

ISTEX

L'excellence documentaire pour tous

SÉMINAIRE TECHNIQUE CHANTIERS D'USAGE

NANCY 6-7 JUIN 2017

Jean-Marie Pierrel – Université de Lorraine



couperin.org
Consortium Universitaire de Publications Numériques

abes
agence bibliographique de l'enseignement supérieur



- **Faire le point** sur les résultats des travaux menés au sein d'ISTEX au cours des dernières 18 mois
 - **enrichissement** des données ISTEX
 - **services à valeur ajoutée**
- **Faire un bilan des projets « Chantiers d'usage »** sélectionnés suite à l'appel à propositions ISTEX
- Permettre aux différents contributeurs de **prendre connaissance** de l'ensemble des résultats des projets en cours
- **Préparer la fin du projet ISTEX**
 - La fin du projet ISTEX est reportée au 31/12/2017

- **Conforter une dynamique de recherche/développement** autour de la plateforme ISTEX qui puisse servir de déclencheur à des activités plus larges d'appropriation par les chercheurs des contenus d'ISTEX pour développer des recherches de Text and Data Mining (TDM) de qualité.
- **Préparer la seconde phase** du projet ISTEX qui devrait amplifier les usages de TDM sur les corpus ISTEX



**Mise en place de coopérations
entre les divers acteurs d'ISTEX et
les équipes de recherche
impliquées**

- **Aujourd'hui : présentation des derniers travaux menés au sein d'ISTEX**
 - Ce matin : les projets d'**enrichissement plein texte** des données ISTEX
 - Cet après-midi : les projets « **Services à valeur ajoutée** »



Présentation des résultats acquis exploitables par tous

- **Aujourd'hui en fin d'après midi et demain présentation des résultats des projets « Chantiers d'usage »**

- **Les projets d'enrichissement des données**
 - Projets d'**enrichissement** menés à l'INIST
 - **Structuration du plein texte et XMLisation** de documents
 - **Balisage** des principaux champs **des références bibliographiques**
 - **Catégorisation** des documents
 - Détection d'**entités nommées** (LDI Tours)
 - Détection en plein texte de **termes et de leurs variantes** (LINA Nantes)

- Pour mémoire le projet **ISTEX-SNU a développé des Widgets** d'intégration d'ISTEX dans les services numériques des établissements

- **CILLEX : classification et outils de métrologie** greffés sur un moteur de recherche ISTEX (CLLE/IRIT Toulouse)
- **LorExplor : Bibliothèque open source de composants Xml** d'exploitation du corpus ISTEX (Université de Lorraine)
- **ISTEX_R : Caractérisation de l'évolution des recherches et des connaissances** dans le temps grâce à la construction de cartes diachroniques (LORIA, ATILF, INIST)

PRÉSENTATION DES RÉSULTATS DES PROJETS « CHANTIERS D'USAGE »

En fin d'après midi

- **ALPAGE** : annotation de corpus ISTEX et codage en TEI
- **FULLLAB** : Comparaison fulltext et abstract en sciences environnementales

Demain

- **Biosystémique** : Recherche de résultats expérimentaux dans les publications scientifiques concernant la biologie systémique
- **Terre-ISTEX** : Identification et analyse des TERRains d'Études dans les corpus ISTEX

- **3ST** : Surligneur Sémantique de Textes Scientifiques
- **OTELO** : chantier thématique sur les Terres rares, de l'exploration à l'impact environnemental
- **NEOTEX** : Exploitation de documents textuels d'un domaine par un néophyte
- **RelTEX** : Usage du corpus ISTEX pour l'exploitation de méthodes d'extraction de connaissances à partir de textes

- **Valoriser les résultats** des projets « Chantiers d'usage »
 - Nécessité de rendre visibles les résultats obtenus
 - Enrichissement du site web d'ISTEX
- ⇒ **Démontrer l'apport en recherche de la disponibilité des ressources ISTEX**
- **Terminer proprement ISTEX 1**
 - Report de la fin du projet au 31/12/2017
 - Ouverture vers quelques aménagements budgétaires
 - ⇒ Quels besoins supplémentaires pour les 4 derniers mois et pour quels objectifs ?

PRÉPARER LA SUITE : VERS UN ISTEX 2

- **Un projet ISTEX 2 devrait voir le jour** qui devrait amplifier les usages de TDM sur les corpus ISTEX
- Analyser les **possibilités actuelles d'ISTEX mais aussi les limites et difficultés** qu'il conviendrait de dépasser
- Nous donner des **éléments pour préparer ce futur projet ISTEX 2**
 - Quelles nouvelles orientations faudrait-il privilégier ?

- **Respect du timing** : le programme est chargé
 - => veiller à respecter les temps de parole
 - ⇒ préserver les temps de discussions
 - ⇒ **Projets d'enrichissement et de SAV** :
 - 20 mn de présentation et 10mn de discussion
 - **Projets « Chantiers d'usage »** :
 - 40 mn de présentation, 10 mn de discussion
- Nous aurons 1 heure de discussion générale sur les actions de communication à entreprendre à partir des résultats de ces projets et conclure ce séminaire
- Merci de transmettre une copie de vos présentations à Marc.ROUX@cnsr-dir.fr et à Jean-Marie.Pierrel@atilf.fr



<http://www.istex.fr>

<http://api.istex.fr/documentation>



istex@inist.fr



[@istexdev](https://twitter.com/istexdev)

Merci de votre attention.