ASTN

Service ASTN/LLL

Appui Scientifique et Technique au Numérique

Y. Stroppa/F. Badin/A. Ly

13 février - 2023

Sommaire

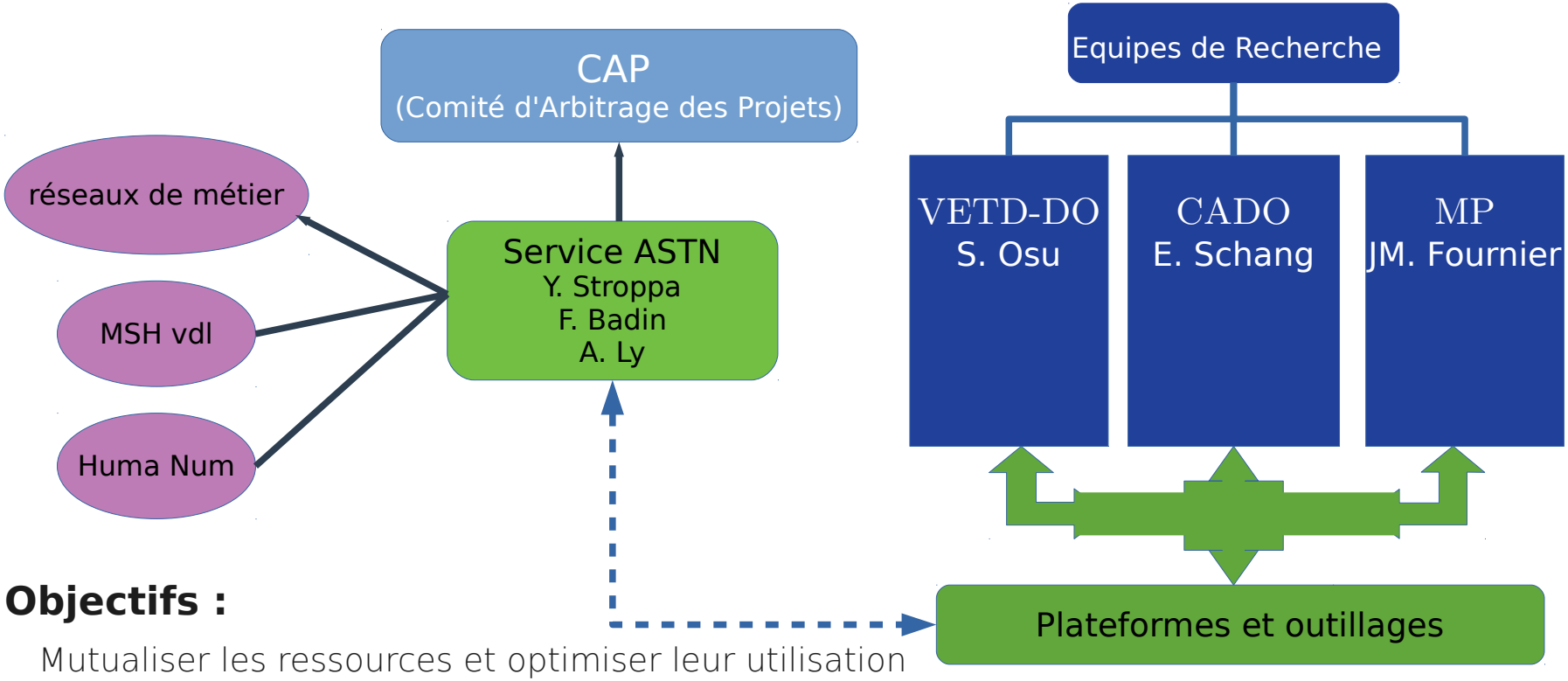
Service ASTN

Activités :

- Architecture globale : SILLL
- Plateformes -- outillages dédiés
  - AlimCorp : Alimentation de corpus
  - DeepCorp : exploitation et traitements corpus
  - IAML : pipeline de traitements en IA avec des ML
  - DeepBdd : plateforme de structuration de données
  - ....

Synthèse

Service ASTN : https://astn.lll-projects.org/



Objectifs :

Mutualiser les ressources et optimiser leur utilisation

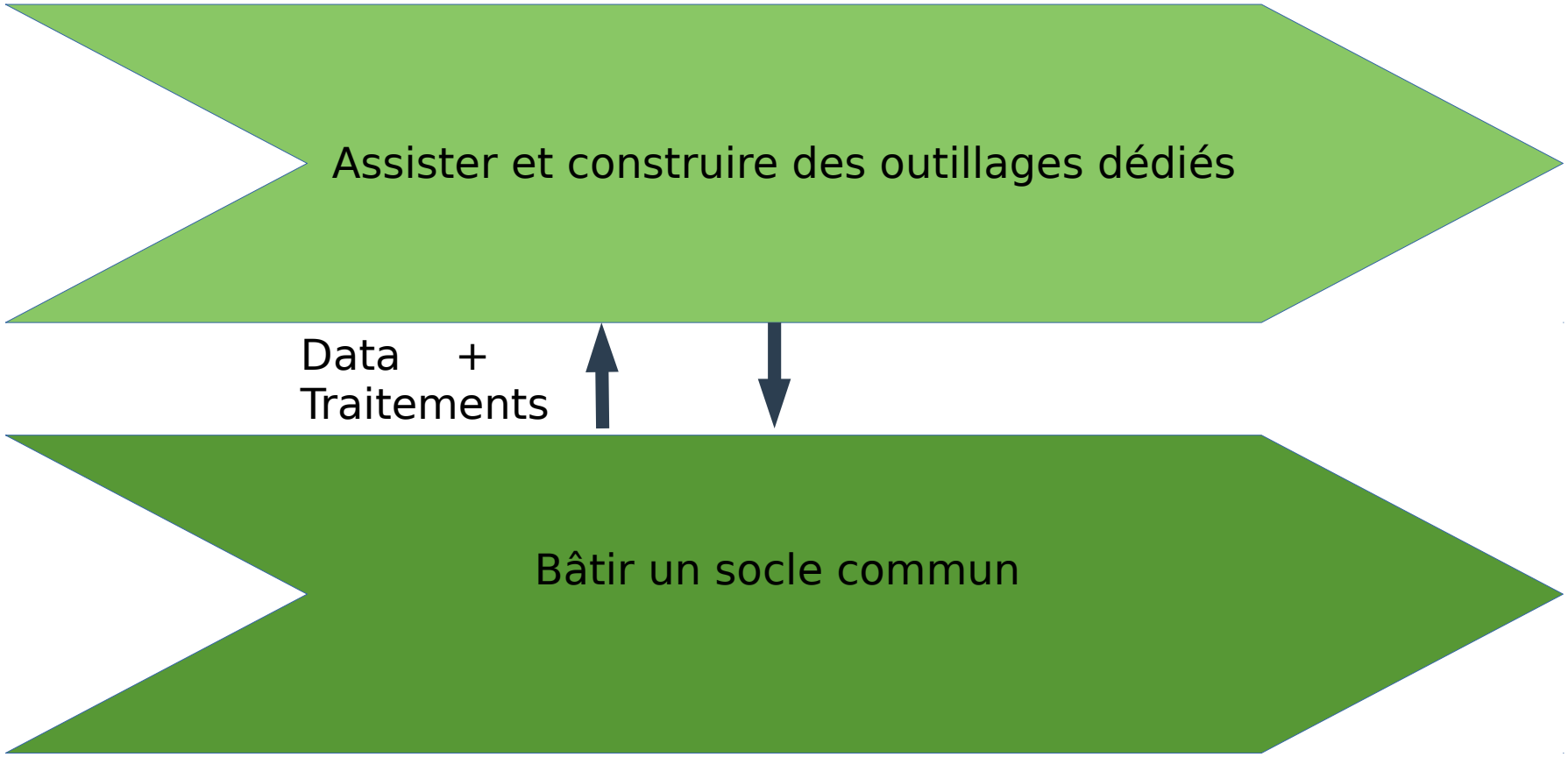
Pour :

- Répondre aux besoins des équipes dans leur projet de recherche
- Interlocuteur avec les réseaux de métiers
- Développer et déployer de l'outillage linguistique
- Maintenir l'ensemble des plateformes

Compétences :

linguistique,
infrastructure, développement,
ASR, Data scientist ...

Deux axes principaux :



Architecture de SILLL :

ASTN

Objectifs/Mots clés :

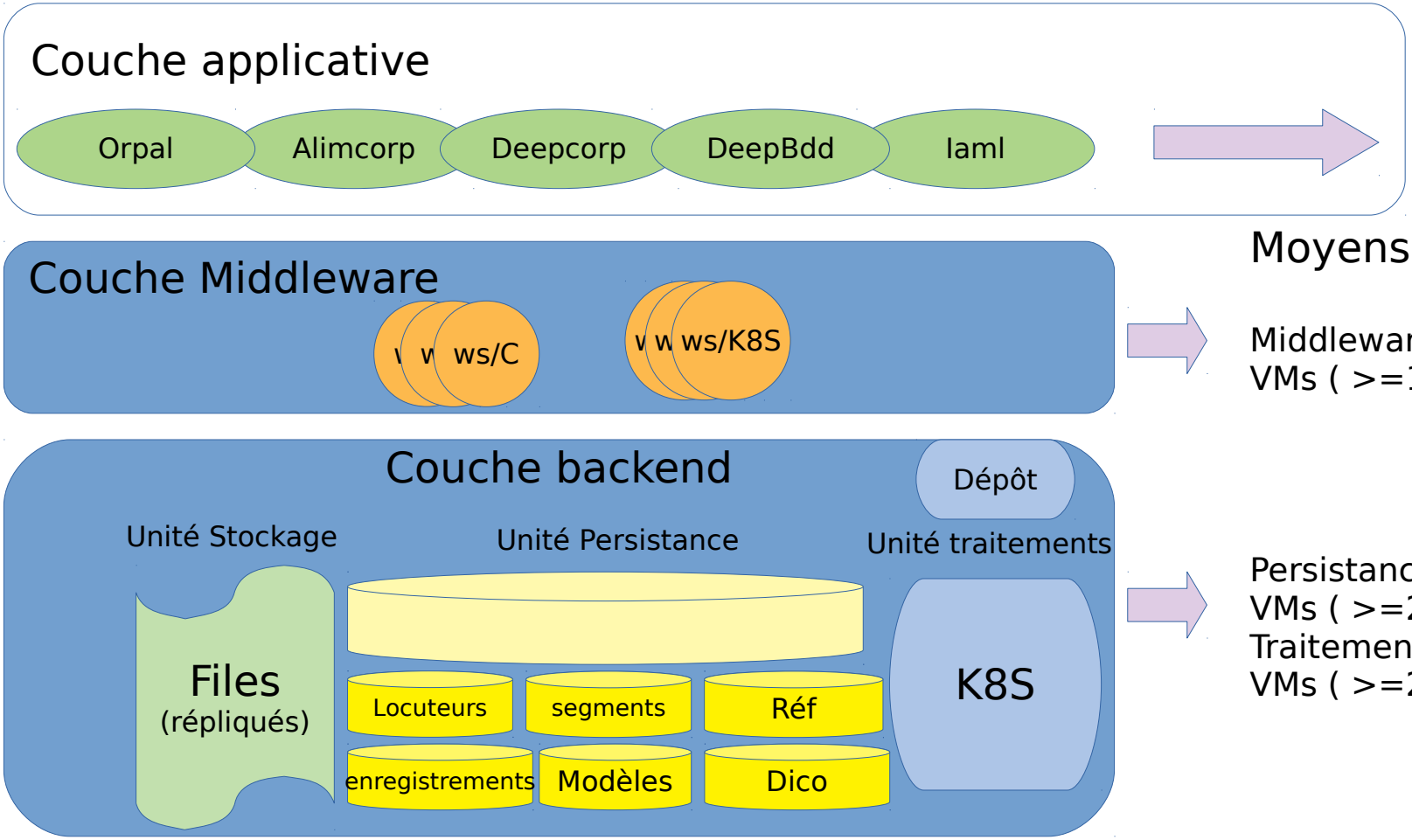
Capitaliser
Sécuriser
Partager
Réutiliser

Mutualiser
Exécuter

Pérenniser
Stocker

Partenaires :
LLL

Acteurs :
ASTN



Moyens :

Middleware
VMs (>=1)

Persistance
VMs (>=2)
Traitements
VMs (>=2)

Plateforme : AlimCorp

Objectifs/Mots clés :

Structure personnalisée
Alimentation simple
Collaboration
Savoir faire
Capitaliser
Organiser
Echanger
Assistance/
Accompagnement
Conserver

Partenaires :

LLL

Utilisateurs visés :

Chercheurs
Doctorants

Définition de la structure de données

- Modèle
- Workflow
- Définition des échanges

Production du modèle de corpus

Instanciation du modèle de corpus
=> création du corpus

Alimentation du corpus

- Formulaires dédiés - Assistance à la saisie
- Gestion des droits
- Partage et échange de message
- Suivi et contrôle

Exports

SILLL

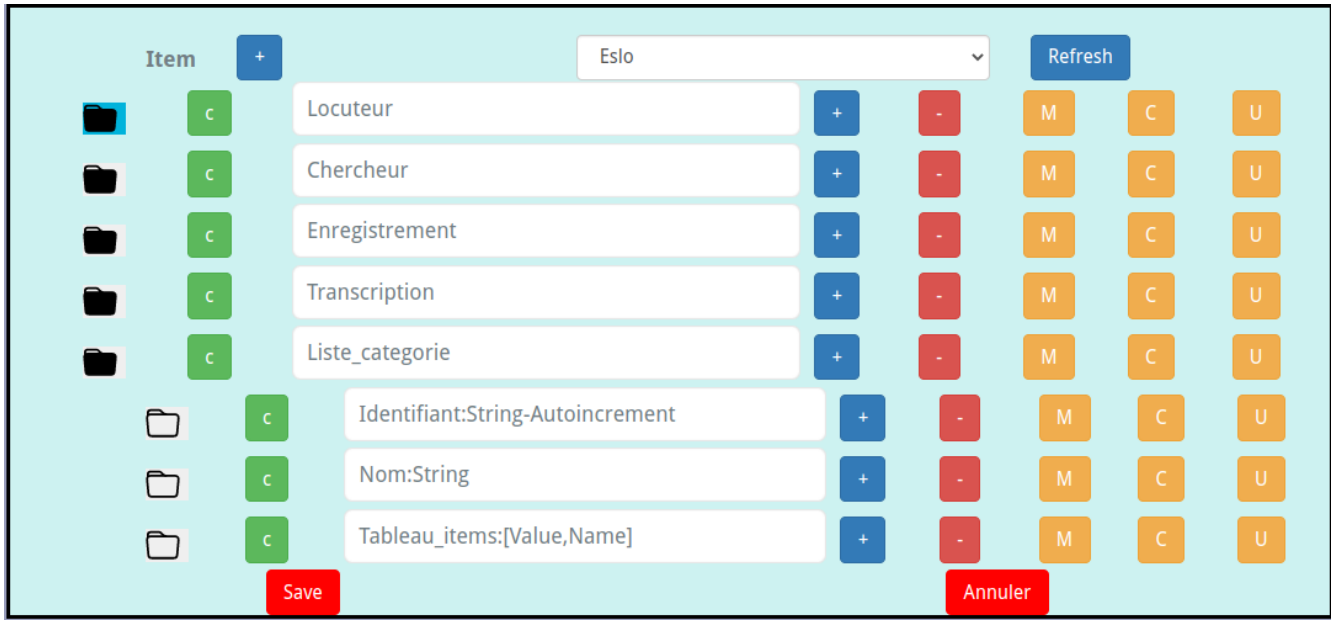
Autres

Moyens :

VMs (>=1)

peut-être préciser le domaine d'application

Structuration du modèle

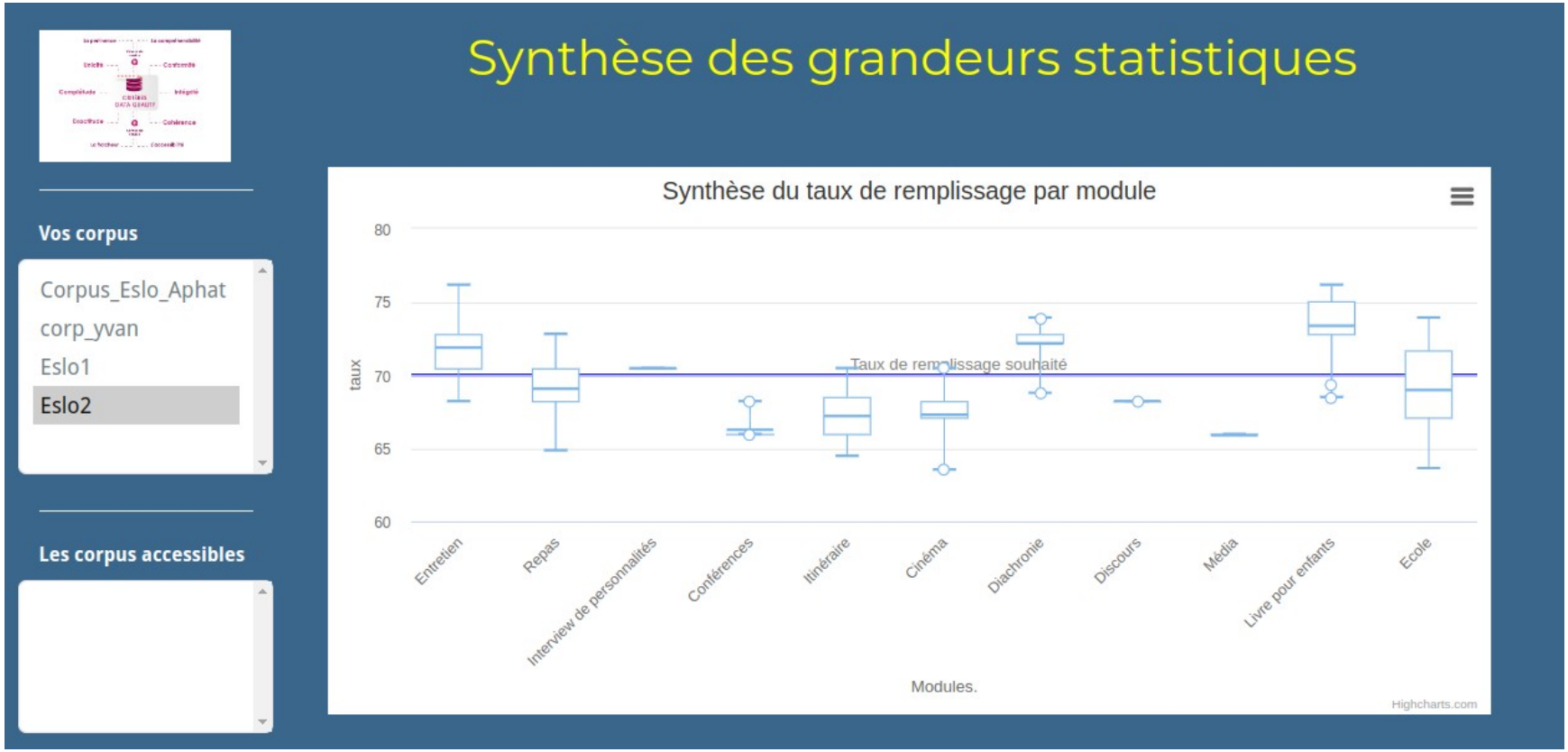


L'utilisateur aura la possibilité de définir et d'organiser son modèle de données en utilisant une forme descriptive hiérarchique.

Alimentation

Formulaire de saisie des différents enregistrements avec suivi instantané des taux de remplissage (méta-données).

Suivi alimentation



Plateforme : DeepCorp et al

ASTN

Objectifs/Mots clés :

Diffusion
Communication
Capitaliser savoir-faire
Capitaliser des méthodes
Traitements et extractions

Portail : DeepCorp
- Communication, diffusion
- Exploration des sources
- Traitements
- Espace de travail

multi-corpus

Site : Eslo
- Communication, diffusion
- Exploration des sources
- Traitements
- Espace de travail

mono-corpus

Site : DeepTunisien
- Communication, diffusion
- Exploration des sources
- Traitements
- Espace de travail

mono-corpus

Moyens :

VMs (>=1)
MiddleWare
Backend
persistance

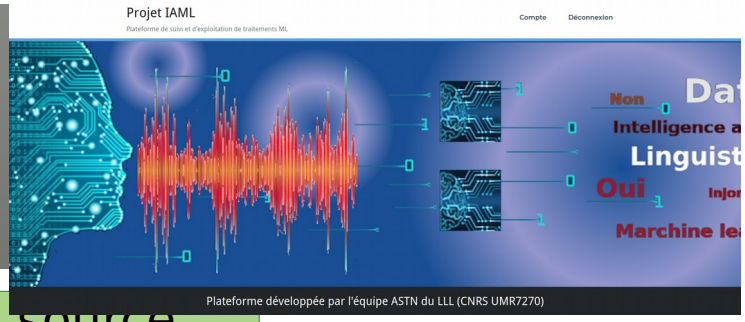
Partenaires :

LLL

Utilisateurs visés :

Chercheurs
Public

Plateforme : IAML



Objectifs/Mots clés :

Collaboration
ML/IA
Capitaliser savoir faire
Capitaliser méthode
Suivi
Assistance
Disponibilité
Expertise

Partenaires :

LLL
PRISME

Utilisateurs visés :

Chercheurs
Public

itération

Sélection et construction source
- Sélection des segments

Exploration et analyse des sources
- Analyse statistique et linguistique des segments de la source

Construction des plans d'expérience
- Définition des objectifs
- Définition des membres

Alimentation des ML

Traitements avancés
- Sélections et constitutions des lots
- Appariement lots, phases et ML

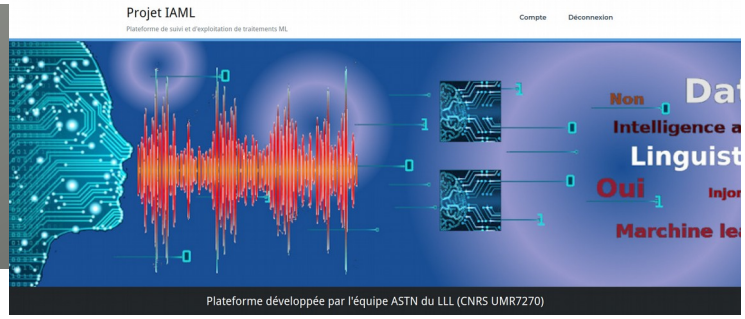
Exécution/K8S

Exploitation des résultats
- Analyse et expertise

Moyens :

VMs (>=1)
MiddleWare
Backend
traitements

Interfaces



Les services

Liste des services offerts par cette plateforme

Présentation

Description de l'objectif de cette plateforme et des moteurs d'analyse

Gestion des sources

Constitution et création des sources de données provenant des autres plateformes SILLL

Exploration des données

Formulaire d'extraction et de consultation des données à partir de requêtes élaborées.

Gestion des solveurs

Dépôt des solveurs de calcul pour la plateforme - description et sources.

Gestion des plans d'expérience

Gestion des plans d'expérience et campagne de tests

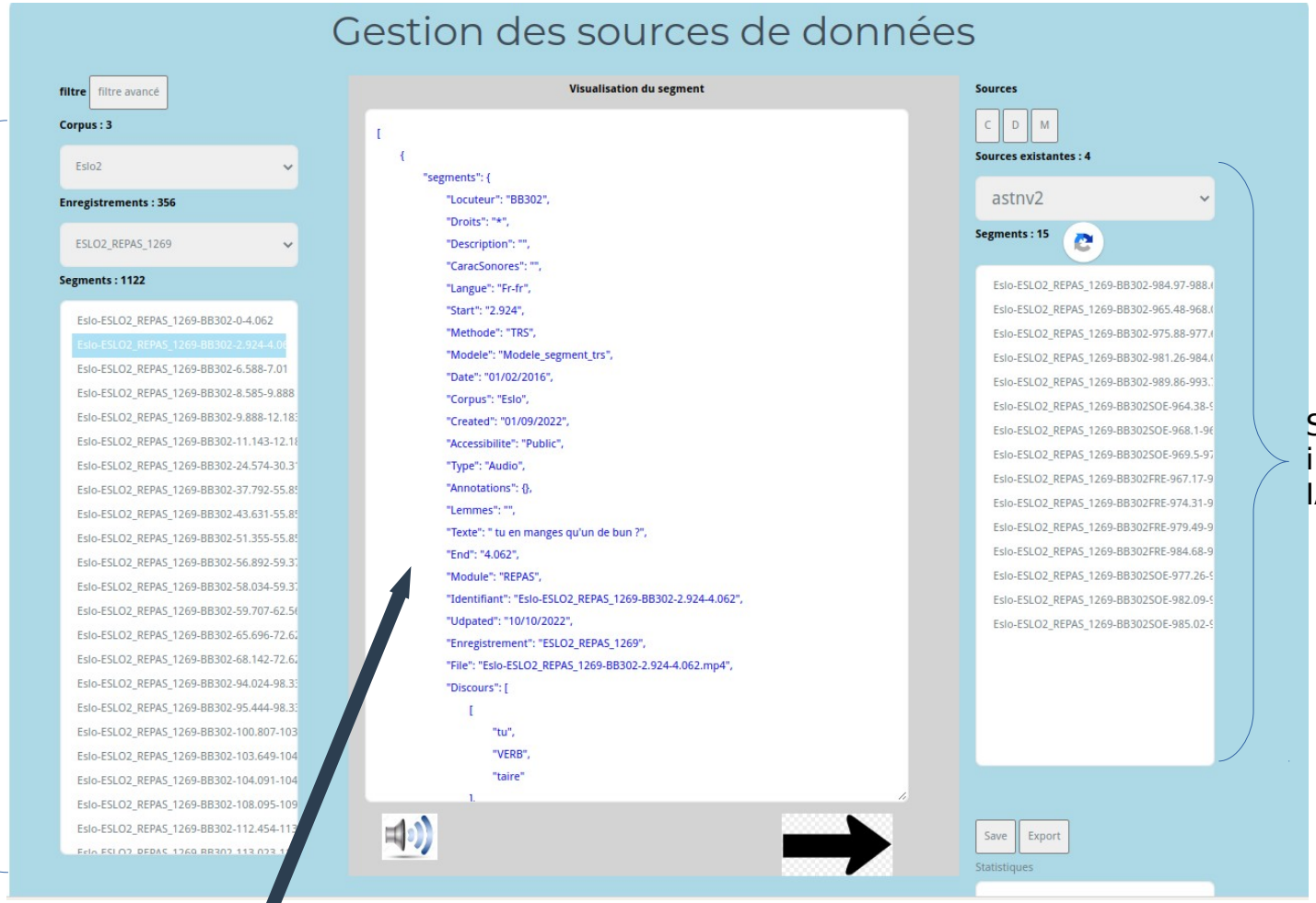
Traitements avancés

Constitution des lots et préparation des demandes d'exécution (jobs de calculs)

Suivi des jobs et résultats

Suivi des jobs (pending, success) et exploitation des résultats de traitements

Création des sources de données à partir de filtres évolués



Segments provenant de SILLL

Structure interne IAML

SILLL

Contenu du segment sélectionné

Constitution du modèle de données

Modèle issu des segments stockés dans SILLL

Nouvelle structure de données pour la source et règles d'appariement.

Plateforme : DeepBdd

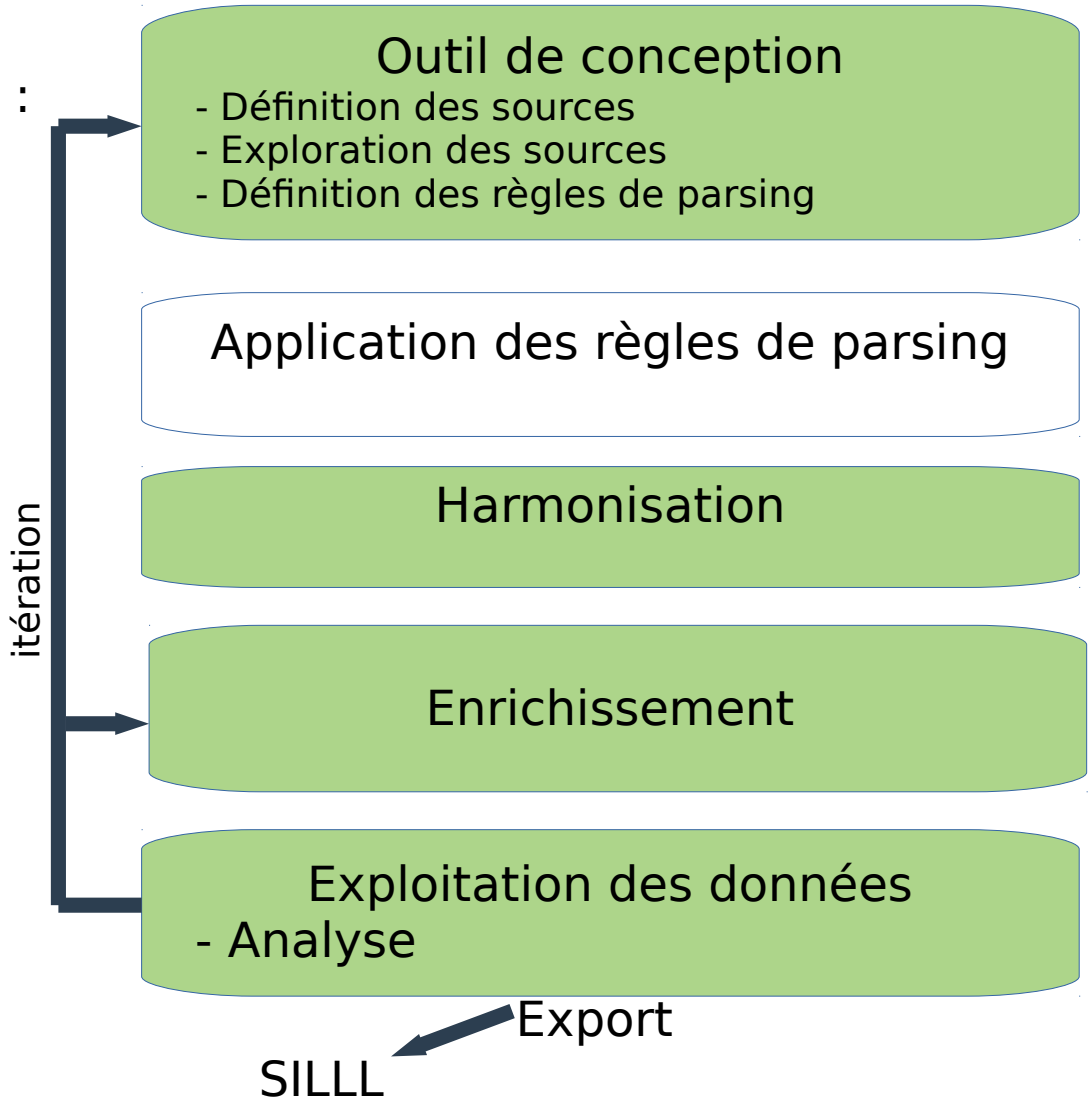
ASTN

Objectifs/Mots clés :

Collaboration
"Phonétique"
Méthodologie
Suivi
Assistance
Langage
Reporting

Partenaires :
LLL

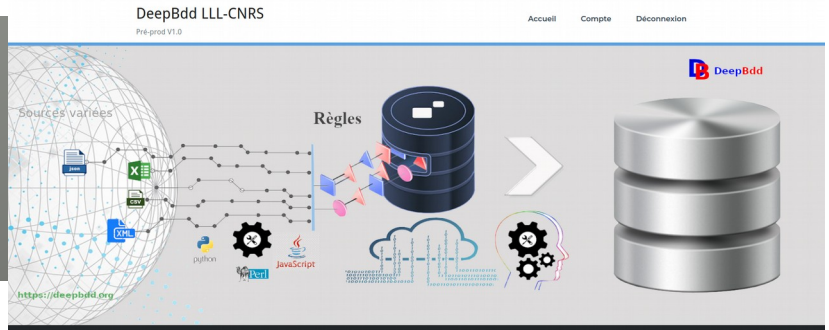
Utilisateurs visés :
Chercheurs
Public



Moyens :

VMs (>=1)
MiddleWare
Backend
persistance

Présentations



Plateforme développée par l'équipe ASTN du LLL (Orléans)

DeepBdd - Bddictionnairique

Plateforme élaborée avec l'équipe de Tours du LLL

Fondation

Détail de la construction et des règles utilisées pour la construction de cette version de plateforme....

Sources

Constitution à partir de fichiers externes (importation) d'une source de données et description de son contenu..

Visualisation et exploitation

Accès au contenu original avec possibilité d'interroger et d'extraire les données.

Pré-traitements

Définition et activation des pré-traitements sur les sources de votre projet (module restreint).

Parser

Description des règles de reconstruction (parsing) à appliquer aux données issues de vos sources(module restreint)

Méta-Données

Règles d'association et de dissociation des mots pour constituer le dictionnaire complet.

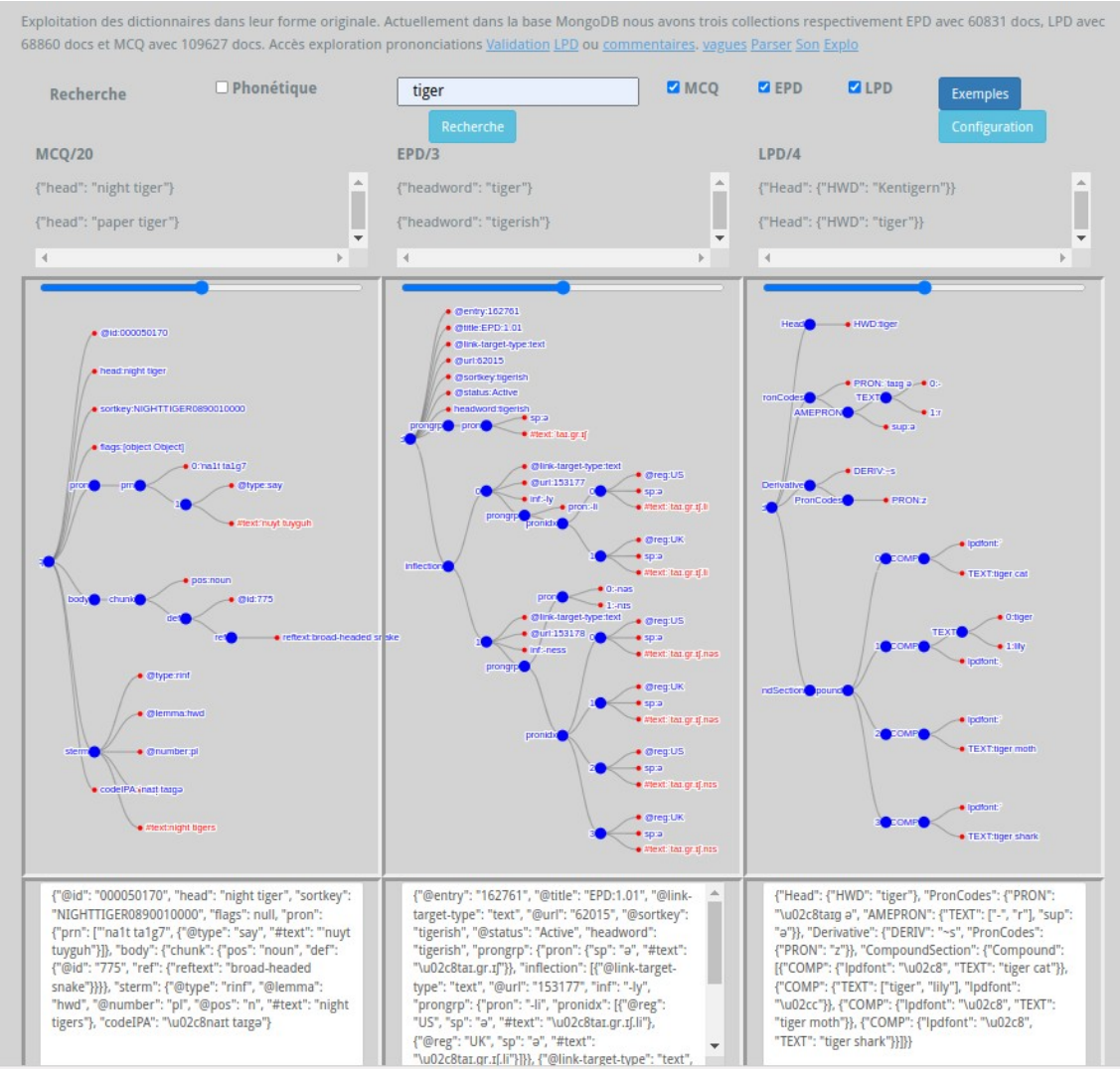
Exploitation/Traitements

Module d'exploitation des sources constitués par la plateforme (filtrage pour exploration et exportation)

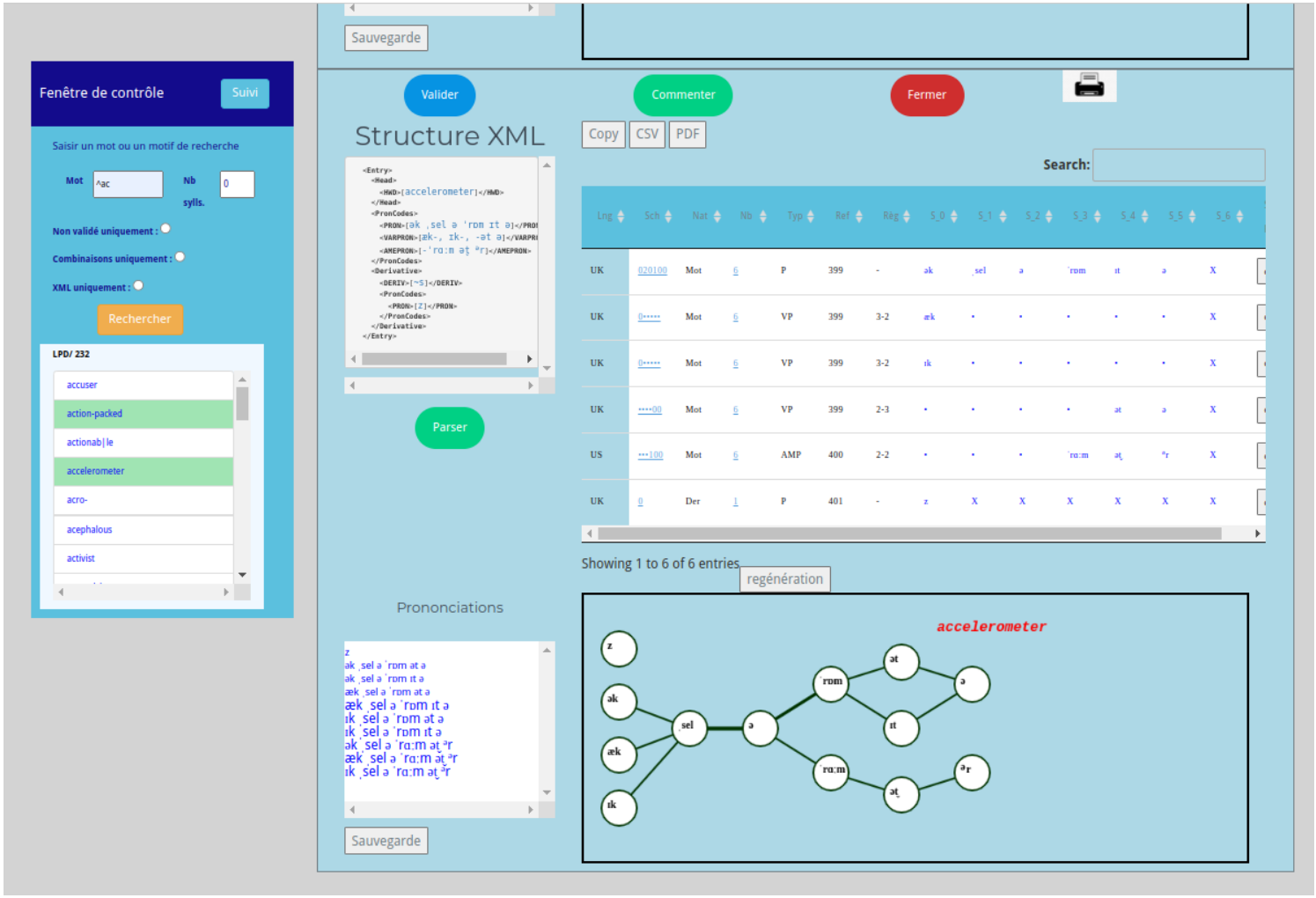
Enrichissement

Enrichissement de la plateforme par l'ajout de différentes informations complémentaires (module réservé)

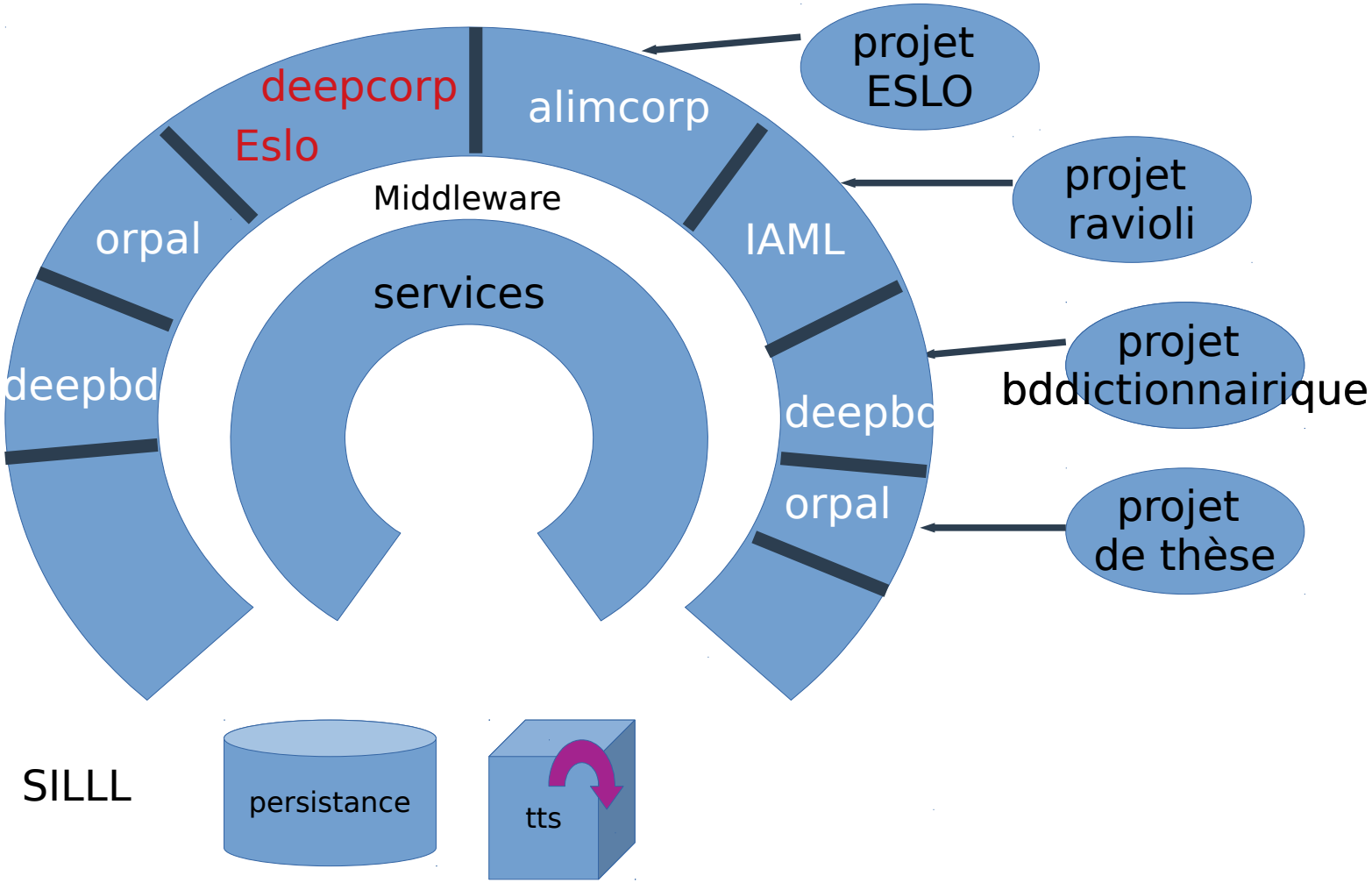
Illustrations



Interprétation des règles phonétique LPD

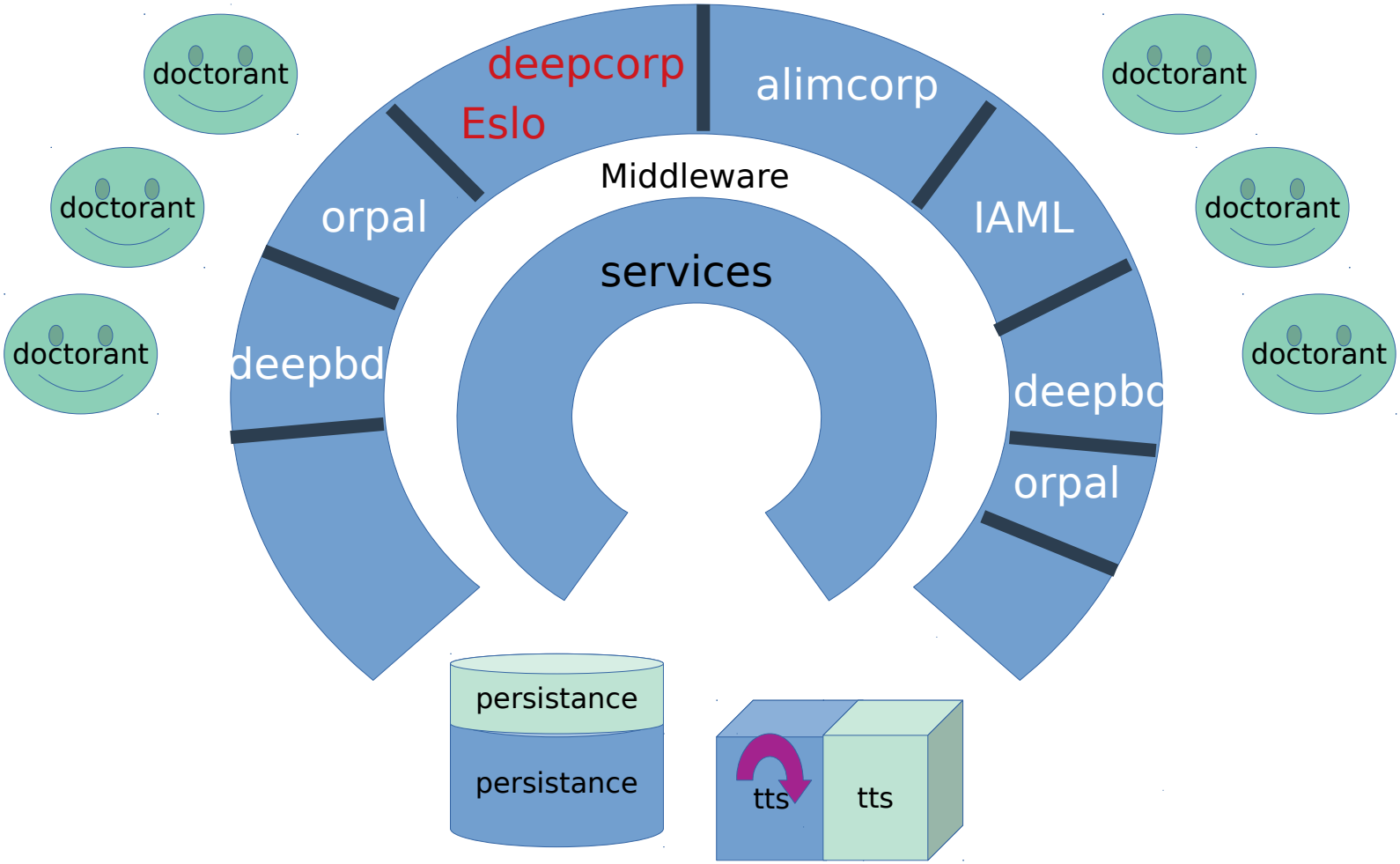


Synthèse : Phase conception



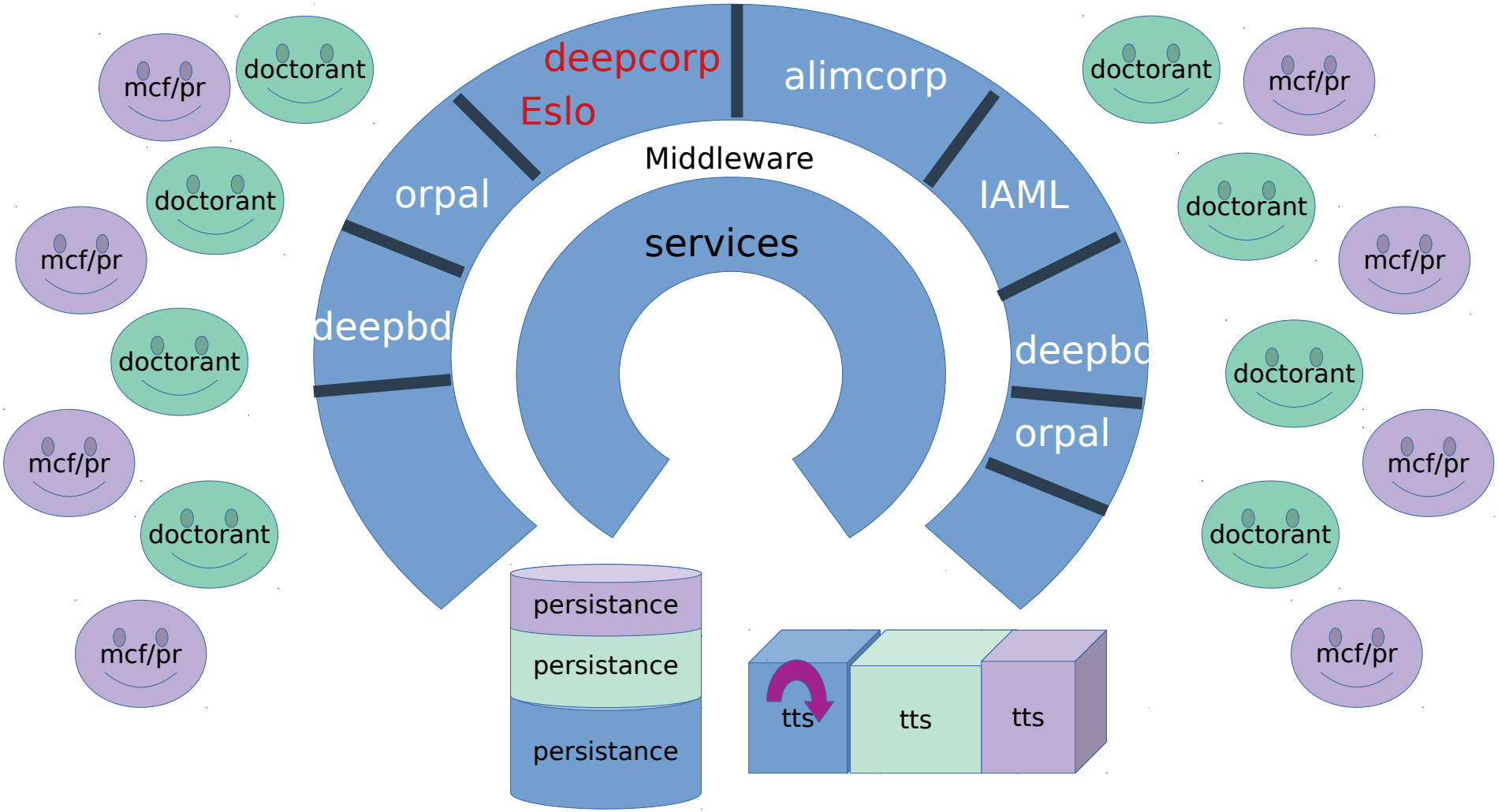
2021-->2023 : élaboration de l'architecture/Applis

Synthèse : Phase déploiement 1



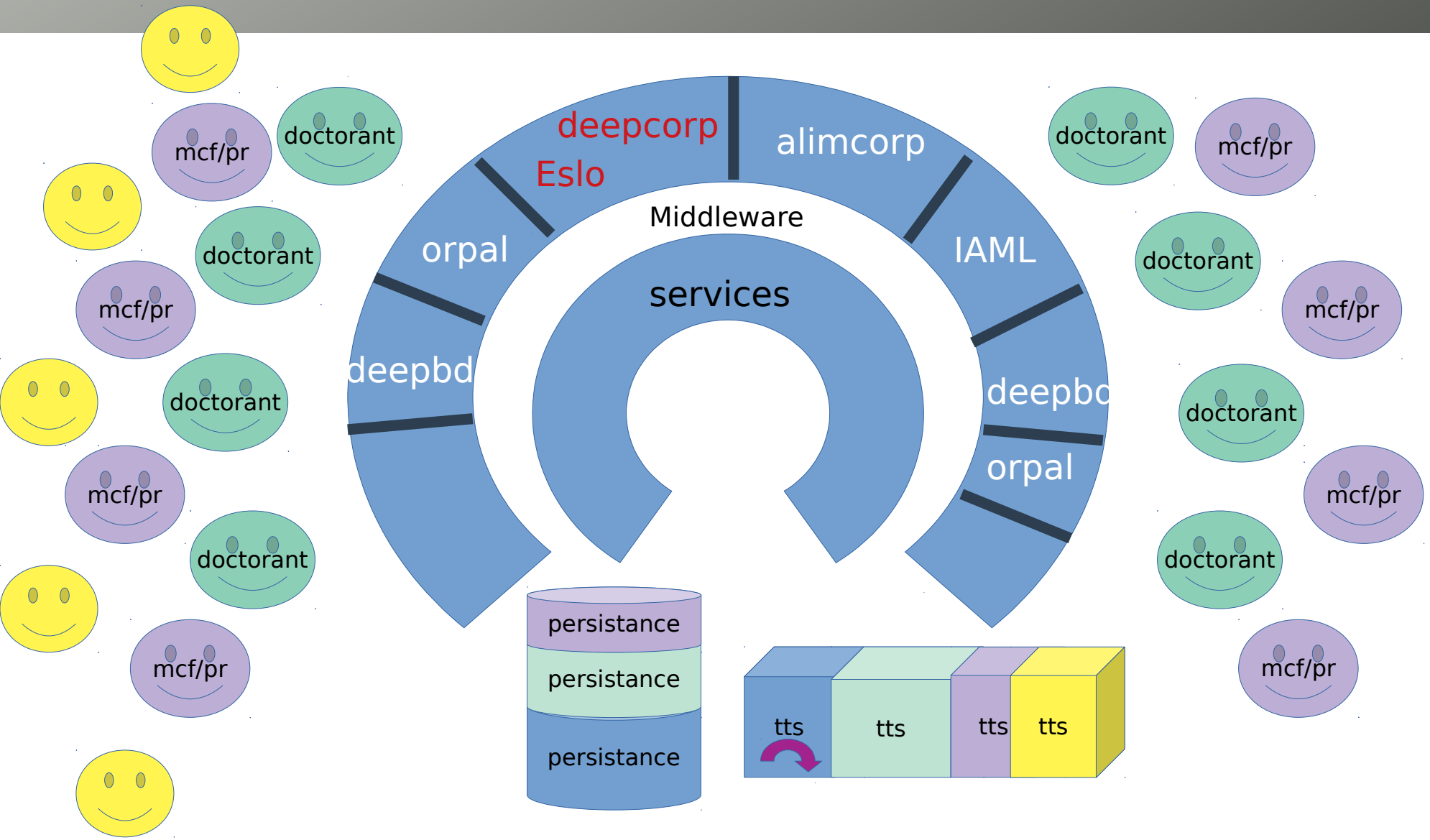
2023 -2024 : alimentation et exploitation docs

Synthèse : Phase déploiement 2



2025 -2026 : alimentation et exploitation mcf/pr

Synthèse : Phase déploiement 3



2026 - : alimentation et exploitation public

Intérêts scientifiques : pour Lotfi à compléter

Meilleure attractivité

Ouverture sur des domaines connexes

Facilité l'interaction avec d'autres disciplines

Exploration innovante avec croisement des données et des observations

Accumulation d'une richesse sans commune mesure - data - traitements

Contexte favorable pour l'émergence d'approche croisée pour la production de nouveaux savoirs ...

Ouverture sur le partage des traitements

Reproductibilité

....