

De la base aux services

Présentation générale de l'infrastructure Istex

The logo for ISTEX, with 'IST' in white and 'EX' in yellow, is centered on a blue background featuring a network of glowing nodes and lines. The background transitions from blue on the left to green on the right.

ISTEX

Journées Istex 2025

16-17 juin 2025 - Nancy

14 ans d'évolution



2 phases

Plateforme documentaire

Phase 1

#1 Plateforme documentaire

2011 - 2018 : un projet créé dans le cadre des PIA
(Programme d'Investissement d'Avenir)

“Construire le socle de la bibliothèque scientifique numérique nationale”



#1 Plateforme documentaire

- **Acquisition massive et centralisée d'archives scientifiques**
 - Issue des Licences Nationales
 - Collections rétrospectives multilingues et multidisciplinaires
- **Mise à disposition d'une base documentaire**
 - Intégration dans les portails documentaires des établissements (API, widgets, Bouton)
 - Intégration dans les outils de découverte et résolveurs commerciaux (primo, eds, alma, etc.)
 - **stockage des textes homogènes et enrichis**

Infrastructure pour la fouille de textes

Phase 2

#2 Infrastructure pour la fouille de textes

2019 - 2025 : Projet d'infrastructure pour la fouille de textes

Depuis le 8 mars 2022, Istex est inscrit dans la feuille de route nationale des infrastructures de recherche du Ministère de l'Enseignement supérieur et de la recherche

“Réservoir d’archives scientifiques, outils et ressources pour la fouille et l’analyse de textes.”

#2 Infrastructure pour la fouille de textes

- **Elargissement du périmètre des ressources mises à disposition**
 - bouquets de documents en libre accès
 - corpus spécialisés, terminologies, etc.
- **Mise à disposition de services spécifiques pour la fouille et l'analyse**
 - Téléchargement massif de documents plein texte
 - Outil d'analyse de corpus (Lodex)
 - Traitements par IA des textes et métadonnées.

Offre de services pour la fouille de textes

Définition

Démarche

*Proposer une offre de services d'analyse et de fouille de textes **de bout en bout***



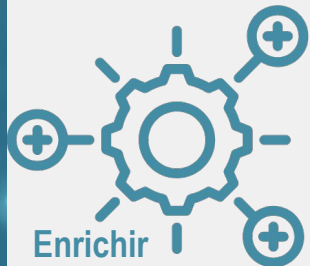
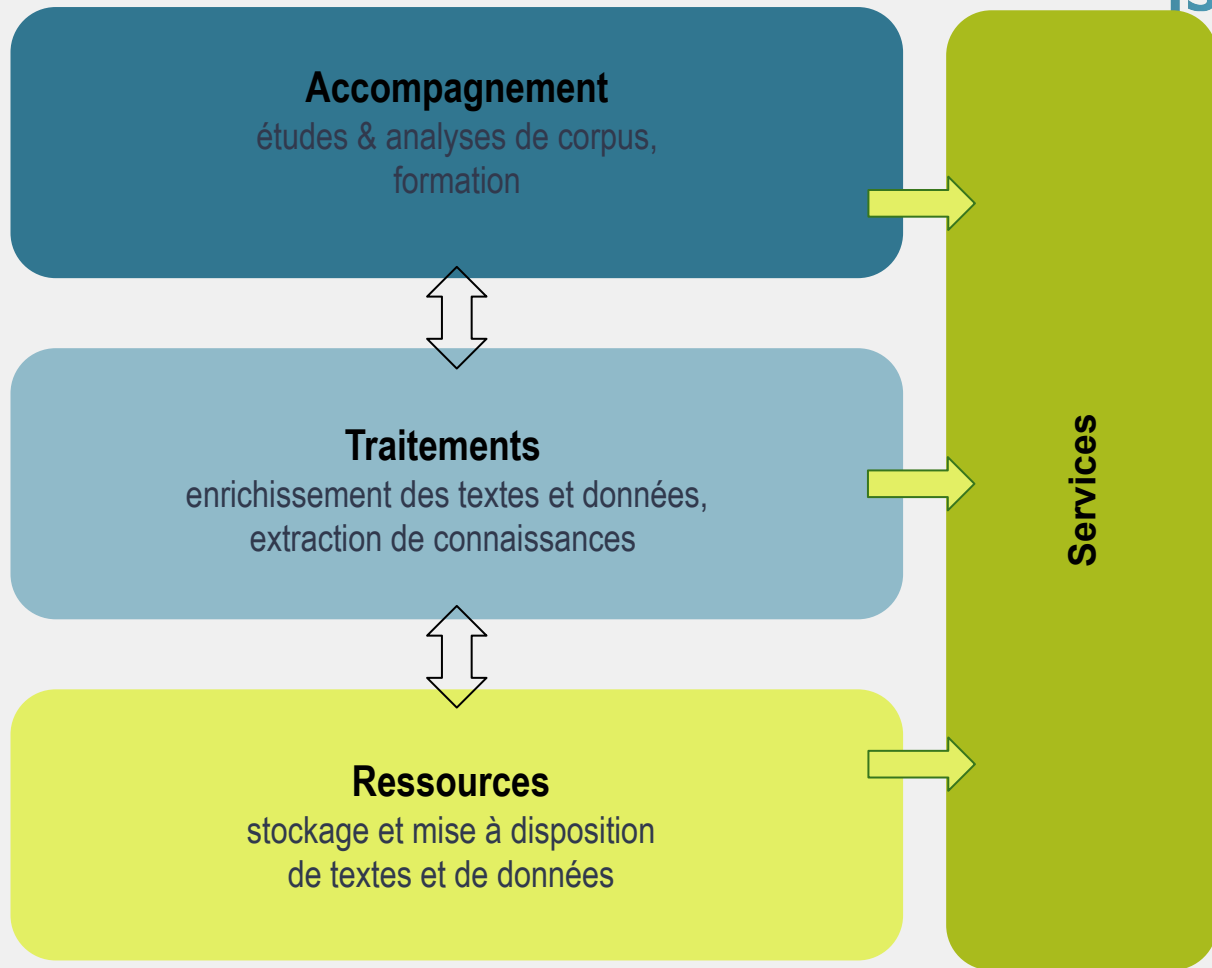
Stocker



Enrichir



Analyser

**Analyser***analysis, analytics***Enrichir***mining***Stocker***storage*

méthodologie de bout en bout

**Contenu et accès
aux publications
intégrées dans votre
espace de travail**

**1. Accéder aux
documents**

**Démarrer votre
projet de fouille
de texte en constituant
votre corpus à l'aide
des ressources Istex**

**2. Utiliser des
corpus de
documents**

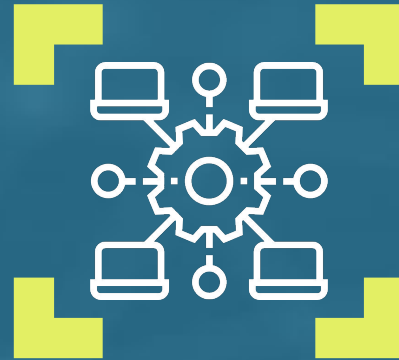
**Les technologies
et les outils de
l'infrastructure Istex
au service de votre
projet de fouille de textes
sur vos propres données**

**3. Explorer et
enrichir des
corpus**

Offre de services pour la fouille de textes

Présentation

textes & services



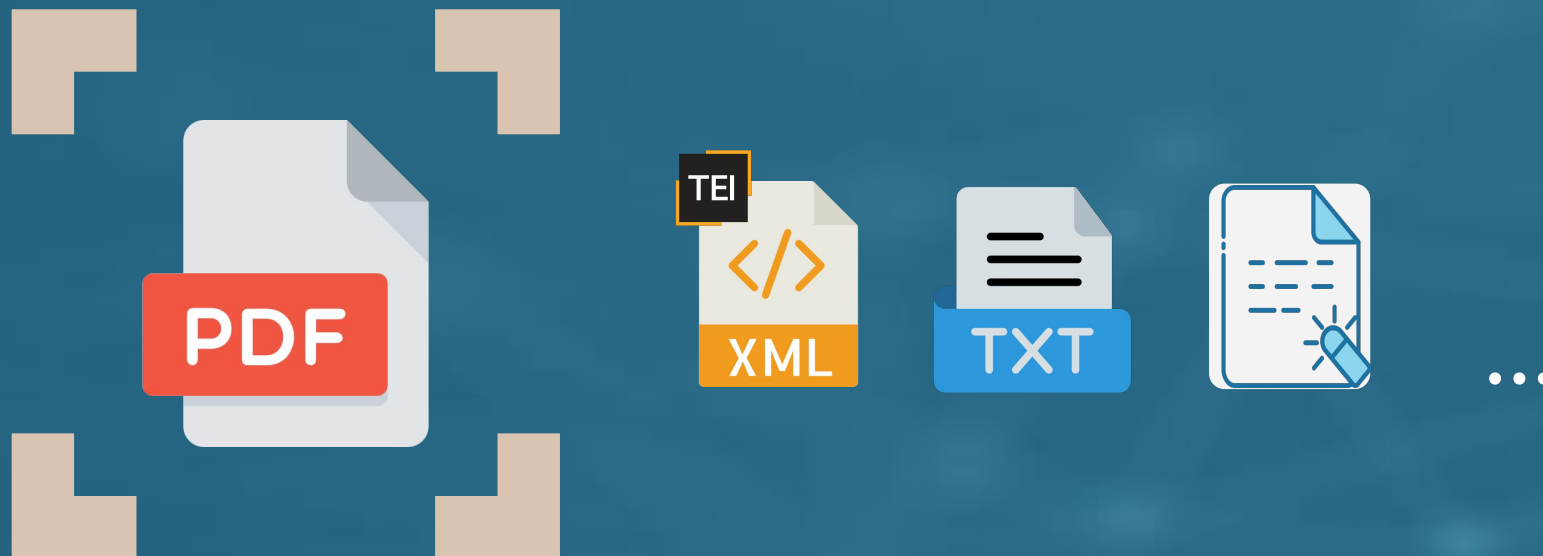


30 064 284

documents

Soit une bibliothèque avec ~14 km de rayonnage

Un document, plusieurs versions du même texte



Plus de 300 000 000 de fichiers

3 types de services



Ressources



Outils



Appui



Web services
pour les personnes
qui connaissent la
programmation.

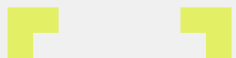


Lodex
pour les personnes
qui connaissent les
outils de **gestion de
données**

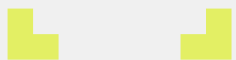


TDM Factory

pour les personnes
qui connaissent les
outils bureautiques



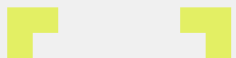
Ressource



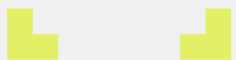
Loterre

pour les personnes
qui s'intéressent aux
**termes
scientifiques**





Ressource



Corpus
pour les personnes
qui s'intéressent à
**une thématique
scientifique**





Appui

Formation

pour les personnes
qui souhaitent
découvrir **la fouille
de textes**

ISTEX

Plusieurs équipes à votre service