompagnement, animation

- Au travers de son réseau de compétences, elle assure le **lien entre la plateforme technique et ses utilisateurs**
- Elle facilite son **appropriation** et s'assure qu'elle **répond aux besoins** des différentes communautés : utilisateurs finaux et intermédiaires, fournisseurs de composants et de contenu.
- Elle contribue à développer **une communauté de pratique** dans le domaine de l'utilisation de la fouille de texte.
- Elle promeut l'harmonisation des standards, représentations et pratiques, dans les **principes FAIR** auprès des différentes communautés concernées (données, publications scientifiques, ontologies, outils logiciels).



# L'e-infrastructure cible de *text mining*

## Composant d'un écosystème international





# E-infrastructure européenne de TDM : OpenMinTeD H2020 (2015-2018)

environnements de TDM

16 partenaires : Droit - Evaluation - Communautés science de la vie, agriculture, biodiversité, SHS, IST. Editeur OA – Formation & communication. Recherche en TDM, apprentissage – *cloud computing*



<h2>Contenu accessible</h2>	<p>Par des interfaces par programme et des requêtes</p>	
<h2>Services identifiables</h2>	<p>Des services et des workflows de FdT documentés qui analysent et annotent le texte</p>	<p><a href="https://services.openminted.eu/home">https://services.openminted.eu/home</a></p>
<h2>Traitement efficace</h2>	<p><i>Cloud computing</i></p>	
<h2>Impliquer les communautés</h2>	<p>Les différentes communautés scientifiques ont des défis différents</p>	
<h2>Applications à forte valeur ajoutée</h2>	<p>Des applications guidées par les besoins des communautés démontrent la puissance de l'infrastructure</p>	



Réponse sous la forme de plateforme offrant des moyens de facilitation de la conception et de la maintenance d'une offre de FdT

**Approche infrastructure**

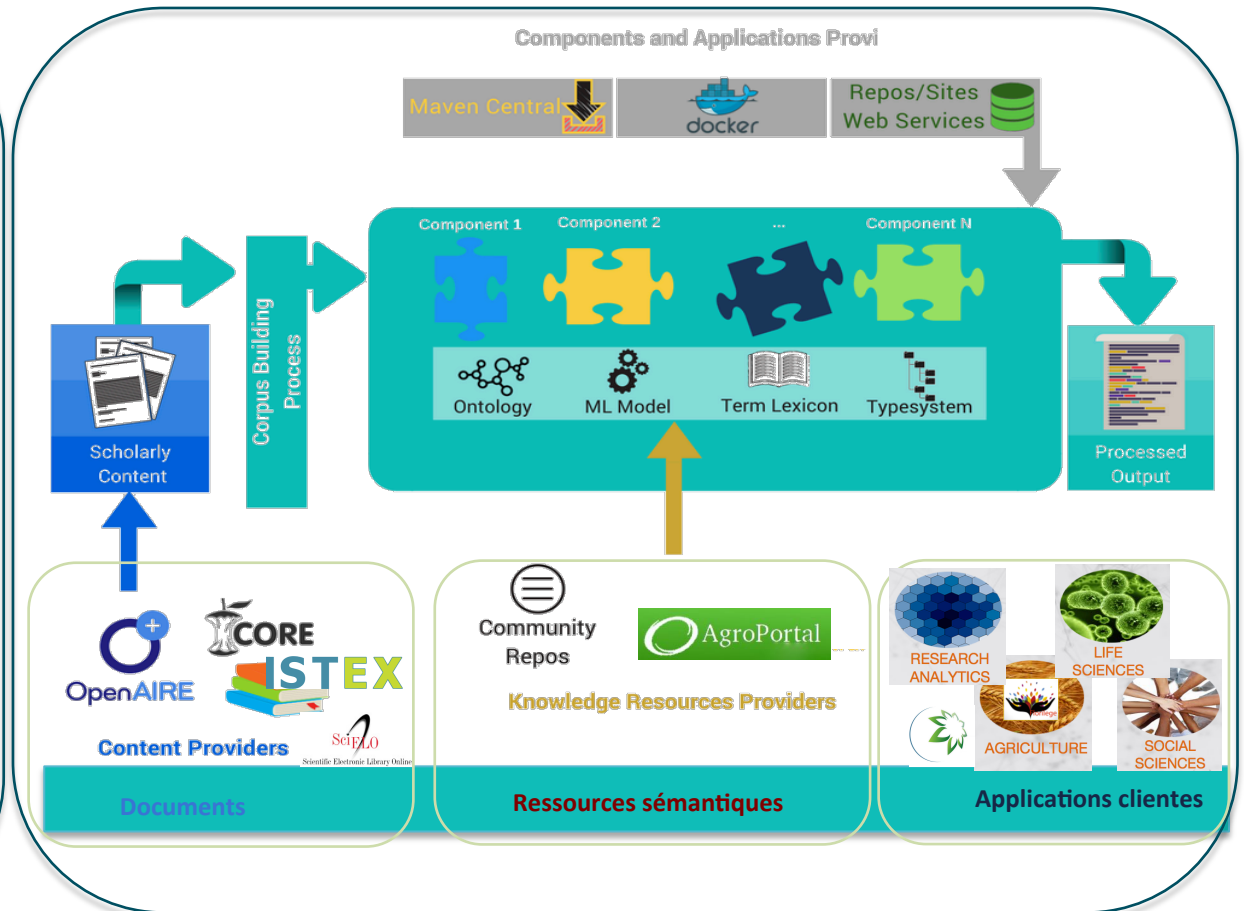
Adopte et adapte les services existants pour de nouveaux services et de nouvelles communautés.

Pour non spécialistes, une interface ergonomique, des catalogues structurés et des cours en ligne

Un accès ouvert dans un cadre unifié au service, aux applications, aux contenus et aux résultats des traitements de TDM.

Pour reproduire, adapter et combiner avec d'autres outils, réutiliser les workflows et remplacer les composants.

Les choix technologiques la rende évolutive et interopérable.



Voir les documents et démonstrations en ligne





## Vers une infrastructure française de service avancées de *text mining* Quelle exploitation d'OpenMinTeD ?

### Un projet du CoSO

Un *OpenMinTeD France* qui s'inscrit dans la stratégie de l'ESR français et réponde aux besoins spécifiques

- Interactions avec les organisations et communautés - Langue
- Formation, acculturation
- Calcul - Accès aux contenus

Pour contribuer à la création et à l'exploitation de l'offre de services

### Axes de travail

- |                            |   |
|----------------------------|---|
| <b>Etude</b>               | Analyse du besoin, organisation, missions, compétences, expertise technique d'OMTD  |
| <b>Conception</b>          | Utilisabilité et faisabilité du couplage avec les bibliothèques numériques (ISTEX) et les portails sémantiques (Agroportal) |
| <b>Pilotes – formation</b> | illustrer la facilité de déploiement et la qualité des résultats pour des défis scientifiques.                              |

### Partenariat

INRA DIST et MaIAGE-Bibliome  
Université de Montpellier, LIRMM-AgroPortal  
CNRS INIST, dépt offre de service, dépt innovation Istex

### Dynamique européenne

Dont OpenMinTeD : convention de partenariat, feuille de route  
Projets européens : infrastructure, EOSC, CEF

Rendez-vous au  
printemps 2019



**Le TDM aujourd'hui „  
des pièces détachées ...**

**... à la réalisation**

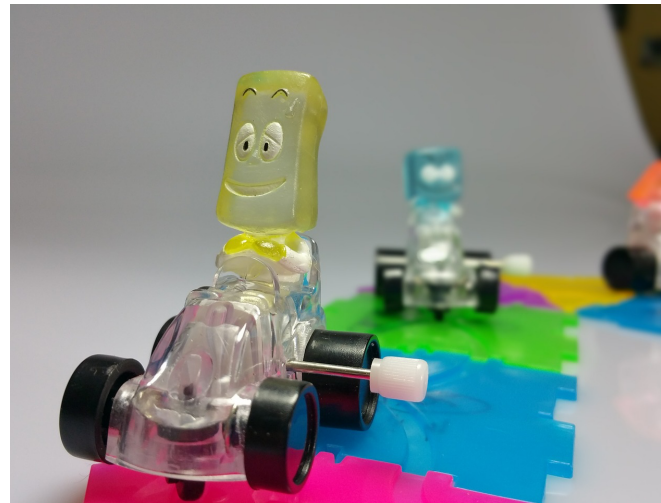
## Conclusion

Le texte n'est *pas une donnée comme une autre*.

La conception d'une infrastructure de services de TDM repose sur

- La réutilisation et la combinaison de nombreuses briques existantes
- et la coordination de nombreux acteurs

En ce sens, elle porte une question structurante  
Dont les enjeux méritent attention







Journées Nationales de la  
Science Ouverte

*Fouille de Texte et  
Science Ouverte*

Ouverture des données  
ouverture des logiciels  
ouverture des services