

Perspectives

Journées Istex 2025

16-17 juin 2025 - Nancy

Projets

AIKO

- Agence de programme numérique INRIA
- Projet de recherche sur 5 ans (16 thèses)
- Une partie infrastructure : ISTEX+
- Budget de 5 millions
 - 500k€ pour la partie infrastructure
- **Objectif pour ISTEX:**
 - Stockage et mise à disposition de grands corpus de publications scientifiques d'origines hétérogènes directement exploitable par les IA (data lake)

LLMs4EU

- Union Européenne (ALT-EDIC)
- Projet européen sur 3 ans (70 partenaires)
- Budget de 20 M€
 - 460k€ pour le CNRS
 - 41k€ pour l'Inist
- **Objectifs pour ISTEX:**
 - Fourniture de données transformées pour des LLMs
 - Possibilité d'utiliser les LLMs produits
 - Génération de requête ?
 - Moteur de réponses ?

GIS Collex-Persée

Programme acquisitions

- Enjeu pour ISTEX
 - Intégrer des nouvelles ressources spécialisées
- Problématiques
 - Valorisation de ressources spécialisées
 - Services et IA spécifiques pour/avec des ressources spécialisées

Programme numérisation

- Co-pilote du GT “Disponibilité des données pour la fouille”
- Enjeu pour ISTEX
 - “Des corpus mixtes”
- Problématiques
 - Mise à disposition et exploitation des données numérisées
 - Valorisation des données numérisées
 - Services et IA spécifiques pour/avec des ressources numérisées

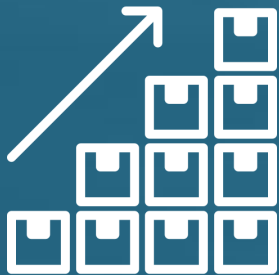
Programme cartographie

- Enjeu pour ISTEX:
 - “Élargir les usages de Lodex”
- Problématiques
 - prise en compte de nouveaux formats de jeux de données
 - enrichissement et web services dédié à la fouille de métadonnées bibliographiques spécifiques.

Développements

Ambitions

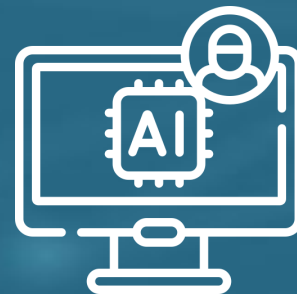
Augmenter la quantité
des textes et de données



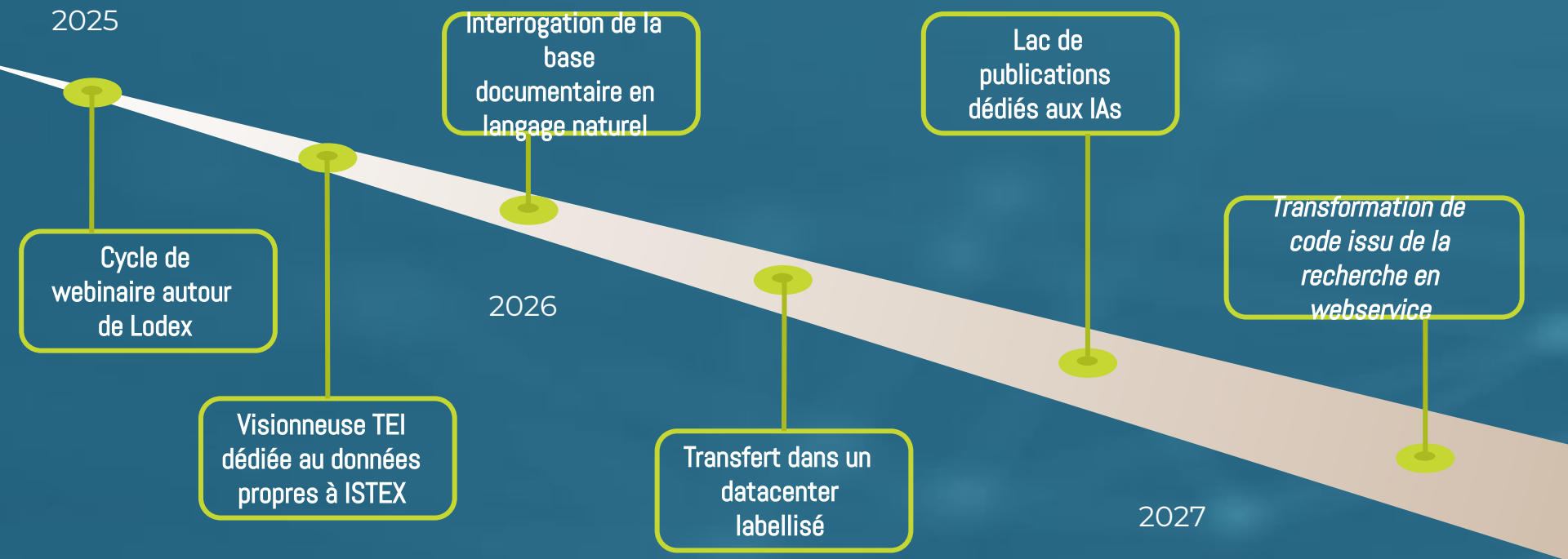
Valoriser la richesse actuelle
des textes et des données



Proposer de nouveaux
services à base d'IA



Projection





ISTEX Infrastructure de recherche

Calendrier

- 2021- 2025: projet d'infrastructure de recherche
- 2025 : candidature vers statut infrastructure
 - mi-avril : dossier préliminaire
 - 18 juillet : dossier définitif
 - fin août : audition
- début 2026 :
Publication feuille de route des infrastructures de recherche

Nouvelle gouvernance

Les instances

- Comité opérationnel
- Conseil scientifique et technique
première réunion : 25 juin 2025
- Comité de pilotage
première réunion : 3 septembre 2025
- Directrice
- Directeur opérationnel

Le comité opérationnel

- Met en oeuvre la stratégie, prépare et suit les réalisations
- Organisation en groupes de travail
 - Alimentation de la plateforme mise à jour, type de données, droits accès
 - stockage gestion et traitement des ressources
 - Usages accompagnement , formation
 - Ressources spécialisées terminologies, jeux entraînement IA
 - Fouille de textes
 - Analyse de données Lodex
- Deux représentants des établissements
élection à venir

Le conseil scientifique et technique

- Patrice Bellot - LIS - CNRS/Univ. Aix-Marseille
- Florian Boudin - LS2N - CNRS/Université de Nantes
- Elena Cabrio - I3S - Univ. Nice, CNRS, INRIA)
- Mathieu Constant - ATILF - CNRS/Université de Lorraine
- Liana Ermakova - HCTI - Univ. Brest
- Cécile Favre - ERIC - Univ. Lyon 2
- Olivier Ferret - LIST - CEA
- Gilles Hubert - IRIT - Univ. Toulouse /CNRS
- Thierry Ménissier - IPHIG - Univ. Grenoble
- Patrice Lopez - Science Miner - membre de la WIPO
- François Yvon - ISIR - CNRS/Sorbonne Université

Le comité de pilotage

- S Rousset : Présidente
Directrice des données ouvertes de la recherche du CNRS
- Représentants des Partenaires
 - Nicolas Morin : ABES
 - Sébastien Perrin : Coordination exécutive de Couperin
 - Nicolas Fressengeas : Université de Lorraine
 - Laurent Romary : INRIA
- Représentant de France Université :
Xavier Leroux - Université de Toulon
- Représentants élus des établissements adhérents :
 - Thierry Fournier - Université de Rennes
 - Marie-madeleine Géroudet : Université de Lille
- Président élu du conseil scientifique et technique (élection à finaliser)