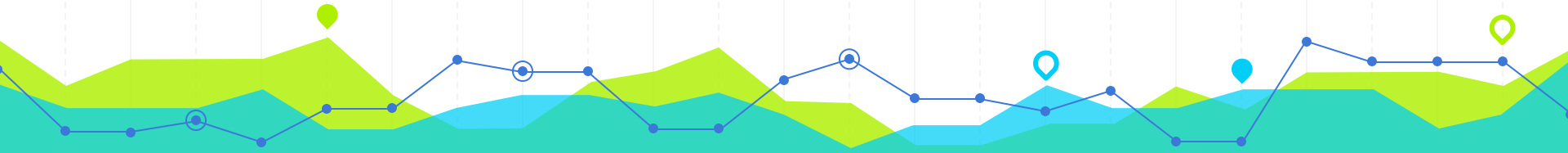


Usage des ressources ISTEX

Ateliers pratiques

cnrs Inist | Institut de l'information scientifique et technique

Programme

1. Présentation générale d'ISTEX
2. Constitution d'un corpus par interrogation d'ISTEX

---------- Une pause ! ----------

3. Téléchargement du corpus avec ISTEX-DL
4. Exploration et valorisation du corpus avec l'outil LODEX
5. Liens utiles
6. Annexes

Documents utiles

Télécharger ce support et les fichiers utiles aux TP

https://formation.lodex.fr

1. *Enregistrer le zip contenant les documents de travail : clic droit, puis" enregistrer la cible du lien sous" → Bureau ou dossier créé*
2. *extraire les 5 fichiers : clic droit, puis "extraire tout" → Bureau ou dossier créé*

Introduction à ISTEX

1

ISTEX
L'excellence documentaire pour tous

ISTEX…

Initiative d'excellence en Information Scientifique et Technique

pour

Construire le socle de la bibliothèque scientifique numérique nationale

ISTEX

L'excellence documentaire pour tous

« Construire le socle de la bibliothèque scientifique numérique nationale. »

est lancé le 30 avril 2012

est né d'une impulsion nationale, dans le cadre du programme « Investissements d'avenir » du Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI), et financé par l'Agence Nationale de la Recherche (ANR)

est réservé aux abonnés de l'Enseignement supérieur et de la recherche française (ESR).

est porté par quatre partenaires : CNRS, ABES, Couperin.org, l'Université de Lorraine pour la Conférence des Présidents d'Université (CPU).
En savoir plus

ANR-10-IDEX-0004-02

ISTEX
L'excellence documentaire pour tous

"Construire le socle de la bibliothèque scientifique numérique nationale."

ISTEX : quels objectifs ?

- Acquisition massive et centralisée d'archives scientifiques
  - collections rétrospectives de la littérature scientifique
  - toutes disciplines
- Construction d'une plateforme nationale d'hébergement et de mise à disposition des données (Inist)

https://www.istex.fr/

ISTEX
L'excellence documentaire pour tous

"Construire le socle de la bibliothèque scientifique numérique nationale."

La plateforme ISTEX : quels services ?

ISTEX
L'excellence documentaire pour tous

"Construire le socle de la bibliothèque scientifique numérique nationale."

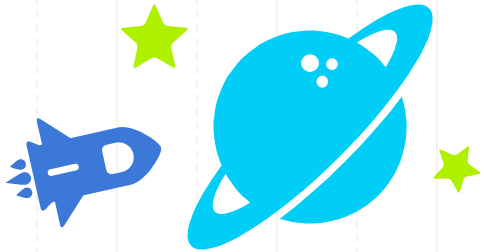
La plateforme ISTEX : quels usages ?

Deux usages : l'un classique, l'autre plus avancé !

- Usage documentaire 
- Usage en fouille de textes (text mining) 

La fouille de textes désigne toute technique de traitement automatisé visant à extraire des connaissances d'un ensemble de textes sous forme numérique, produits par des humains pour des humains. Elle permet d'analyser parallèlement de vastes quantités de données selon un critère de nouveauté ou de similarité, et ainsi de dégager des informations de haute qualité difficiles à appréhender par la simple lecture cursive, telles que des constantes, des tendances et des corrélations.

(Sources : https://cnnumerique.fr/files/2017-10/CNNum_Fiche_TDM.pdf ;
https://fr.wikipedia.org/wiki/Fouille_de_textes ;
https://adbu.fr/competplug/uploads/2016/12/v9-Designed-Exec-Summary-ADBU-8pp_fra_final.pdf)



ISTEX

Quel contenu ? Quelques chiffres...

ISTEX aujourd'hui, c'est :

23 207 399

documents

*Le fonds ISTEX en quelques chiffres**

27
corpus éditeurs

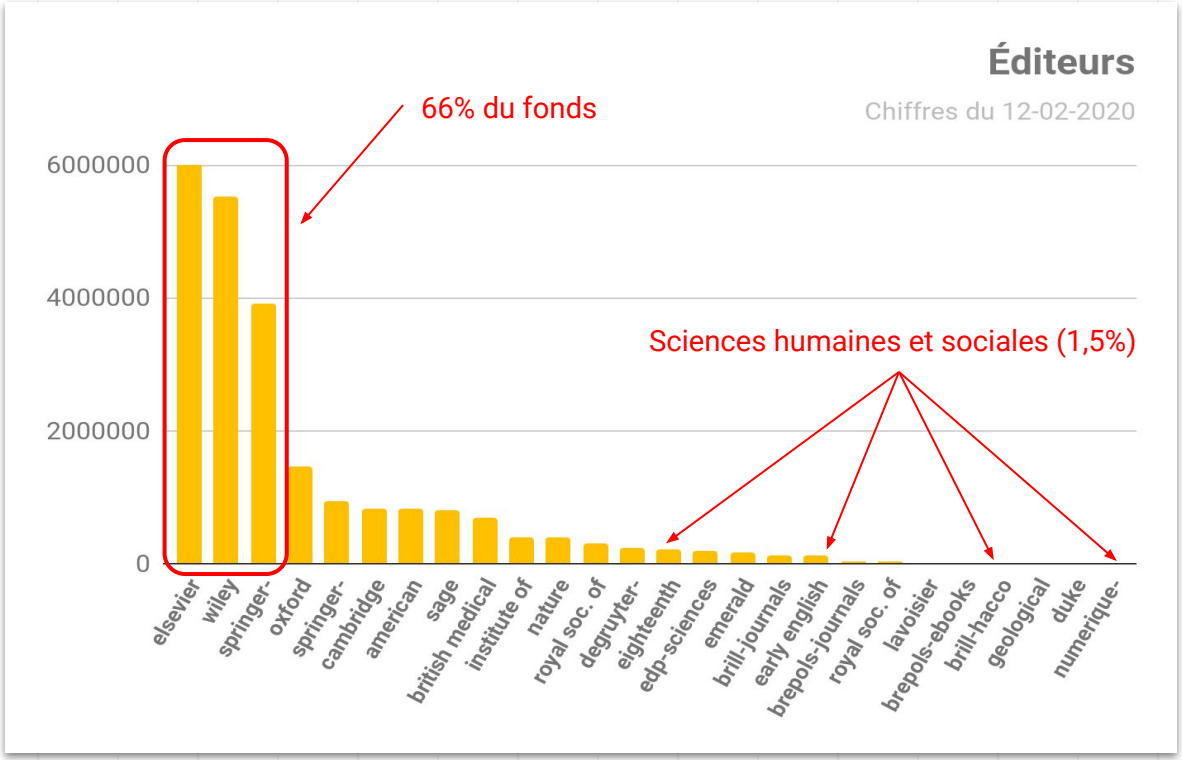
9 279
revues

346 863
monographies

** Chiffres du 10-03-2020*

Le fonds ISTEX > répartition par provenance

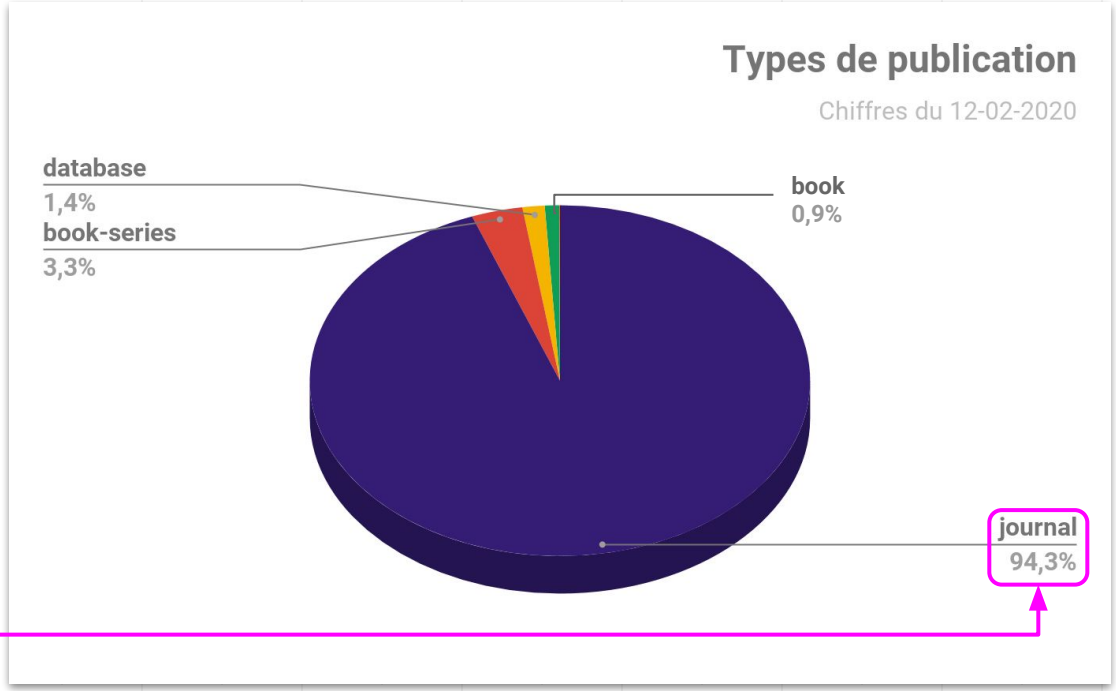
NB : Documents provenant de 27 corpus éditeurs mais chiffres des graphiques datant du 12/02/2020



Le fonds ISTEX > répartition par types de publication

Prépondérance des documents issus de **revues**

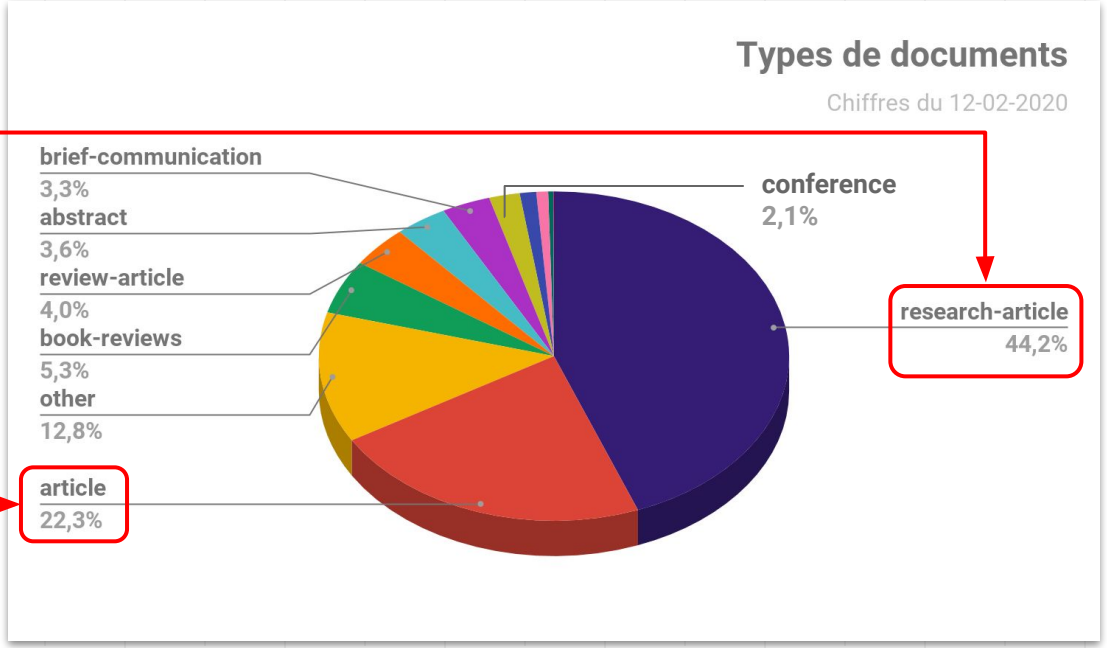
Définitions et chiffres sur data.istex.fr :
Types de publication



Le fonds ISTEX > répartition par types de contenu

Plus de **66%** des documents sont des **articles**

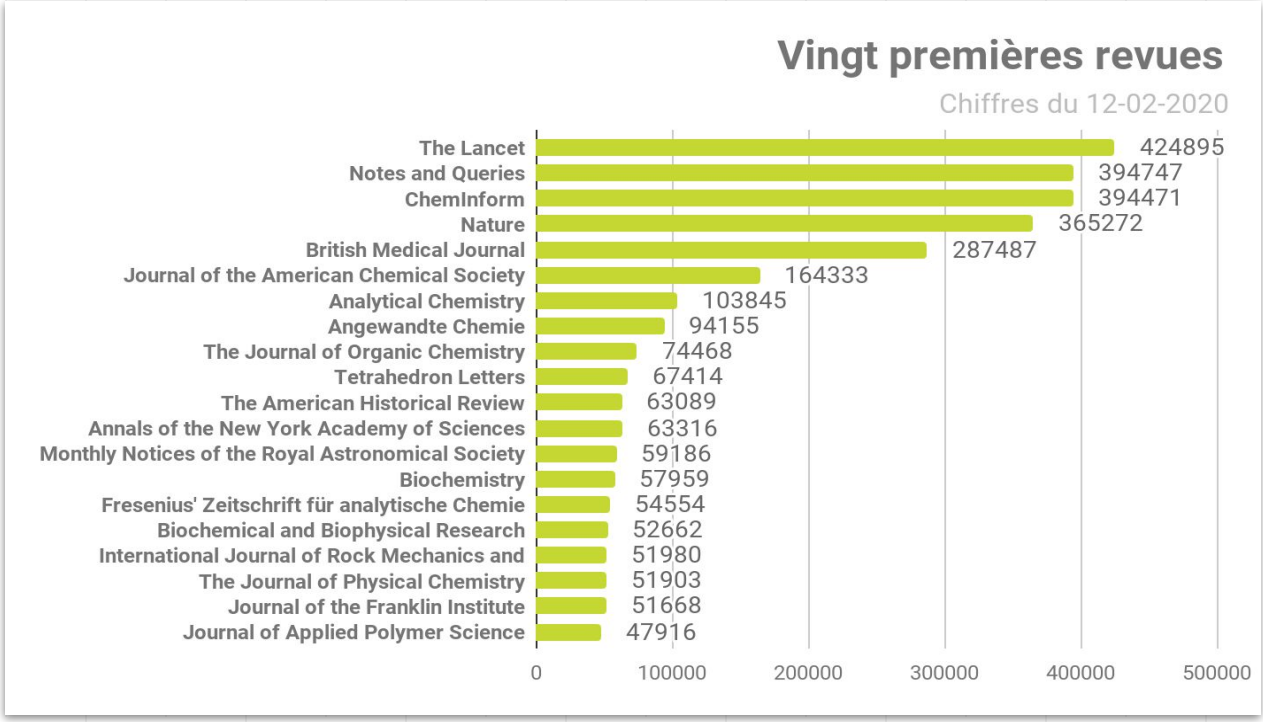
Définitions et chiffres sur data.istex.fr :
Types de contenu



Le fonds ISTEX > répartition par revues

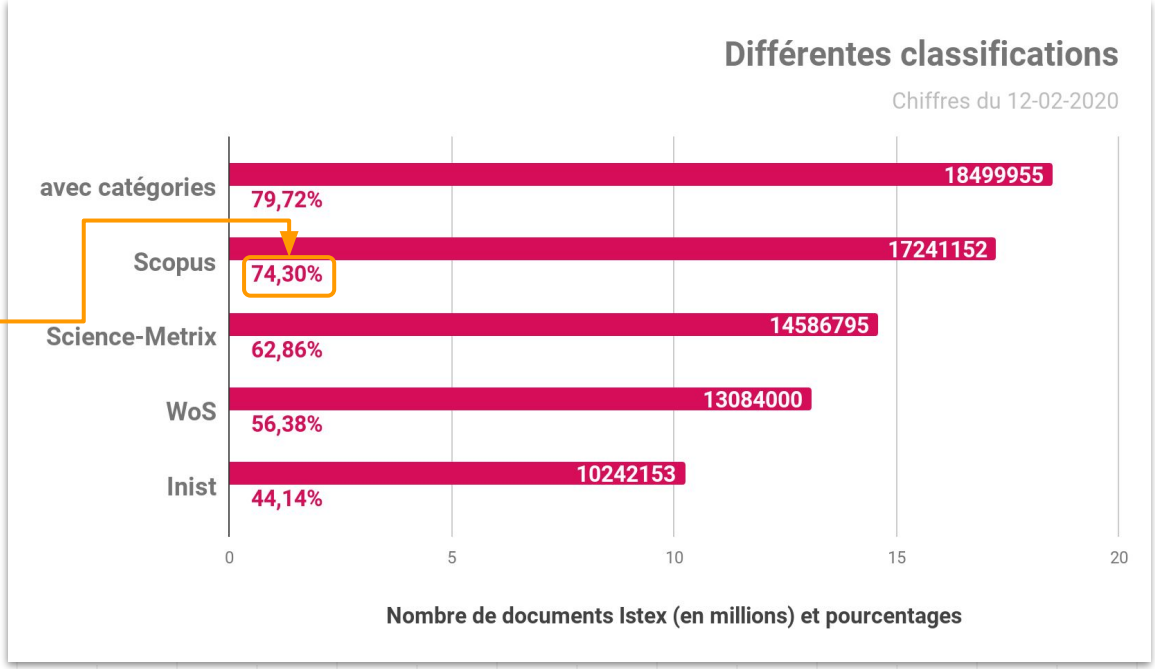
Dans le fonds de plus de **9 000** revues présentes dans ISTEX :

liste des **20** revues les plus importantes en **nombre de documents**



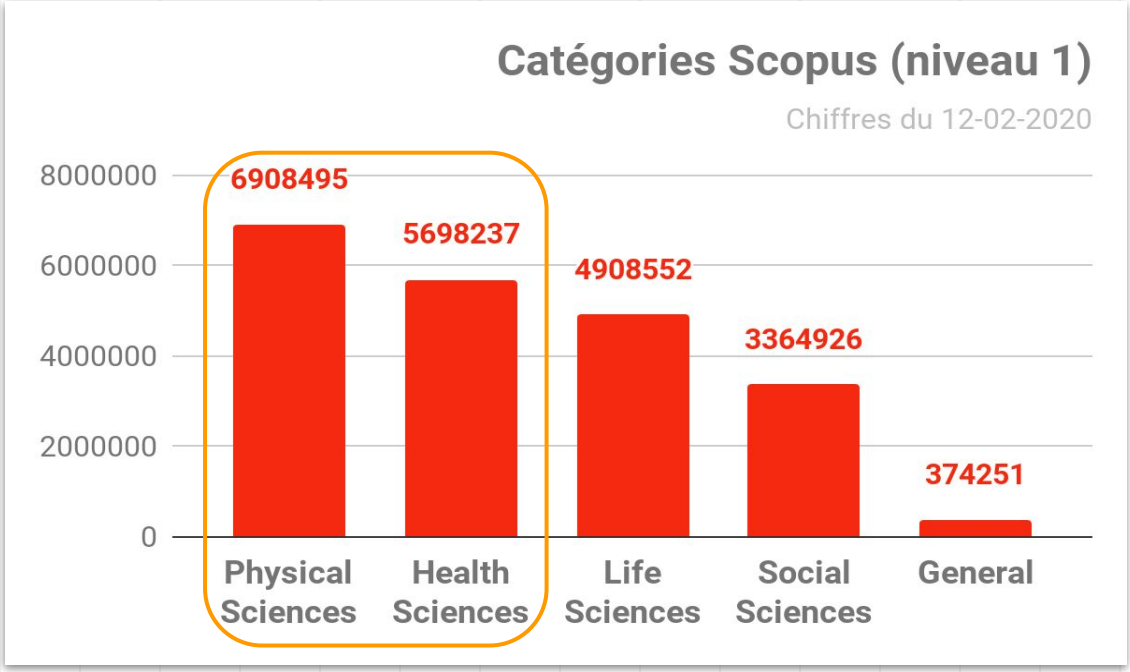
Le fonds ISTEX > répartition par domaines scientifiques - 1

Scopus est la classification la plus représentative des documents Istex, sachant que plus de 20% des documents ne comportent aucune catégorie scientifique



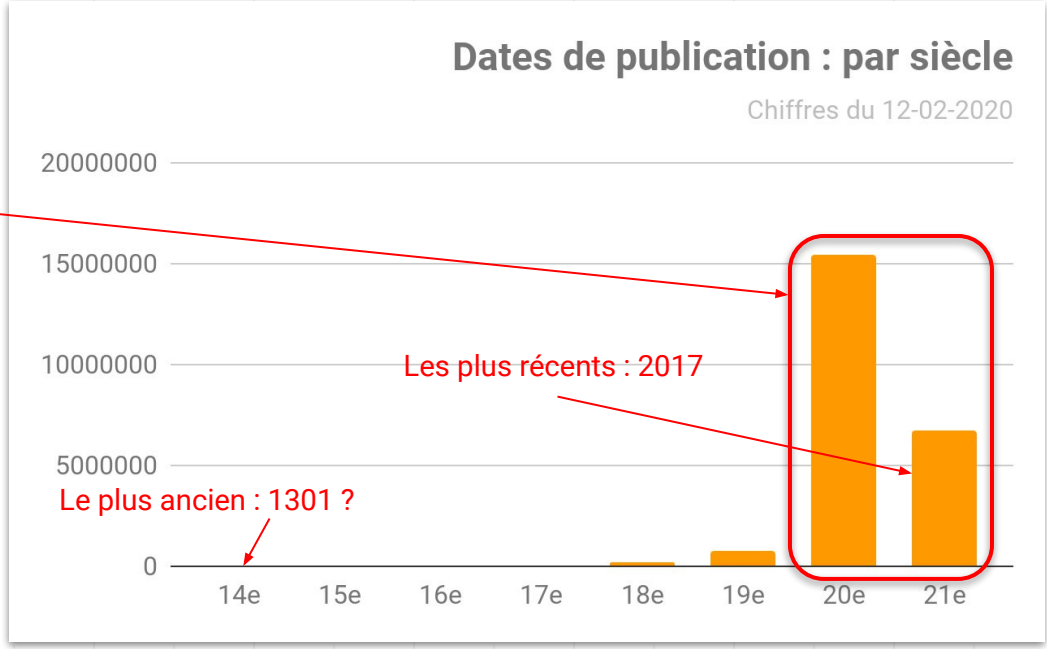
Le fonds ISTEX > répartition par domaines scientifiques - 2

54% des documents Istex sont **classés avec une catégorie Scopus** relative aux **sciences physiques ou de la santé**



Le fonds ISTEX > répartition par dates - 1

95% du fonds a été publié aux **20e et 21e** siècles

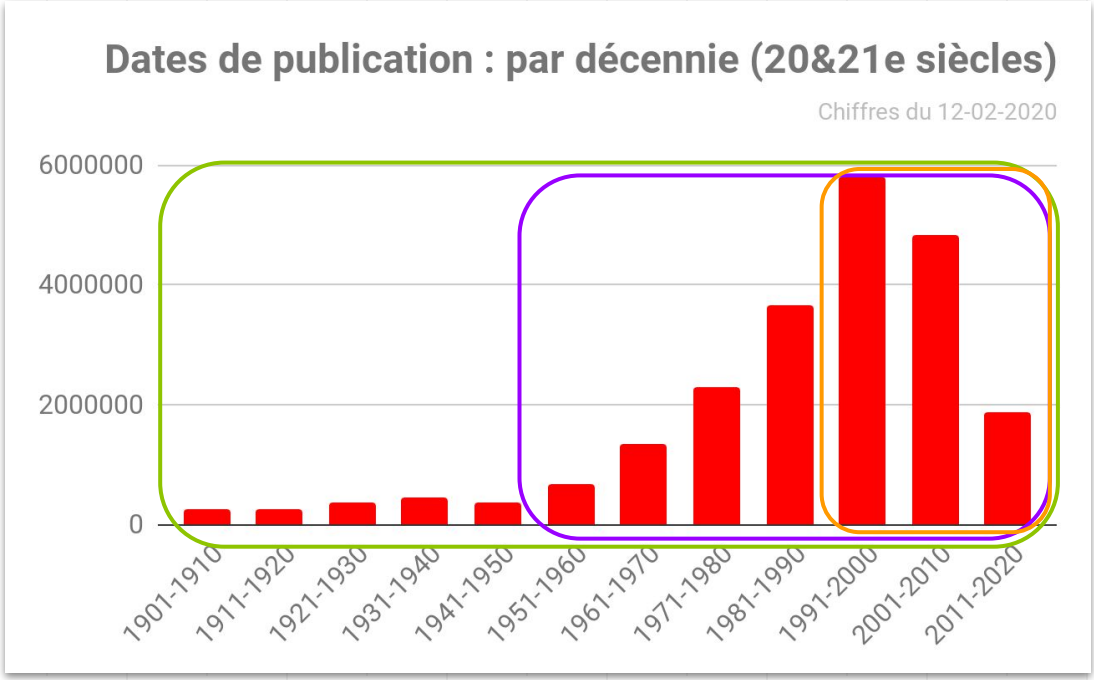


Le fonds ISTEX > répartition par dates - 2

95% des documents publiés entre 1900 et aujourd'hui (2017)

88% des documents publiés depuis 1950

53% des documents publiés sur les 30 dernières années

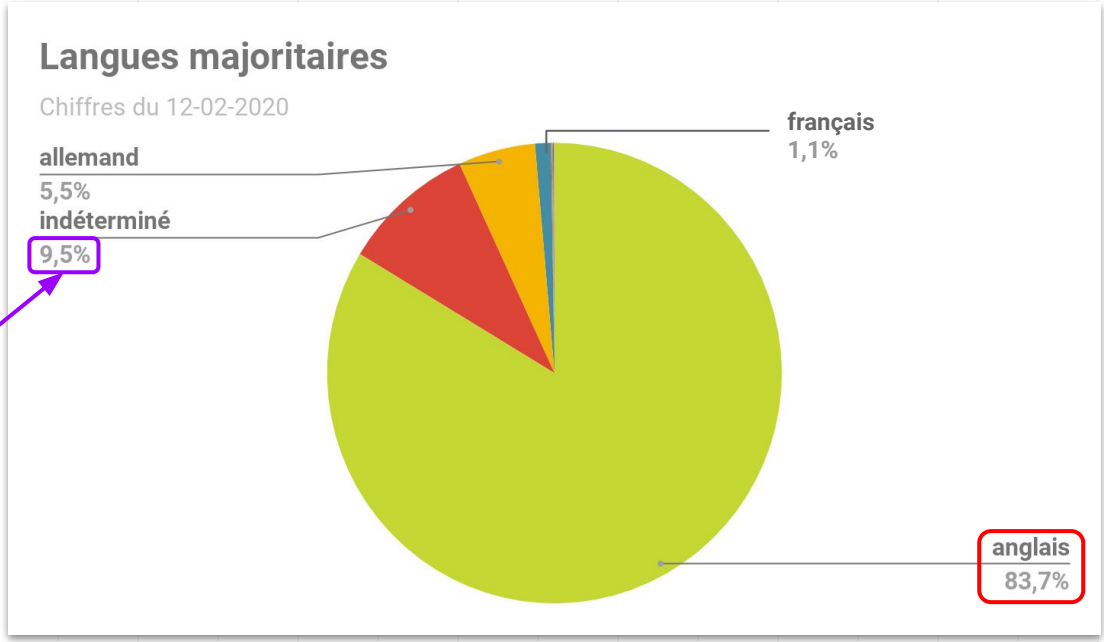


Le fonds ISTEX > répartition par langues - 1

53 langues !

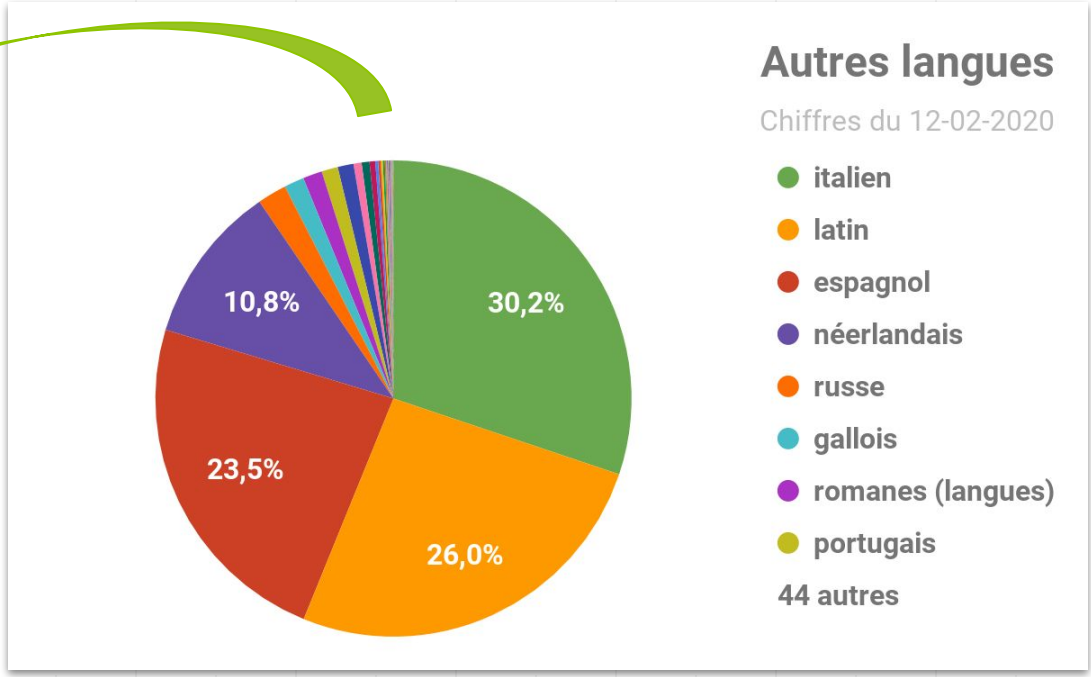
Anglais majoritaire

Information non renseignée par les éditeurs pour près de 2,2 millions de documents !

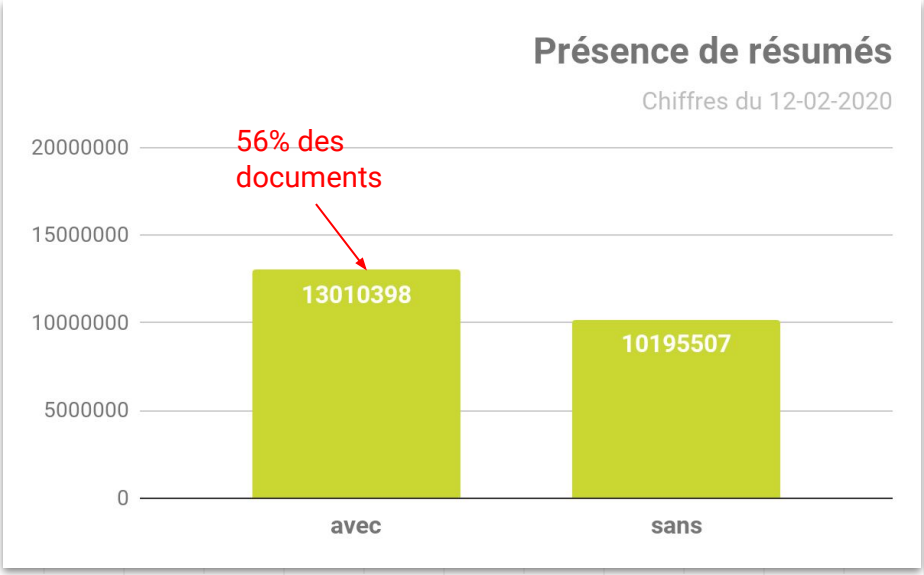


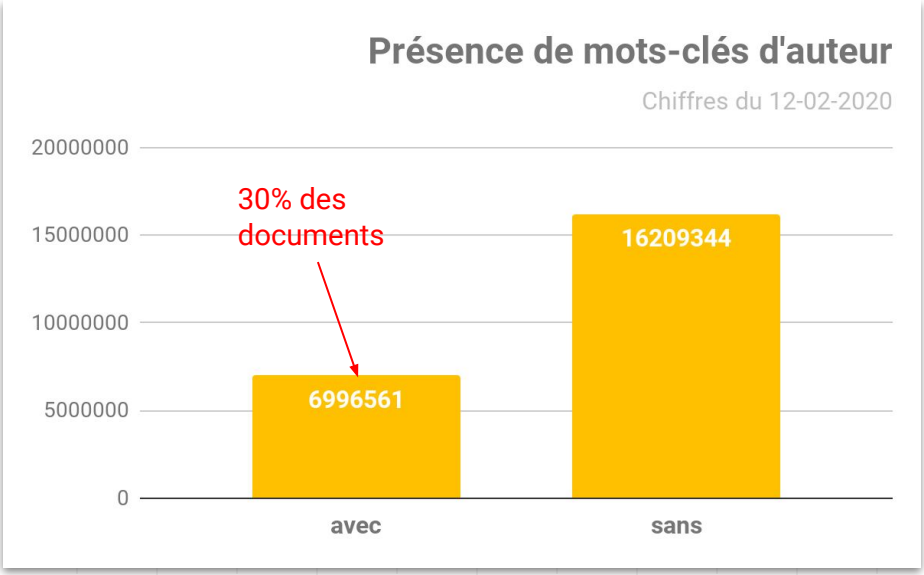
Le fonds ISTEX > répartition par langues - 2

Les **49** autres langues concernent **0,3%** des documents restants (soit près de 57 000 documents)



Le fonds ISTEX > répartition par éléments descriptifs du contenu

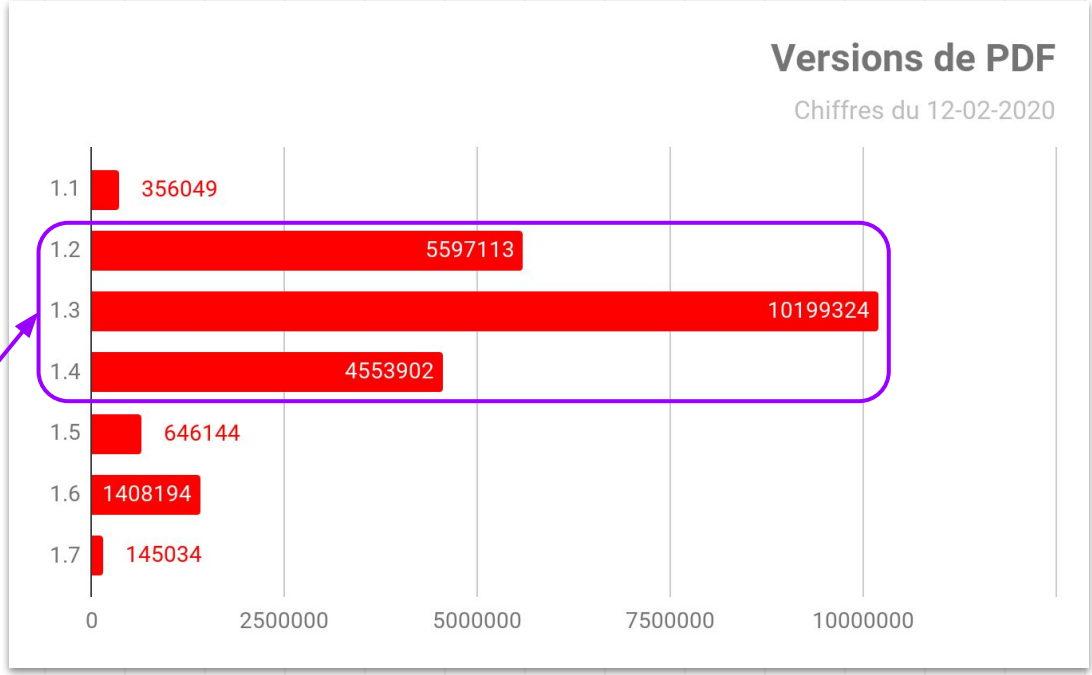




Le fonds ISTEX > répartition par indicateurs - 1

Plus la version de PDF est élevée, meilleure devrait être la qualité du document

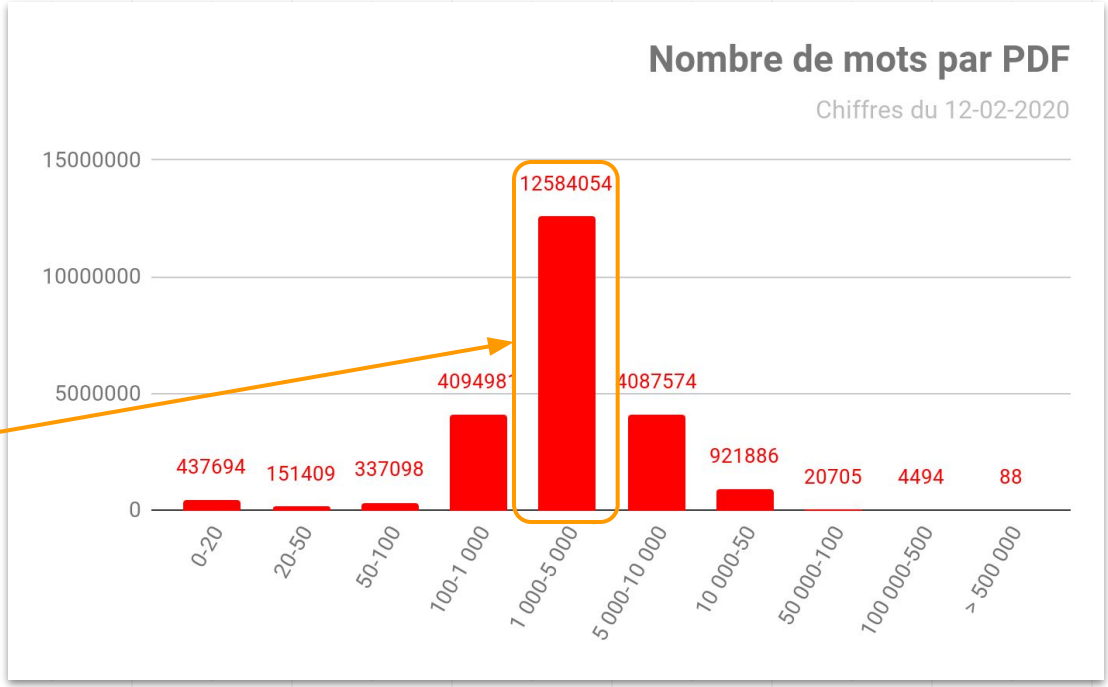
Près de 88% des documents existent dans une version de PDF comprise entre 1.2 et 1.4



Le fonds ISTEX > répartition par indicateurs - 2

Un nombre de mots réduit peut indiquer un PDF image ou bien un court résumé

La majorité des documents PDF (54%) ont un nombre de mots compris entre 1 000 et 5 000

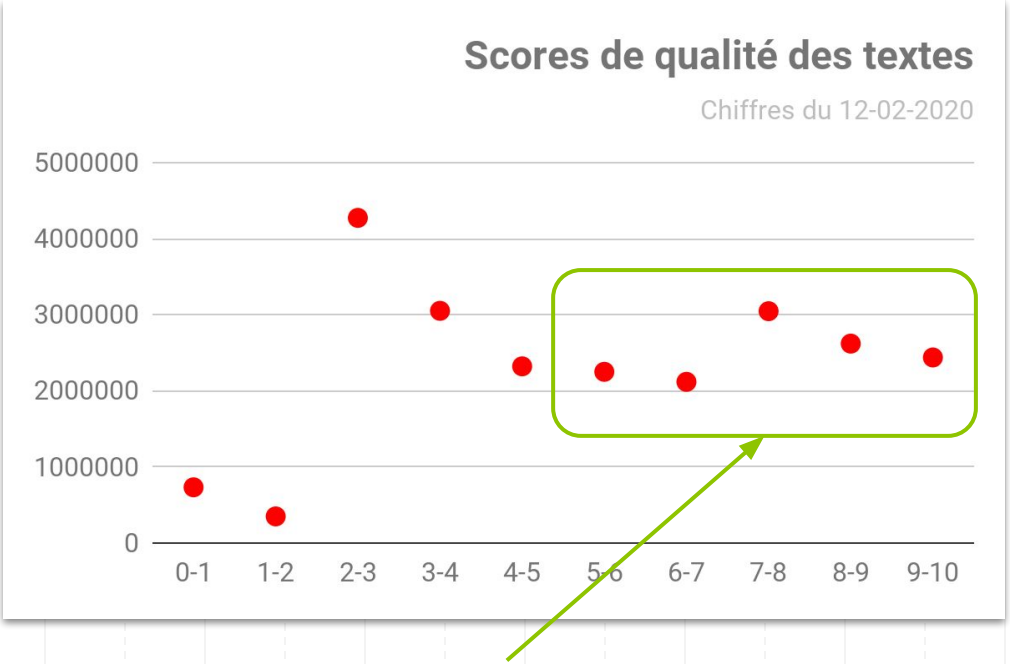


Le fonds ISTEX > répartition par indicateurs - 3

Le score de qualité est un nombre décimal compris entre 0 et 10.

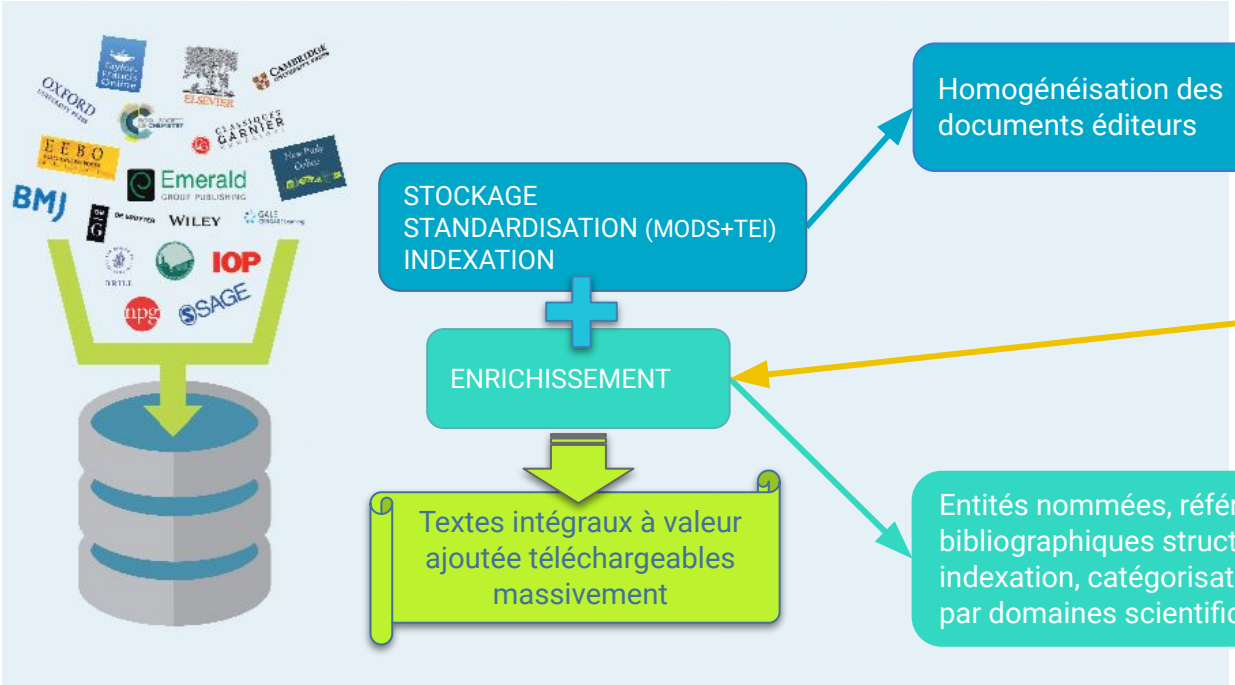
Son mode de calcul tient compte de la version du PDF, du nombre de mots du résumé, ainsi que du nombre de mots du texte intégral (PDF).

Voir la formule dans : Calcul du score de qualité



54 % des documents (12 475 415) ont un score supérieur à 5

La plateforme ISTEX > la chaîne de traitements



La plateforme ISTEX : les enrichissements

Unitex

Détection de 9 types d'entités nommées (lieux géographiques, organismes financeurs, etc.)

enrichment-process.data.istex.fr/unitex

⇨ **68,6%** des docs

Bayésien naïf (nb)

Catégorisation de l'article par apprentissage avec Pascal/Francis

enrichment-process.data.istex.fr/nb

⇨ **43,4%** des docs

Grobid

Structuration des références bibliographiques

enrichment-process.data.istex.fr/grobid

⇨ **58,1%** des docs

Multicat

Catégorisation de la revue par appariement (WoS / Science Metrix / Scopus)

enrichment-process.data.istex.fr/wos

enrichment-process.data.istex.fr/science metrix

enrichment-process.data.istex.fr/scopus

⇨ **75,2%** des docs

Teeft

Indexation des articles par extraction de termes à partir du texte intégral en anglais

enrichment-process.data.istex.fr/teeft

⇨ **72,6%** des docs

Chiffres du 04-02-2020

La plateforme ISTEX : deux usages

Exemples d'usage documentaire

- Accéder à un article
- Rechercher les articles d'un auteur
- Constituer une bibliographie sur un domaine
- Produire une synthèse documentaire

La plateforme ISTEX : deux usages

Exemples d'usage TDM

- Aider au pilotage scientifique & au travail du chercheur
- Répondre à une question scientifique
- Usages hors Recherche

La plateforme ISTEX : exemples d'usage TDM*

Pilotage & Recherche

- Détection de sujets de recherche émergents
- Détection de plagiat
- Recommandation d'articles pour des néophytes ou non

Question scientifique

- Applications inattendues pour un ancien médicament
- Cartographie du génome humain
- Nouveaux matériaux thermoélectriques
- Traitement médiatique des "*Gilets Jaunes*"
- Portrait-robot d'espèces de poissons sauvages domesticables

Autres usages

- Marketing ciblé : détection d'intention d'achats et personnalisation des réponses
- Aide au diagnostic médical
- Prévention de la cybercriminalité
- Gestion des risques de fuite dans les réseaux d'assainissement
- Traduction automatique
- Assistant personnel

** Ces cas d'usage sont documentés dans le fichier "Exemples-usage-TDM.pdf" présent dans le zip "Documents-atelier"*

La plateforme ISTEX : exemples d'usage TDM à l'Inist

Réaliser des analyses scientifiques

- Analyses syntaxiques et terminologiques
- Evolution temporelle de thématiques

Mettre au point ou tester des outils

- **Entity-Fishing/IRC3** : détection de noms d'espèces animales et végétales
- **Unitex** : détection de définitions de concepts scientifiques
- **TermSuite** : détection de termes et de leurs variantes
- **TermSuite/Teeft + Neurodoc :** constitution et analyse de corpus
- **Topic Modeling, Iramuteq** : exploration thématique

Évaluer la performance d'outils

- **Unitex** : Corpus gold annoté manuellement en entités nommées pour évaluer les résultats de détection de l'outil

La plateforme ISTEX : exemples d'usage TDM à l'Inist

Corpus réalisés à l'Inist :

- Sur des **domaines** et des **thématiques** variés :
  - Astrophysique, orthophonie, botanique, zoologie, géosciences, sciences et techniques alimentaires
  - Zone polaire arctique, vieillissement

- De petits et de gros corpus :
  - De **34** à **52 238** documents

Accessibles sur data.istex.fr

Interrogation d’ISTEX 2

Interroger ISTEX > Comprendre la structuration du réservoir

Le réservoir ISTEX contient des **objets documentaires**

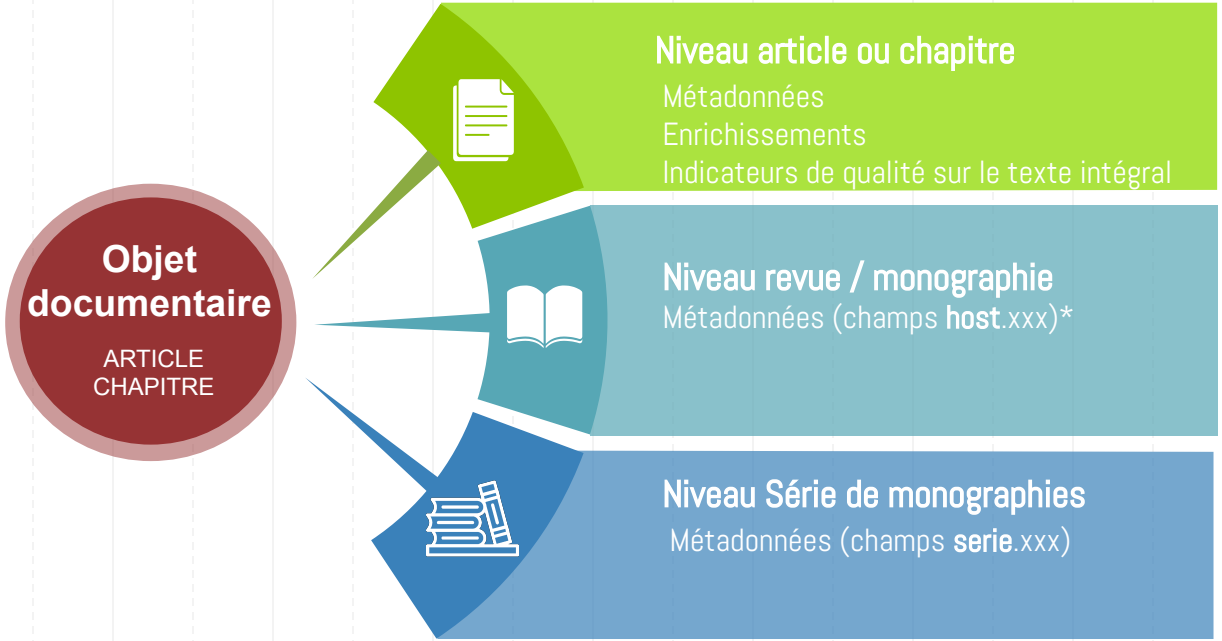
Ces objets documentaires sont :

- Soit des **articles** (de revue)
- Soit des **chapitres** (de monographie, d'e-book, de monographies en série)

Différents niveaux de granularité

L'interrogation peut se faire sur des informations concernant :

- Le niveau article ou chapitre
- Le niveau monographie
- Le niveau série de monographies



** Ces informations seront identiques pour **tous les articles d'une même revue** ou pour **tous les chapitres d'une même monographie***

Où interroger ISTEX ?

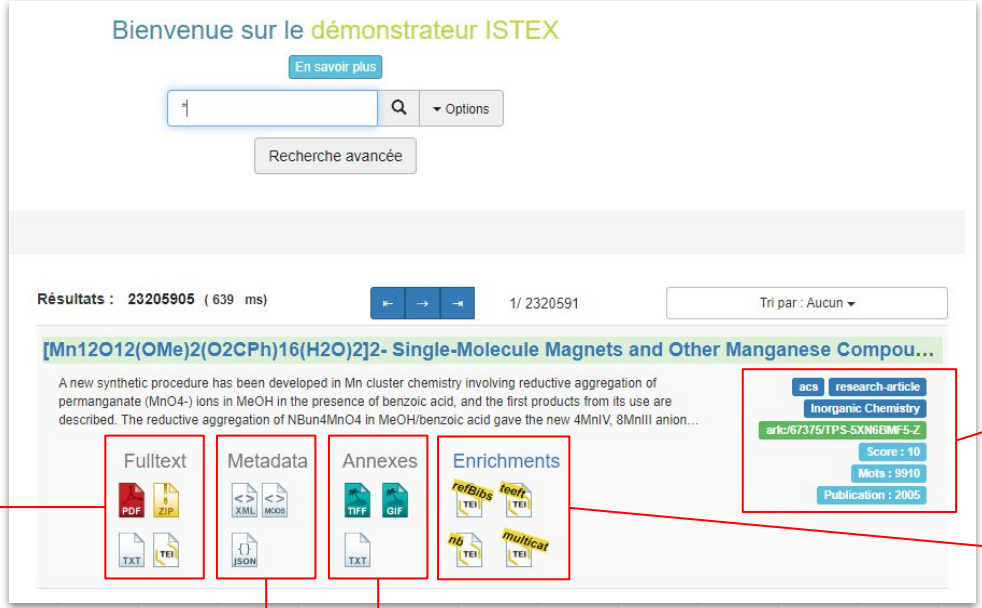
Le démonstrateur est une interface à **vocation pédagogique** branchée sur l'API ISTEX qui permet de :

- construire sa requête (en mode simple ou en mode avancé)
- visualiser et filtrer les résultats

https://demo.istex.fr

Interroger ISTEX > le démonstrateur

Les formats disponibles pour le texte intégral, les métadonnées décrivant le document et les annexes/couvertures

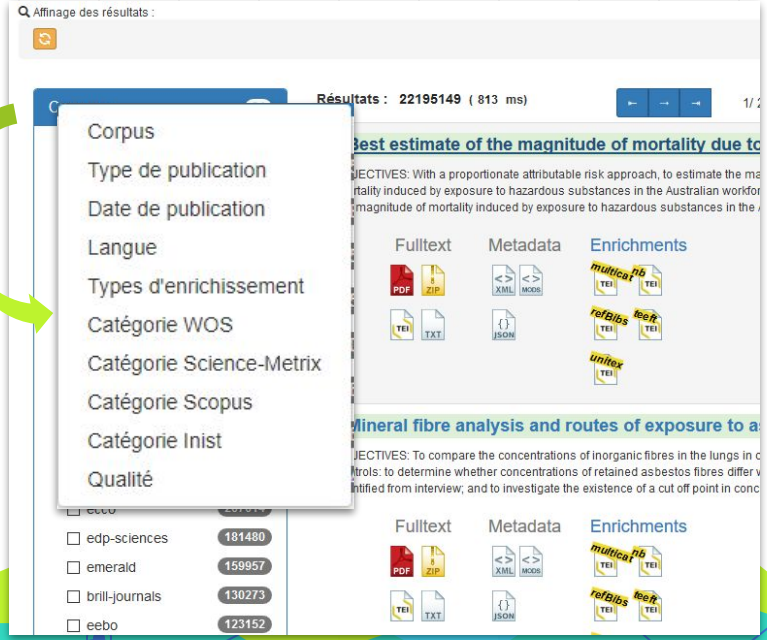


Accès rapide à différentes infos bibliographiques du document

Les différents types d'enrichissements disponibles en TEI

Interroger ISTEX > le démonstrateur

Facettes pré-définies dans l'interface

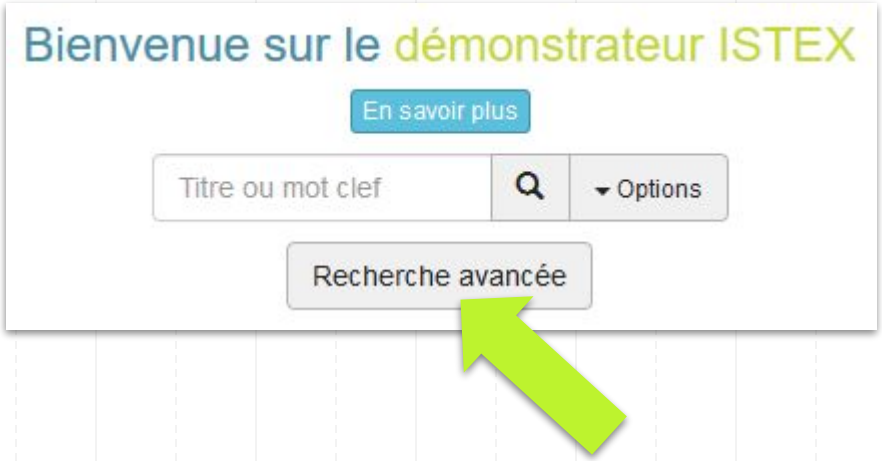


→ donne une vision synthétique du corpus

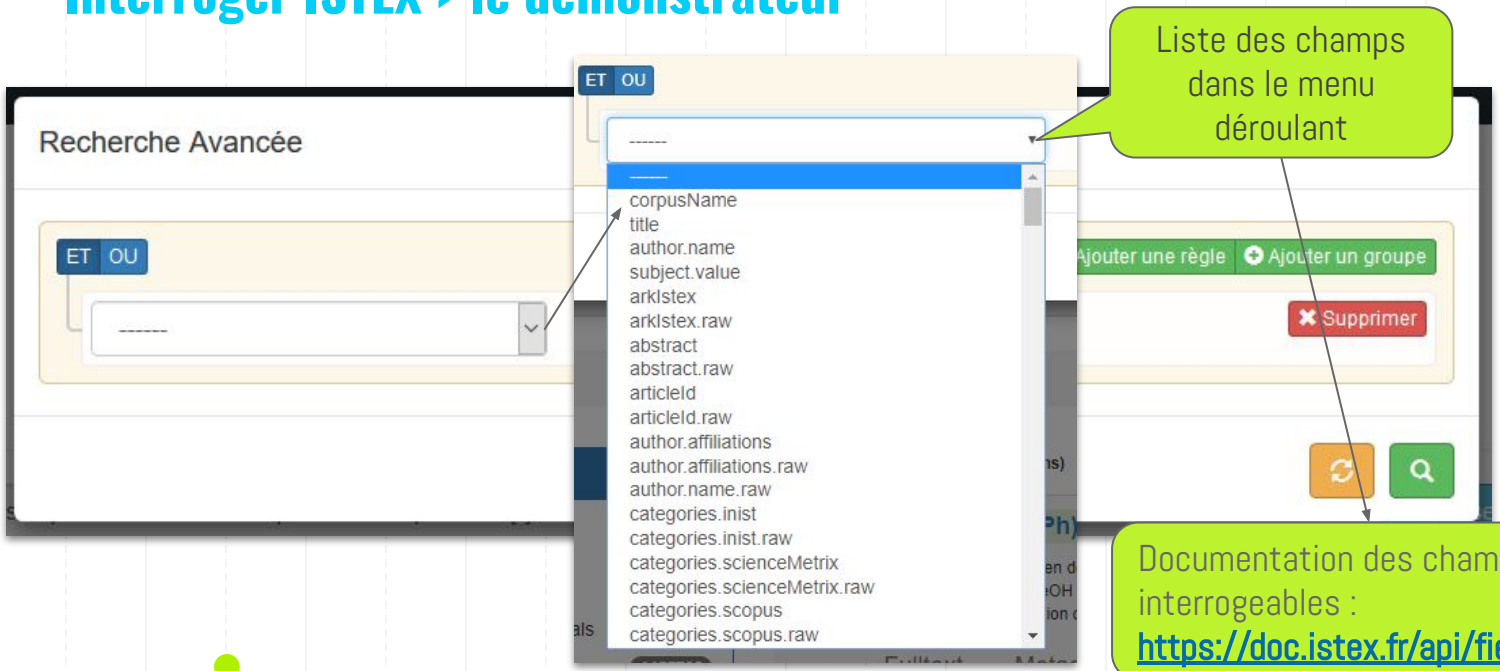
→ permet de filtrer les résultats de la requête

→ mais possibilités limitées - exploratoire

Interroger ISTEX > le démonstrateur



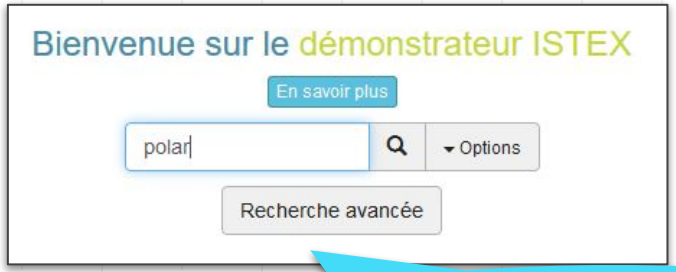
Interroger ISTEX > le démonstrateur



Interroger ISTEX > la recherche basique

Petit exercice pour démarrer :

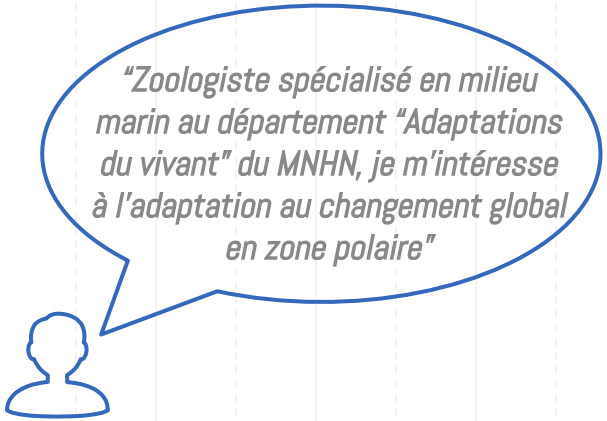
- *On souhaite rechercher les documents possédant le terme "polar"*



Recherche sur :
- *Les **métadonnées***
- *Le texte **intégral***
- *Les **références** bibliographiques*
- *Les **enrichissements***

Résultats : + de 1 000 000 docs !!!

Cas d'usage



Objectif pédagogique :

Écrire une équation, testée pas à pas, utilisant un certain nombre d'opérateurs, d'astuces et de syntaxes, associés à une recherche sur un certain nombre de champs utiles pour délimiter un corpus pertinent et de taille raisonnable

Interroger ISTEX > Cas d'usage

Application TDM

- Outil de reconnaissance d'**entités nommées** scientifiques
- **Détecter** les **espèces animales**
- Dans le texte intégral

Thématique

- zoologie : **espèces animales** de type **poisson ou mollusque**
- zone géographique : zone **polaire** ou **subpolaire arctique**

Contraintes liées à l'application

- L'outil traite les documents en anglais
- Les documents devront être en format PDF
- Attention : l'outil ne traite pas les PDF image

*Rédiger sa requête dans un bloc-note type "NotePad". Ou copier successivement les exemples de "**TP.txt**", et coller les requêtes dans le démonstrateur*

Interroger ISTEX > Principes de base

1. je recherche les termes suivants
   - fish, mollusc
   - arctic, subarctic, Svalbard, Barents sea

2. dans les champs
   - titre
   - résumé
   - mots-clés d'auteurs

3. je limite mon corpus aux documents
   - langue de publication = anglais

Notions abordées

- **Opérateurs booléens**
- **Parenthésage**
- **Recherche sur champs**
- **Recherche insensible à la casse**
- **Expressions multitermes / mots-composés**

Interroger ISTEX > Principes de base

Equation étape 1

Recherche sur des termes de zoologie

```
fish OR mollusc
```

Résultats (10 mars 2020) : 800 717 docs

Explications :

- **OR** cumule les résultats associés à “fish” et “mollusc”
- Les mots sont recherchés sur tout le document
- Si pas d’opérateur utilisé, l’opérateur par défaut OR s’applique

*Les opérateurs doivent s’écrire **en MAJUSCULES***

Interroger ISTEX > Principes de base

Equation étape 1

Recherche sur des termes de zoologie et de géographie

```
( fish OR mollusc ) AND

( arctic OR subarctic )
```

Résultats (10 mars 2020) : 50 614 docs

Explications :

- OR cumule les résultats associés à “fish” et “mollusc”
- Les mots sont recherchés sur tout le document
- Si pas d’opérateur utilisé, l’opérateur par défaut OR s’applique
- **AND + parenthèses** permet d’associer 2 critères, ici termes zoologiques et termes géographiques

Interroger ISTEX > Principes de base

Equation étape 1

Recherche sur des termes de zoologie et de géographie contenant **majuscule**

```
( fish OR mollusc ) AND

( arctic OR subarctic OR svalbard )
```

Résultats (10 mars 2020) : 51 100 docs

Explications :

- OR cumule les résultats associés à “fish” et “mollusc”
- Les mots sont recherchés sur tout le document
- Si pas d’opérateur utilisé, l’opérateur par défaut OR s’applique
- AND + parenthèses permet d’associer 2 critères, ici termes zoologiques et termes géographiques
- **insensibilité à la casse**
  - **"svalbard"** cherche aussi **"Svalbard"**

Interroger ISTEX > Principes de base

Equation étape 1

Recherche sur des termes de zoologie et de géographie contenant **majuscule, expression multiterme**

```
( fish OR mollusc ) AND

( arctic OR subarctic OR svalbard OR
"barents sea" )
```

Résultats (10 mars 2020) : 52 885 docs

Explications :

- OR cumule les résultats associés à “fish” et “mollusc”
- Les mots sont recherchés sur tout le document
- Si pas d’opérateur utilisé, l’opérateur par défaut OR s’applique
- AND + parenthèses permet d’associer 2 critères, ici termes zoologiques et termes géographiques
- insensibilité à la casse
  - "svalbard" cherche aussi "Svalbard"
- utilisation d’une **expression multiterme** entre guillemets
  - **barents sea** = 1 346 725 docs
  - **"barents sea"** = 12 752 docs

Interroger ISTEX > Principes de base

Equation étape 1

Recherche sur des termes de zoologie et de géographie contenant **majuscule, expression multiterme, mot-composé**

```
( fish OR mollusc ) AND

( arctic OR subarctic OR "sub-arctic"
 OR svalbard OR "barents sea")
```

Résultats (10 mars 2020) : 52 885 docs

Explications :

- OR cumule les résultats associés à “fish” et “mollusc”
- Les mots sont recherchés sur tout le document
- Si pas d’opérateur utilisé, l’opérateur par défaut OR s’applique
- AND + parenthèses permet d’associer 2 critères, ici termes zoologiques et termes géographiques
- insensibilité à la casse
  - "svalbard" cherche aussi "Svalbard"
- utilisation d’une expression multiterme entre guillemets
  - barents sea = 1 347 835 docs
  - "barents sea" = 12 778 docs
- utilisation d’un **mot-composé** entre guillemets
  - **sub-arctic** = 2 116 617 docs
  - **"sub-arctic"** = 7 808 docs

Interroger ISTEX > Principes de base

Equation étape 1

Recherche sur des termes de zoologie et de géographie contenant **majuscule, expression multiterme, mot-composé**

```
( fish OR mollusc ) AND

( arctic OR subarctic OR "sub-arctic"
OR svalbard OR "barents sea" )
```

Langues autres que anglais : **sensibilité aux diacritiques**
Ex : “mammifère” ne cherche pas “mammifere”

Résultats (10 mars 2020) : 52 885 docs

Explications :

- OR cumule les résultats associés à “fish” et “mollusc”
- Les mots sont recherchés sur tout le document
- Si pas d’opérateur utilisé, l’opérateur par défaut OR s’applique
- AND + parenthèses permet d’associer 2 critères, ici termes zoologiques et termes géographiques
- insensibilité à la casse
  - "svalbard" cherche aussi "Svalbard"
- utilisation d’une expression multiterme entre guillemets
  - barents sea = 1 347 835 docs
  - "barents sea" = 12 778 docs
- utilisation d’un mot-composé entre guillemets
  - sub-arctic = 2 116 617 docs
  - "sub-arctic" = 7 808 docs

Plus de détails : opérateurs / astuces

Interroger ISTEX > Principes de base

Equation étape 2

Recherche restreinte sur certains champs

```
title:((fish OR mollusc) AND (arctic OR subarctic OR "sub-arctic" OR svalbard OR "barents sea"))
OR abstract:(idem)
OR subject.value:(idem)
```

Résultats (10 mars 2020) : 52 885 docs ➢ 1 243 docs

Explications :

- Recherche restreinte sur les champs :
  - titre de l'article : "**title**"
  - résumé : "**abstract**"
  - mots-clés d'auteur : "**subject.value**"
- Les noms de champs sont introduits par **:**
- Ajout de **parenthèses** pour regrouper les valeurs recherchées dans un même champ
- **Pas de factorisation** des noms de champs. Il faut répéter les termes pour chaque champ interrogé

Plus de détails : exemples de contenus / recherche sur champs

Interroger ISTEX > Principes de base

Equation étape 3

Restriction aux documents publiés **en anglais**

```
(idem) AND language:eng
```

Résultats (10 mars 2020) : 1 243 docs ➢ 1 133 docs

Explications :

- Ajout de **parenthèses** à l'ensemble de l'équation car on ajoute un critère qui porte sur toute l'équation précédente
- Recherche sur le champ : **language**

  Liste des codes langues

Interroger ISTEX > Principes de base

Équation complète à cette étape

```
(title:((fish OR mollusc) AND (arctic OR subarctic OR svalbard OR "sub-arctic" OR "barents sea"))
OR abstract:((fish OR mollusc) AND (arctic OR subarctic OR svalbard OR "sub-arctic" OR "barents sea"))
OR subject.value:((fish OR mollusc) AND (arctic OR subarctic OR svalbard OR "sub-arctic" OR "barents sea")))
AND language:eng
```

Résultats (10 mars 2020) : 1 133 docs

Interroger ISTEX > Niveau Intermédiaire - affiner les valeurs

4. je limite mon corpus aux documents
   - qui ne sont pas des PDF image

5. je cherche à ajouter un maximum de variantes pertinentes aux termes suivants
   - fish, mollusc
   - svalbard

Les variantes peuvent être de plusieurs natures :

- *formes au **pluriel***
- *formes **non accentuées***
- *variantes **orthographiques***
- *formes **composées***

Notions abordées

Interrogation sur des intervalles de nombres

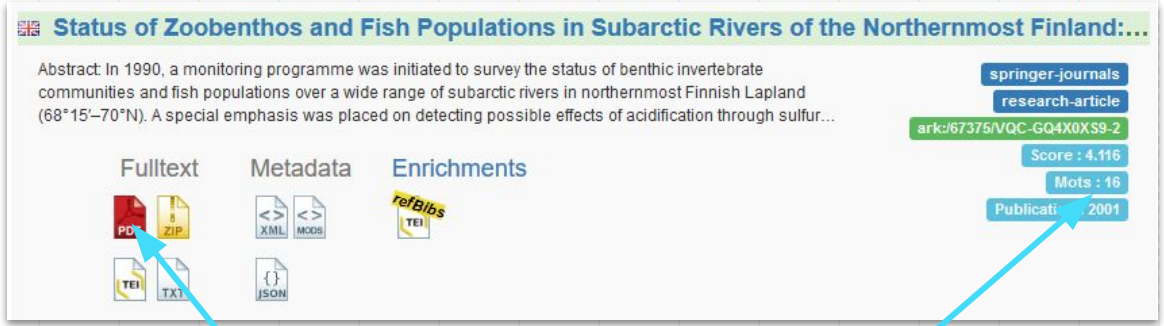
Utilisation de troncatures

Utilisation d'expressions régulières

Interroger ISTEX > Niveau Intermédiaire - affiner les valeurs

Equation étape 4

Un petit test pour commencer : Exploration de la facette “Qualité” dans le démonstrateur



3 - Ouverture du PDF et sélection du texte

2 - Vérification du nombre de mots

1 - Réglage du curseur à 500

Interroger ISTEX > Niveau Intermédiaire - affiner les valeurs

Equation étape 4

Recherche restreinte sur certains champs avec des **intervalles de valeurs**

```
qualityIndicators.pdfWordCount:[500 TO 10000]
qualityIndicators.pdfVersion:[1.2 TO *]
```

Explications :

Recherche sur des intervalles de valeurs à l'aide de crochets **[ ]**

[500 TO 10000] ⟵ valeurs limites inférieure et supérieure

[1.2 TO *] ⟵ valeur limite supérieure infinie

*TO s'écrit obligatoirement **en MAJUSCULES***

Plus de détails : intervalles

Interroger ISTEX > Niveau Intermédiaire - affiner les valeurs

Equation étape 5

Recherche sur des formes au pluriel et des formes non accentuées à l'aide de **troncatures**

```
fish* OR mollus*
```

Explications :

- **?** remplace **1** caractère
- ***** remplace **0 à n** caractère(s)

mollus* = mollusque(s) mollusca mollusc(s) mollusk(s) molluscan...

fish* = fish fishes fishing fishery fisheries fisherman fishermen...
et Tully-Fisher ! ⚠️ **variante hors domaine**

Plus de détails : troncatures

Interroger ISTEX > Niveau Intermédiaire - affiner les valeurs

Equation étape 5 (suite)

Recherche plus ciblée sur des formes au pluriel à l'aide d'**expressions régulières**

```
/fish(es)?/
```

sensibilité à la casse :
aucune majuscule entre les délimiteurs //

Explications :

La troncature par * ne convient pas ici car ramène des variantes hors domaine

Les expressions régulières permettent de rendre compte de toutes les possibilités d'écriture d'un terme en ciblant des variantes spécifiques

fish OR fish**es** = /fish**(es)?**/

terminaison "es" optionnelle

Plus de détails : expressions régulières

Interroger ISTEX > Niveau Intermédiaire - affiner les valeurs

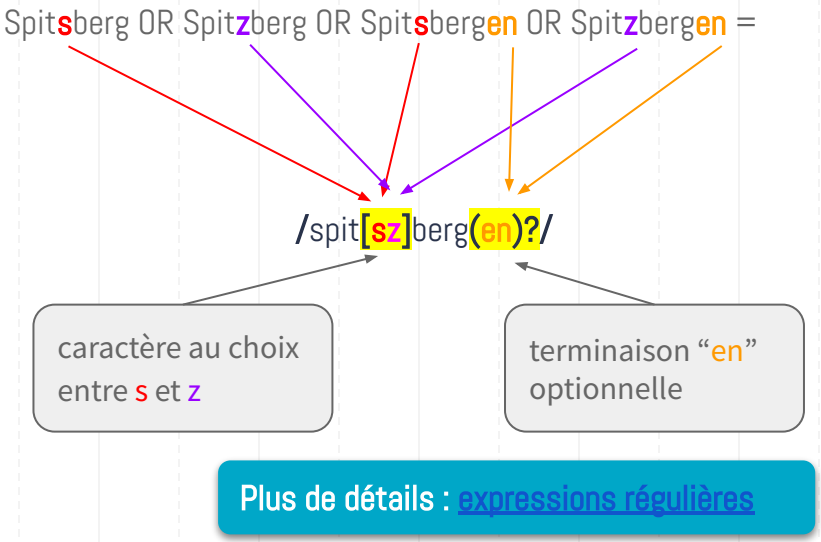
Equation étape 5 (suite)

On peut aller beaucoup plus loin avec la recherche sur des termes avec variantes orthographiques

Exemple : Spitzberg, île principale de l'archipel du Svalbard, s'écrit avec 4 orthographes possibles

```
svalbard OR /spit[sz]berg(en)?/
```

Explications :



Interroger ISTEX > Niveau Intermédiaire - affiner les valeurs

Équation complète à cette étape

```
(title:((/fish(es)?/ OR mollus*) AND (arctic OR subarctic OR "sub-arctic" OR svalbard OR
/spit[sz]berg(en)?/ OR "barents sea"))

OR abstract:((/fish(es)?/ OR mollus*) AND (arctic OR subarctic OR "sub-arctic" OR svalbard OR
/spit[sz]berg(en)?/ OR "barents sea"))

OR subject.value:((/fish(es)?/ OR mollus*) AND (arctic OR subarctic OR "sub-arctic" OR svalbard OR
/spit[sz]berg(en)?/ OR "barents sea")))

AND language:(eng)

AND qualityIndicators.pdfWordCount:[500 TO 10000]
```

Résultats (10 mars 2020) : 1 133 ➤ 1 192 docs

Interroger ISTEX > pour les plus avancés

6. je cherche à ajouter des variantes supplémentaires aux termes suivants :
   - subarctic, barents sea

Notions abordées

Recherche floue

Recherche de proximité

Interroger ISTEX > pour les plus avancés

Equation étape 6

Recherche sur des variantes d'écriture du terme "subarctic" à l'aide de la **recherche floue** :

```
"sub-arctic" OR subarctic~1
```

*- permet de **récupérer** des articles **pertinents** bien que contenant des erreurs (numérisation)*
*- à utiliser **avec prudence** après **vérification** des résultats*

Explications :

en fonction du nombre entier indiqué après le tilde, compris entre **0 et 2**, on recherche un terme ayant **au maximum 2 caractères de différence** (en plus, en moins ou dissemblables).

- subarctic~0 récupère : subarctic
- subarctic~1 récupère : subarctic + subacrtic, subartic, subaratic, subarctica
- subarctic~2 récupère : subarctic, subacrtic, subartic, subaratic, subarctica + subaratica **mais aussi** subarcsec, subaortic

NB : subarctic~ (sans chiffre précisé) équivaut à subarctic~2

Plus de détails : recherche floue

Interroger ISTEX > pour les plus avancés

Equation étape 6 (suite)

Recherche sur des synonymes de “barents sea” à l'aide d'**opérateurs de proximité**

```
"barents sea" OR "barents seas"~2
```

Explications :

en fonction du nombre entier indiqué après le tilde, compris entre **0 et n**, on recherche une expression comportant deux termes **plus ou moins distants** (distance de **n termes**).

"barents seas"~2 récupère :

- Barents-Kara Seas
- Barents and Kara seas
- Barents, and Kara seas

Le moteur de recherche ne tient pas compte de l'ordre des termes et recherche les termes dans les 2 sens

NB : pas de troncature possible

Plus de détails : recherche de proximité

Interroger ISTEX > pour les plus avancés

Équation complète

```
(title:((/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR
/spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2))

OR abstract:((/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR
/spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2))

OR subject.value:((/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR
/spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2)))

AND language:eng

AND qualityIndicators.pdfWordCount:[500 TO 10000]
```

Résultats (10 mars 2020) : 1 192 ➢ 1 197 docs

Interroger ISTEX > Et après ?

Téléchargement du corpus avec ISTEX-DL

Exploration de l'interface

Réutilisation de la requête constituée

Sélection du format de fichier en fonction du cas d'usage

Exploration du corpus dans LODEX

Import du fichier dans Lodex

Modélisation à l'aide du modèle prédéfini

Exploration des différents graphiques

Téléchargement de corpus

3

ISTEX-DL
& Autres outils

“

ISTEX-DL ou ISTEX-DownLoad : “télécharger un corpus ISTEX en quelques clics”

ISTEX-DL : Objectif

Interface web conviviale de type formulaire pour **extraire un corpus ISTEX** en **3 étapes** et récupérer son corpus sur son poste de travail dans un **format compressé (zip)**...avec un **minimum** de connaissance informatique !

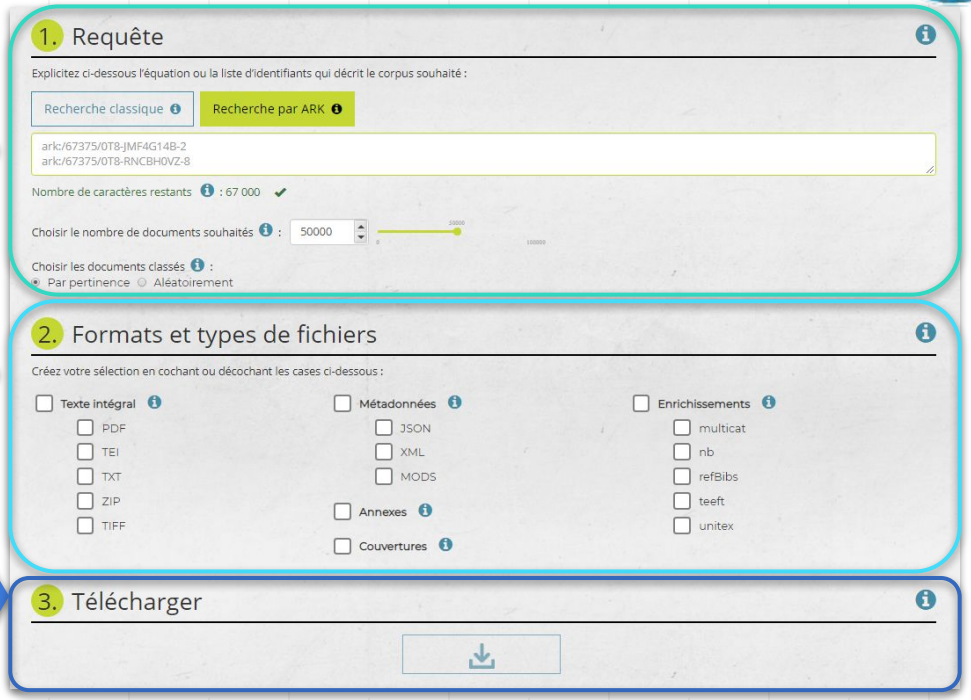
https://dl.istex.fr

TROIS ÉTAPES,...

construction d'une équation

sélection des fichiers

téléchargement du corpus



ISTEX-DL

...CINQ BOUTONS

On gomme...

Réinitialiser

Récupérer

On ne perd pas sa requête en cours...

On partage son corpus...

On teste...

Exemples

Partager

Historique

On peut rejouer le passé...

ISTEX-DL

Deux modes de recherche

ISTEX-DL : étape 1

Recherche par identifiants des documents

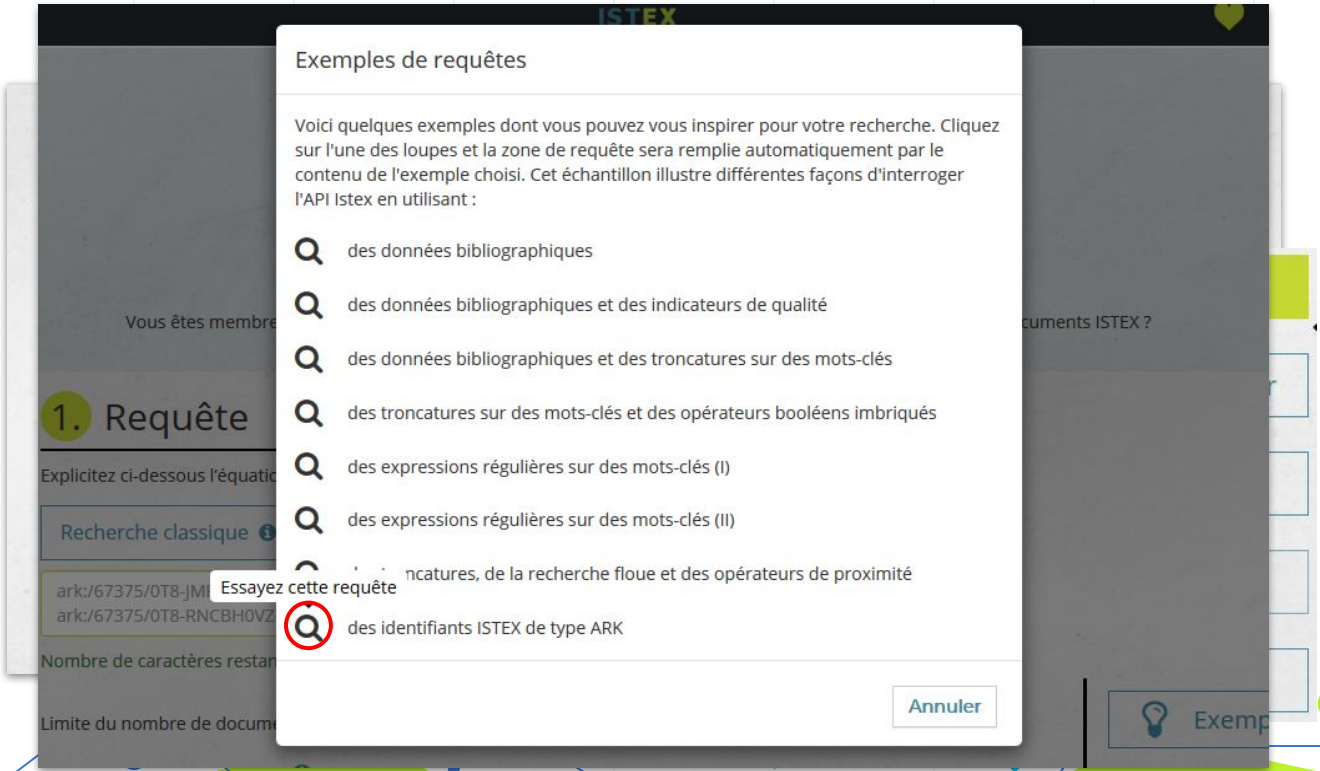
ISTEX-DL : étape 1

Recherche par identifiants des documents

ISTEX-DL : étape 1

Recherche par identifiants des documents

ISTEX-DL : étape 1



Recherche par identifiants des documents

ISTEX-DL : étape 1

Recherche par identifiants des documents

ISTEX-DL : étape 1

Longueur équation : limitation

(nombre & couleur)

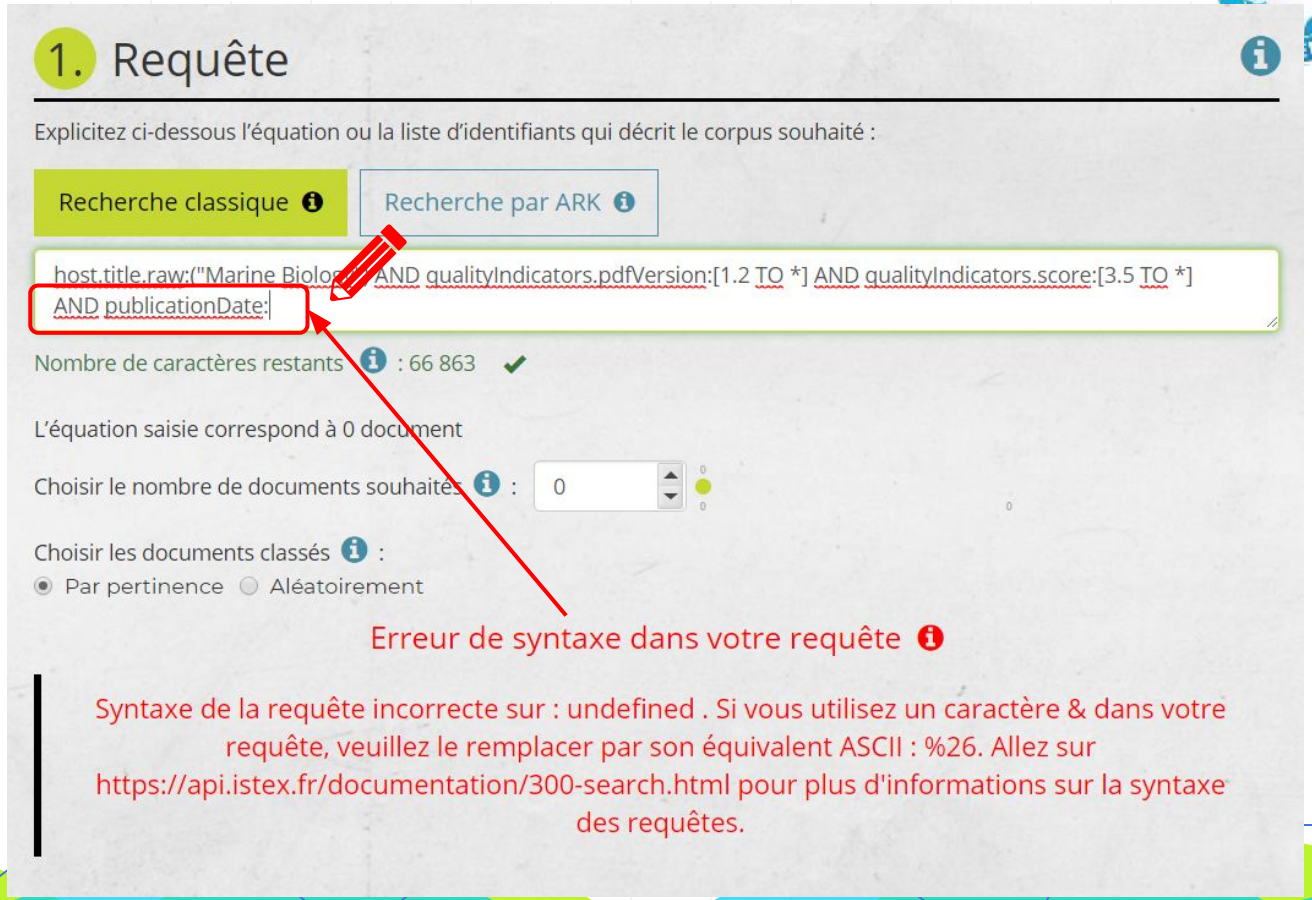
ISTEX-DL : étape 1

Recherche par équation booléenne

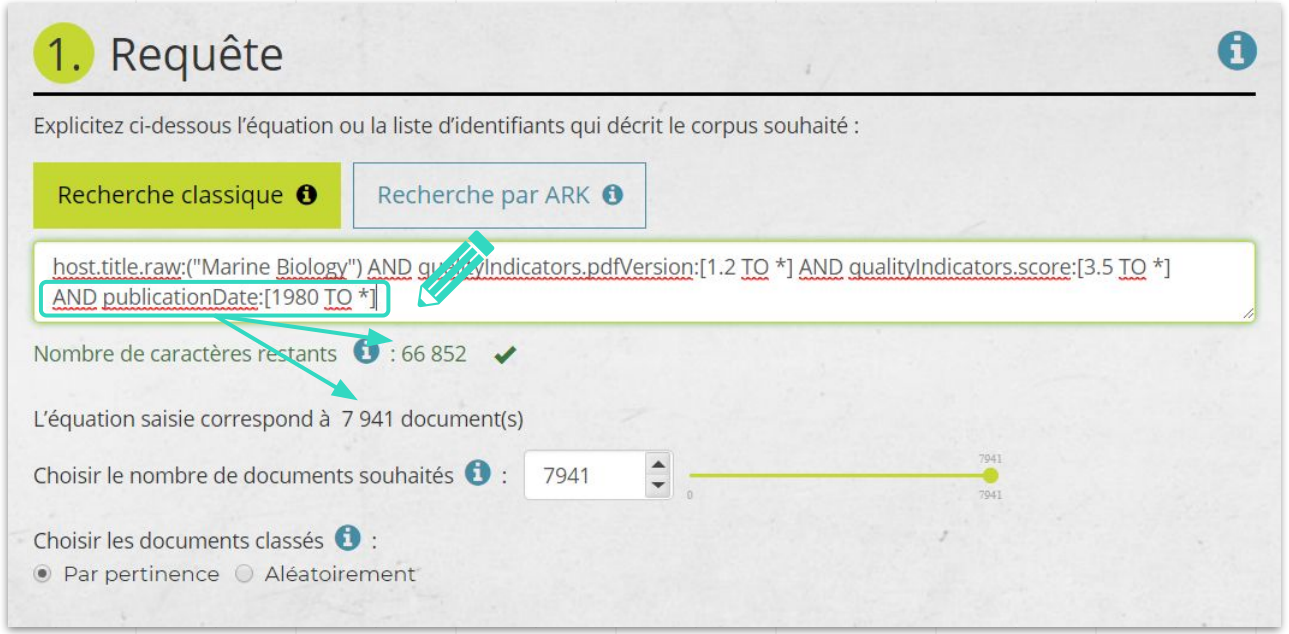
ISTEX-DL : étape 1

Recherche par équation booléenne

ISTEX-DL : étape 1



Recherche par équation booléenne



ISTEX-DL : étape 1

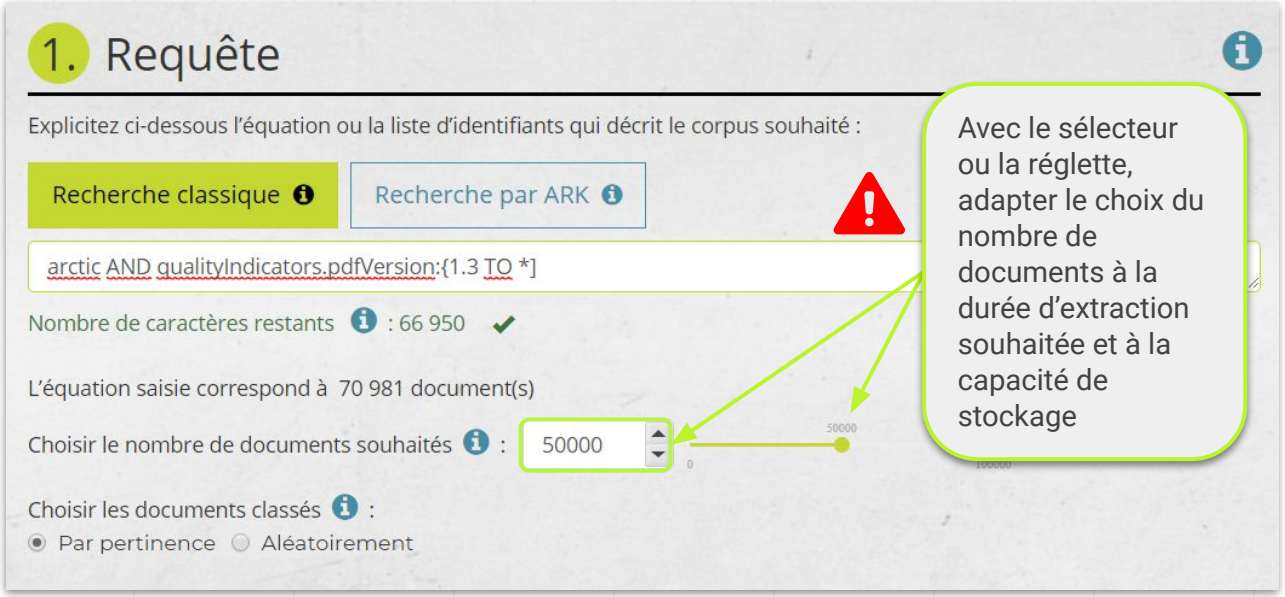
Nombre de documents : limitation

ISTEX-DL : étape 1

Nombre de documents : limitation

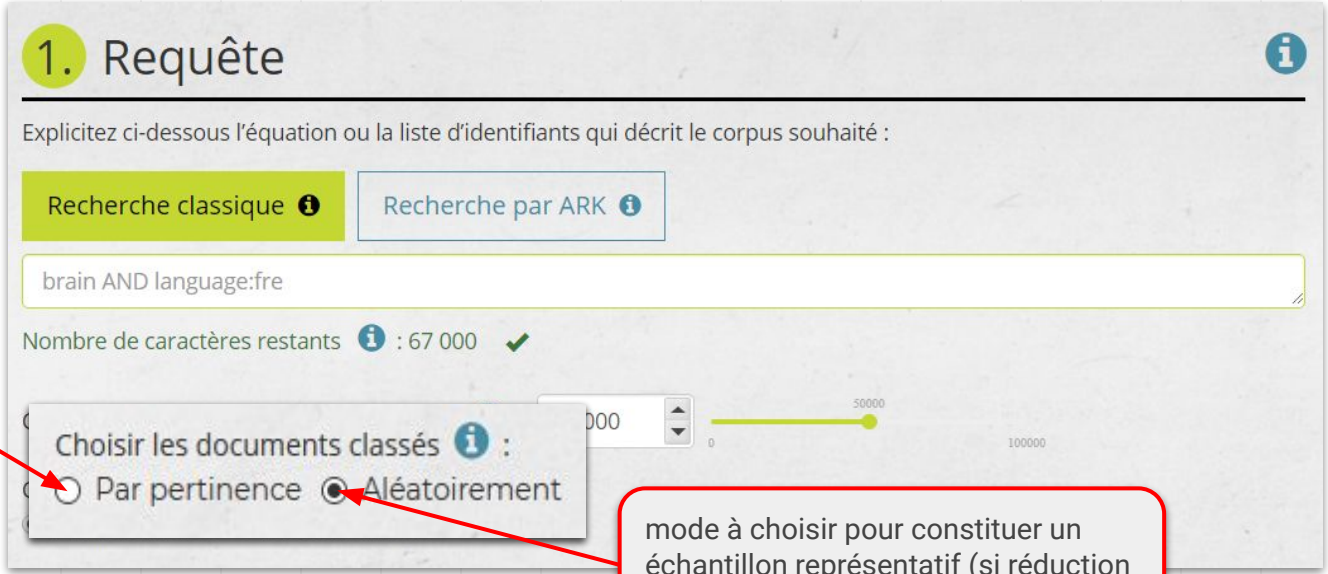
ISTEX-DL : étape 1

Nombre de documents : sélection



ISTEX-DL : étape 1

Deux modes de tri



mode par défaut qui trie selon un score de pertinence des documents en réponse à l'équation

mode à choisir pour constituer un échantillon représentatif (si réduction du nombre de documents)

ISTEX-DL : étape 1

Recherche par équation booléenne

Cas d'usage

ISTEX-DL : étape 1

1. Requête

Explicitez ci-dessous l'équation ou la liste d'identifiants qui décrit le corpus souhaité :

Recherche classique | Recherche par ARK

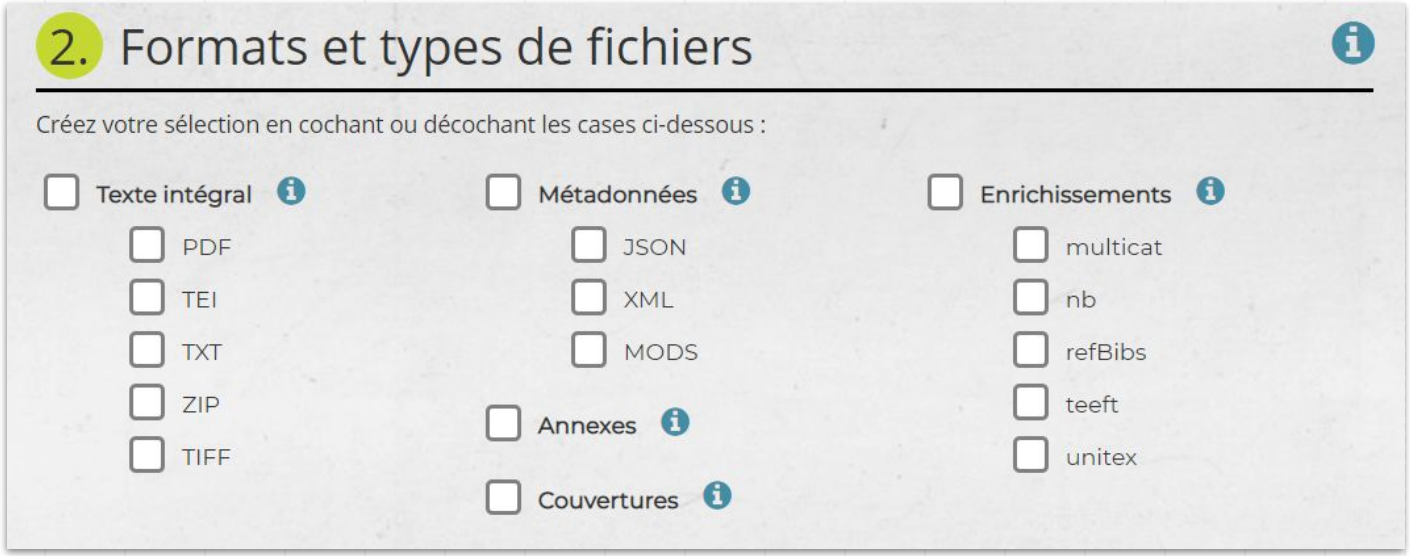
```
(title:((/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR /spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2 ))
 OR abstract:((/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR /spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2 ))
 OR subject.value:((/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR /spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2 )))
 AND language:eng
 AND qualityIndicators.pdfWordCount:[500 TO 10000]
```

Nombre de caractères restants : 66 460

L'équation saisie correspond à 1 197 document(s)

Choisir le nombre de documents souhaités : 1197

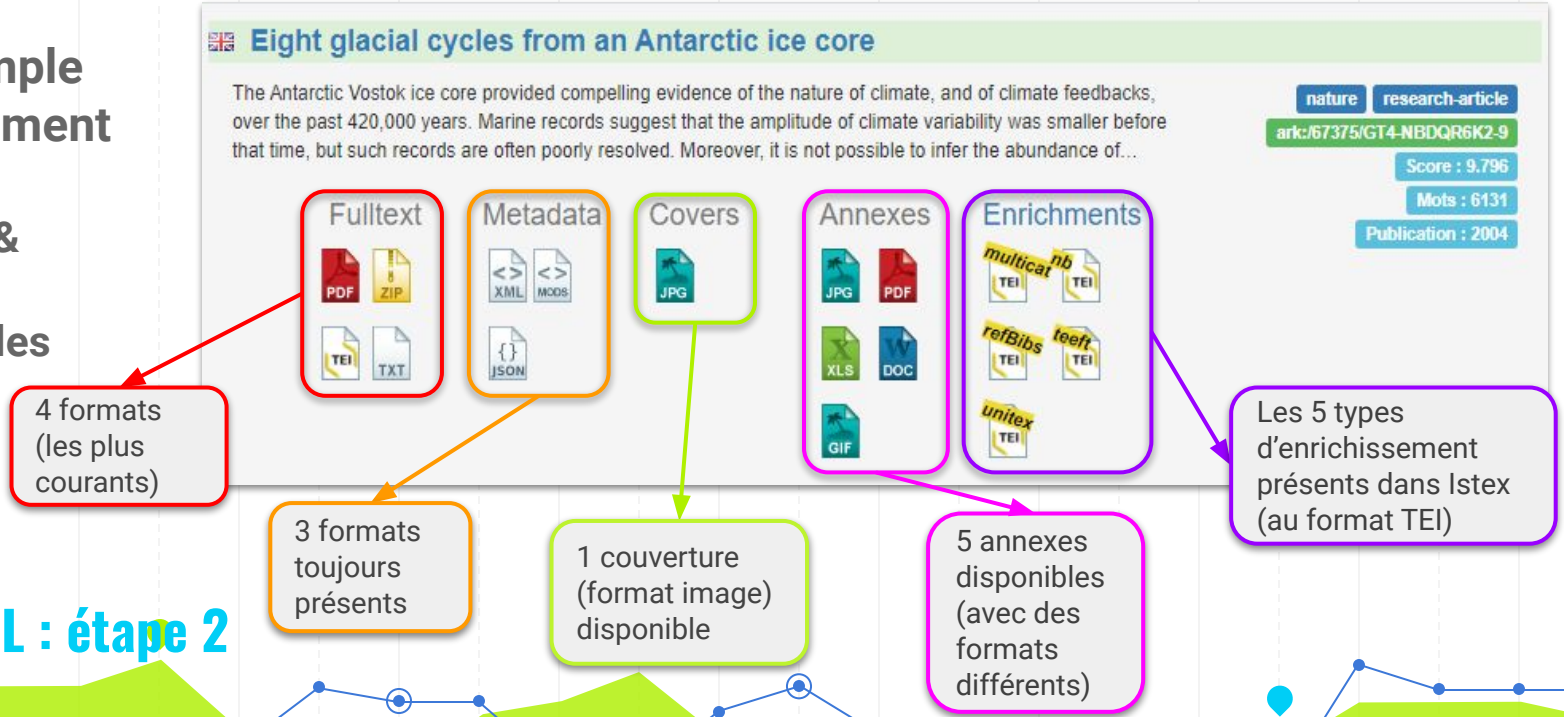
Sélection



ISTEX-DL : étape 2

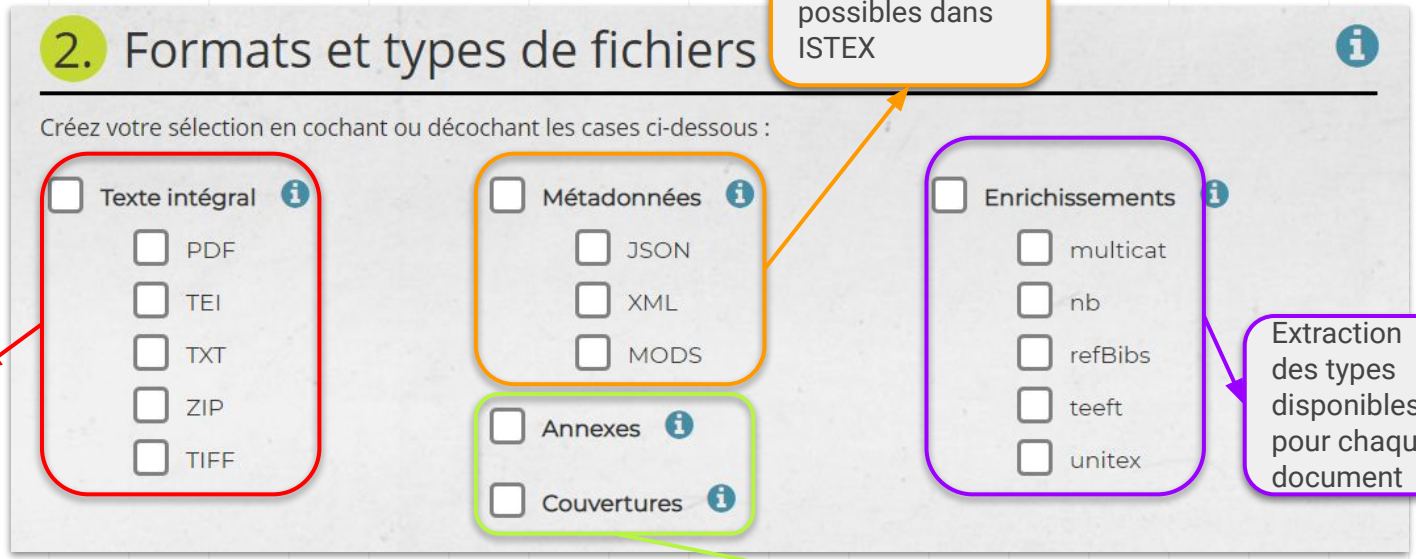
Un exemple de document

formats & fichiers disponibles



ISTEX-DL : étape 2

Sélection des fichiers



Liste de tous les formats possibles dans ISTEX

Le format OCR, s'il existe pour un document, sera préféré au TXT

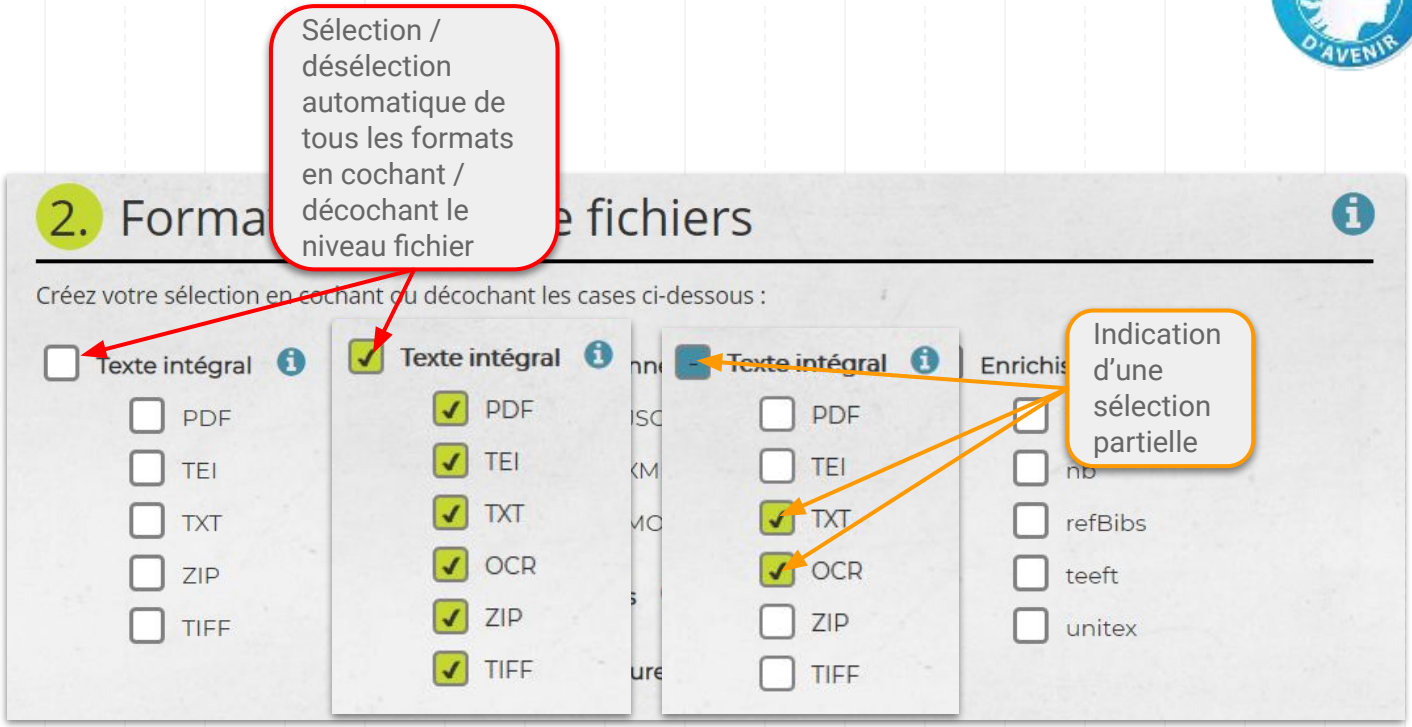
Extraction des types disponibles pour chaque document

Pas de détail des formats

ISTEX-DL : étape 2

Quels fichiers seront extraits ?

Aide des couleurs

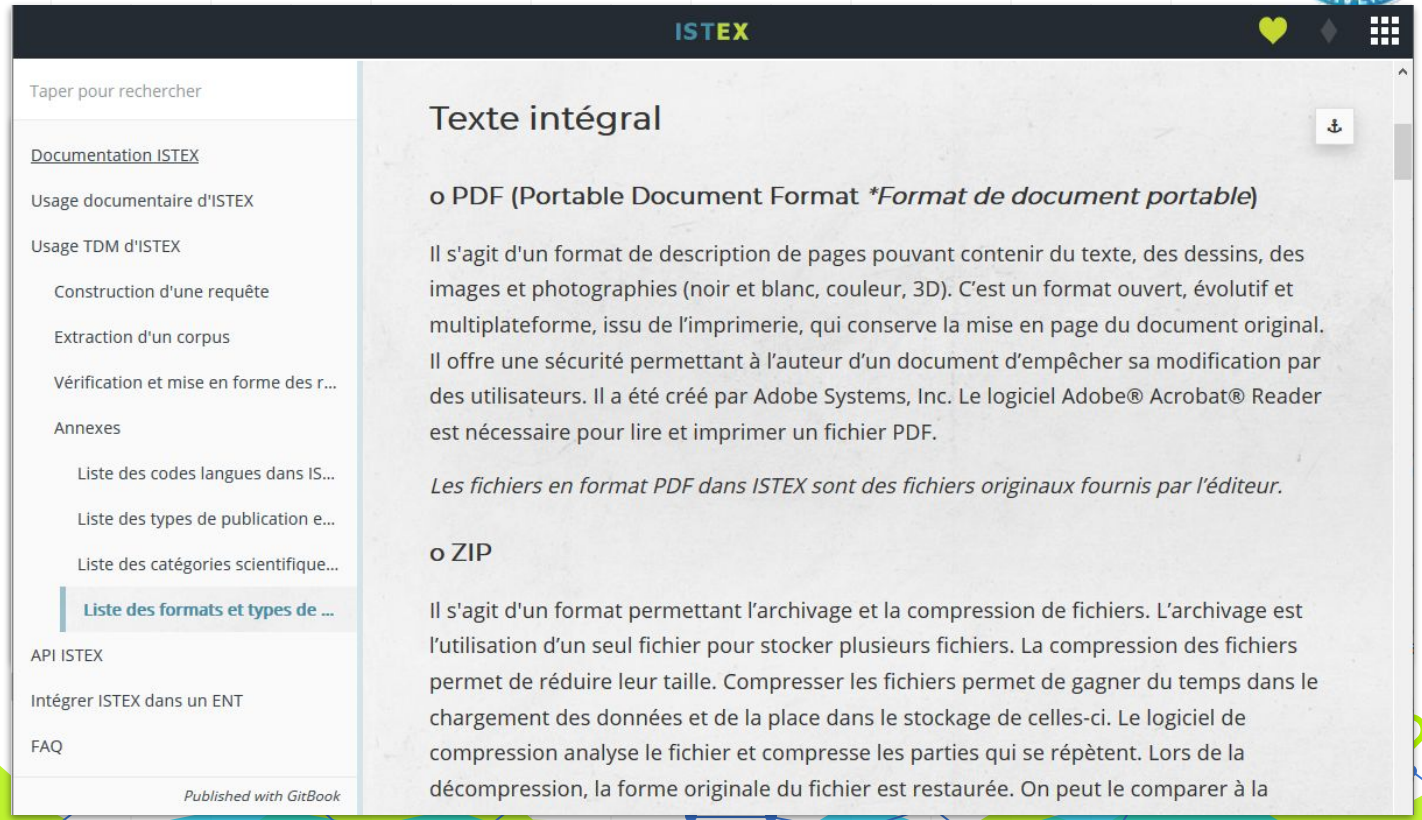


ISTEX-DL : étape 2

Quels fichiers choisir ?

Aide des infobulles

ISTEX-DL : étape



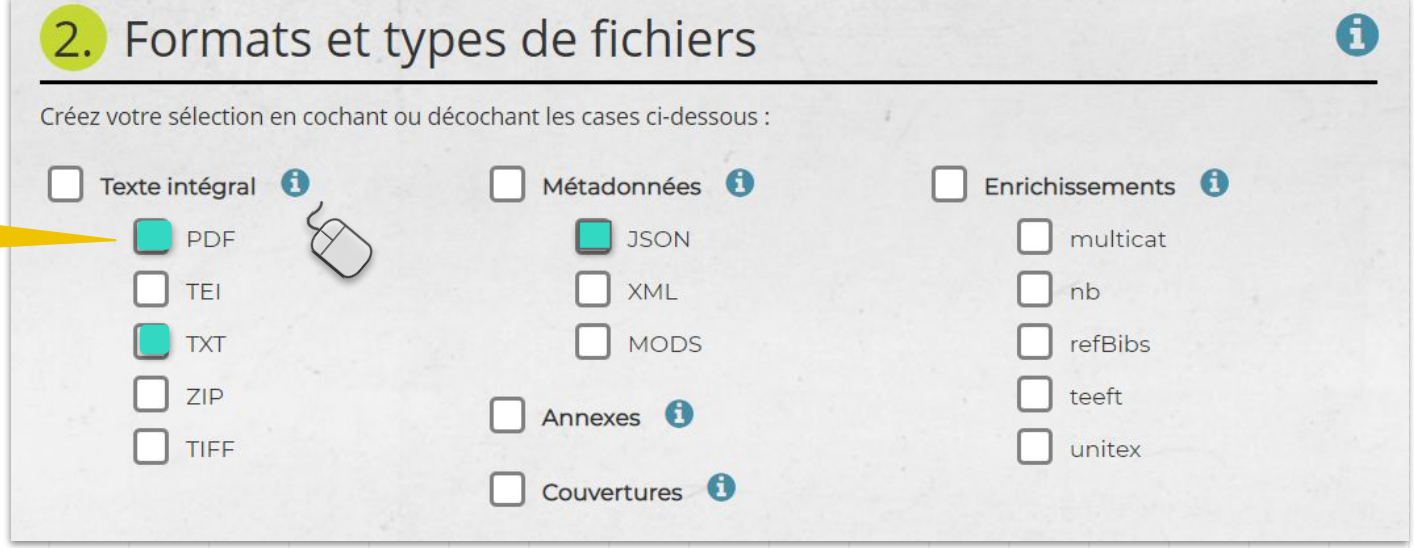
Quels fichiers choisir ?

Aide des infobulles

ISTEX-DL : étape 2

Quels fichiers choisir ?

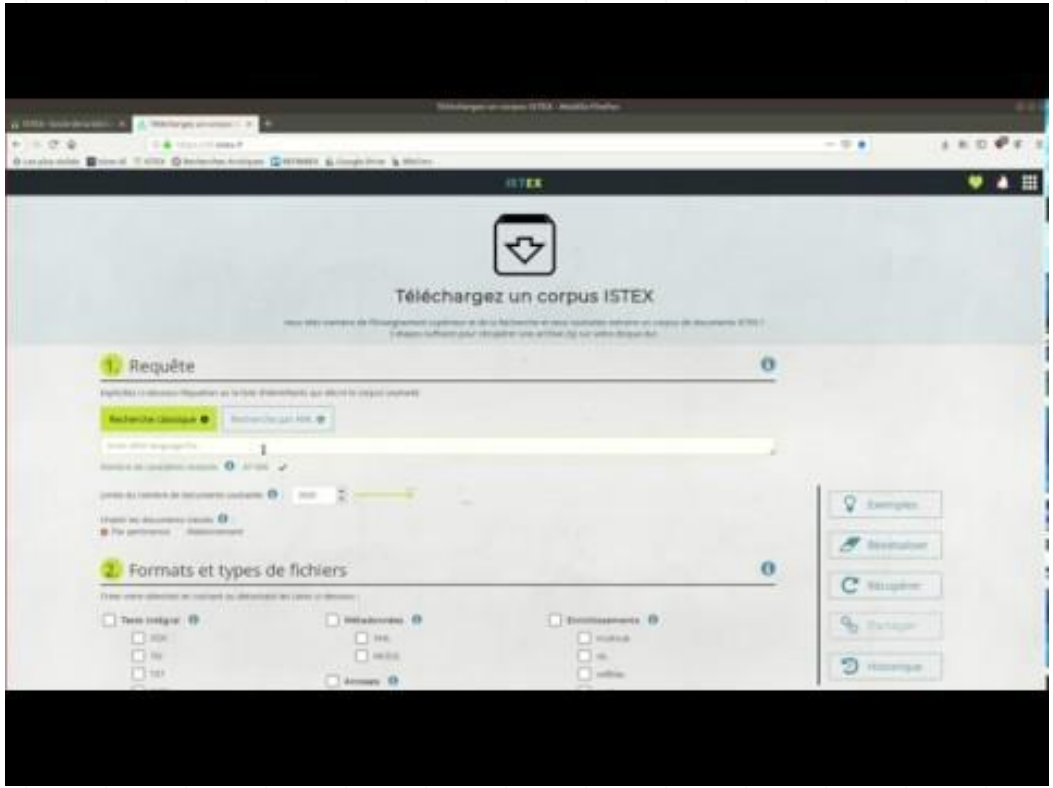
Cas d'usage



ISTEX-DL : étape 2

Fermeture inopinée ?

Bouton “Récupérer”

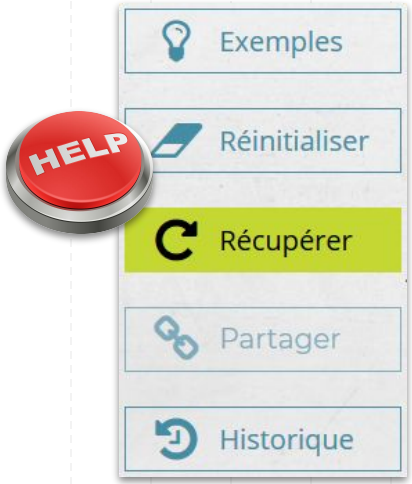


ISTEX-DL : étape 2

ISTEX-DL : étape 2

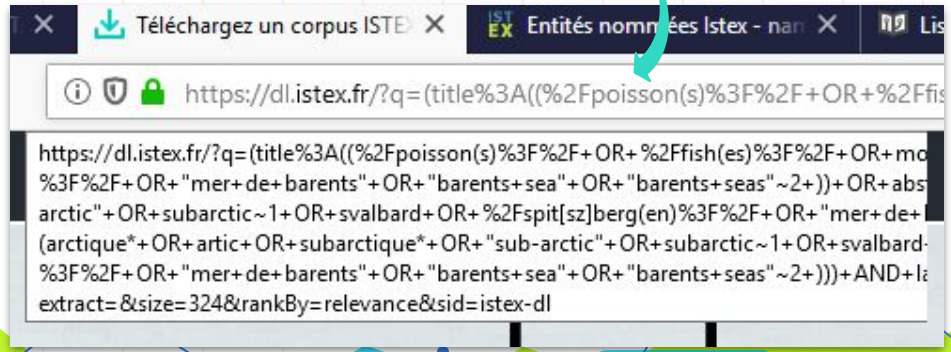
Fermeture inopinée ?

Bouton “Récupérer”



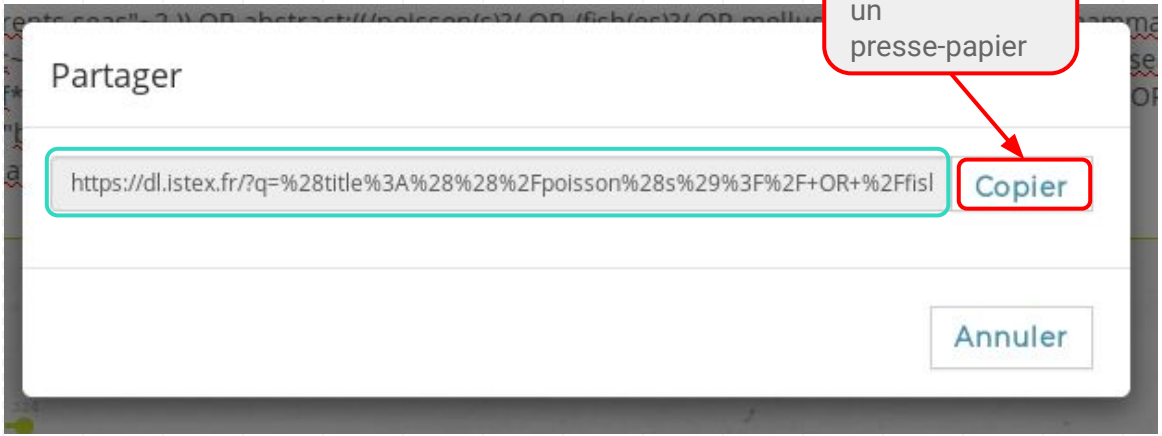
À vérifier :

- Paramétrage navigateur : **autorisation** des cookies et des données de site jusqu'à leur expiration
- Connexion au site : **https://dl.istex.fr** uniquement, sans précision de paramètres (attention à l’adresse enregistrée en favori)



Formulaire complété… 2 boutons s'activent

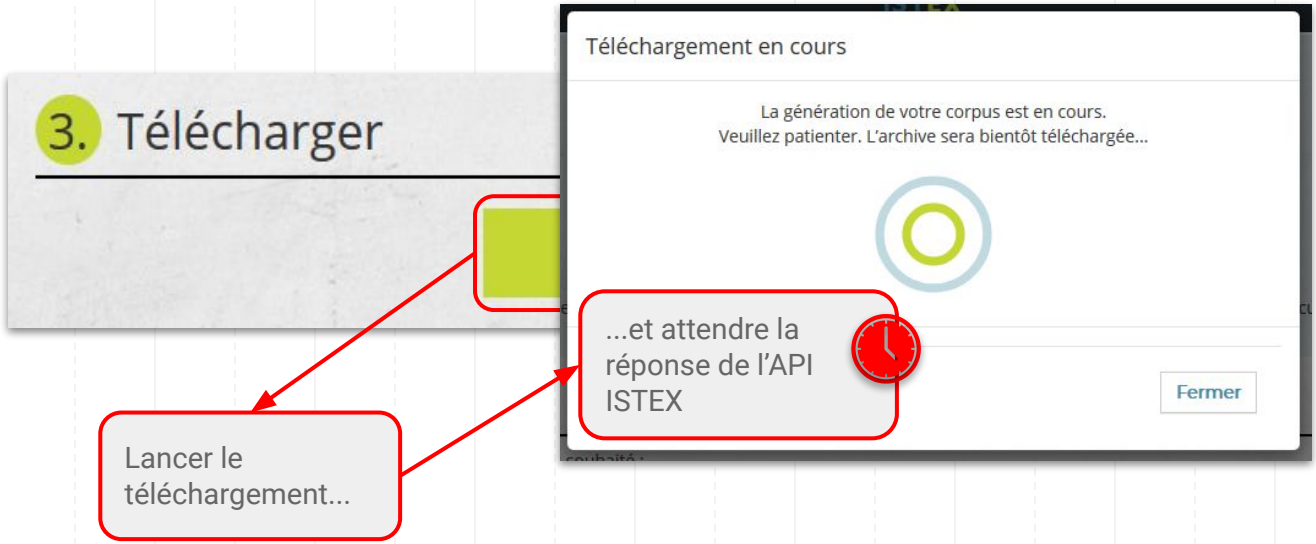
Bouton "Partager"



ISTEX-DL : étape 3

Formulaire complété… 2 boutons s’activent

Bouton “Télécharger”



ISTEX-DL : étape 3

S’authentifier si besoin

mode normal & mode nomade

Authentification dans votre établissement :

- automatique par adresse IP connue

ISTEX-DL : étape 3

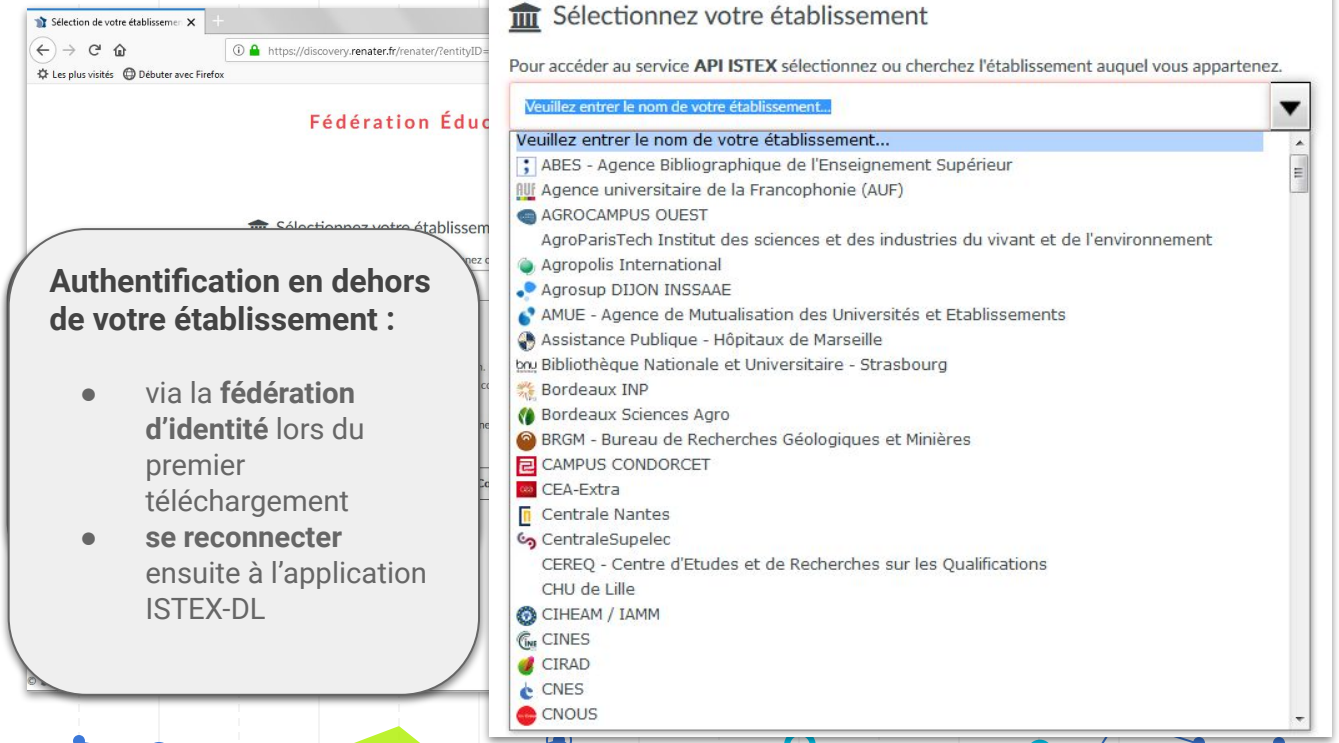
ISTEX-DL : étape 3

S’authentifier si besoin

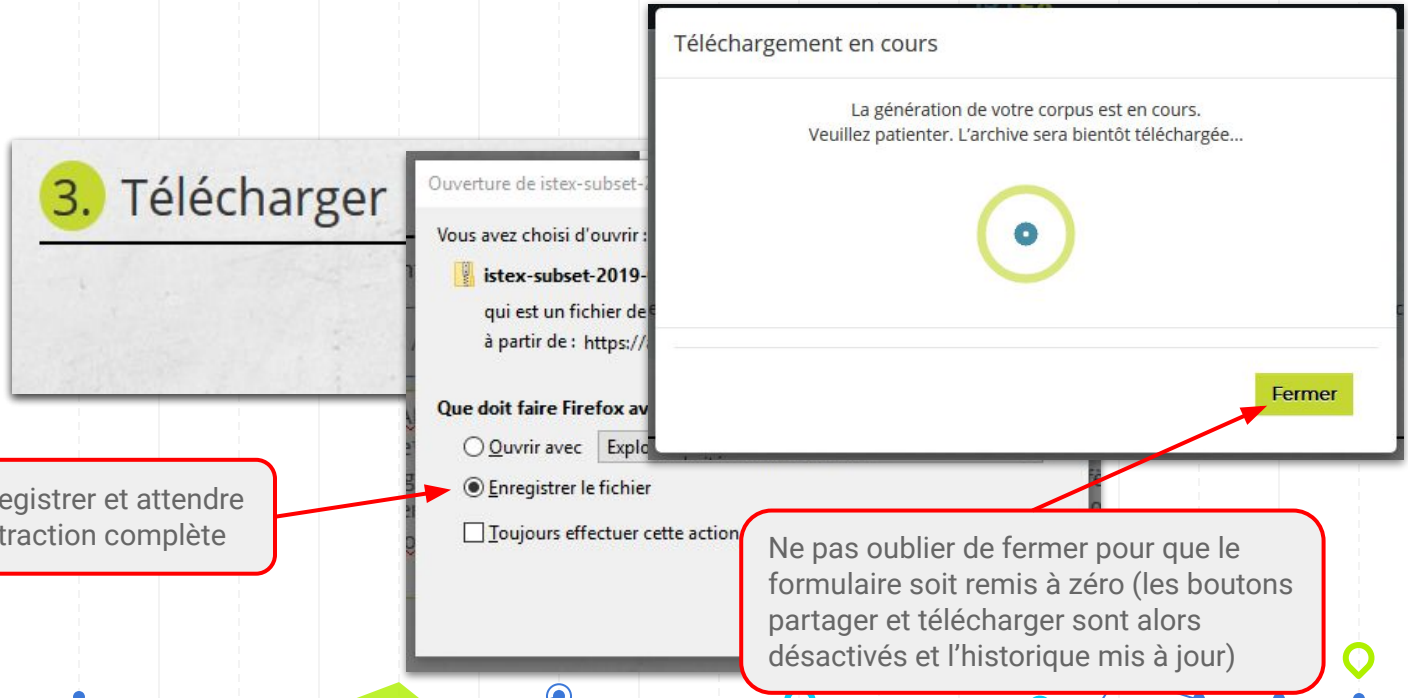
mode normal &
mode nomade

Authentification en dehors de votre établissement :

- via la **fédération d’identité** lors du premier téléchargement
- **se reconnecter** ensuite à l’application ISTEX-DL



Fin du téléchargement

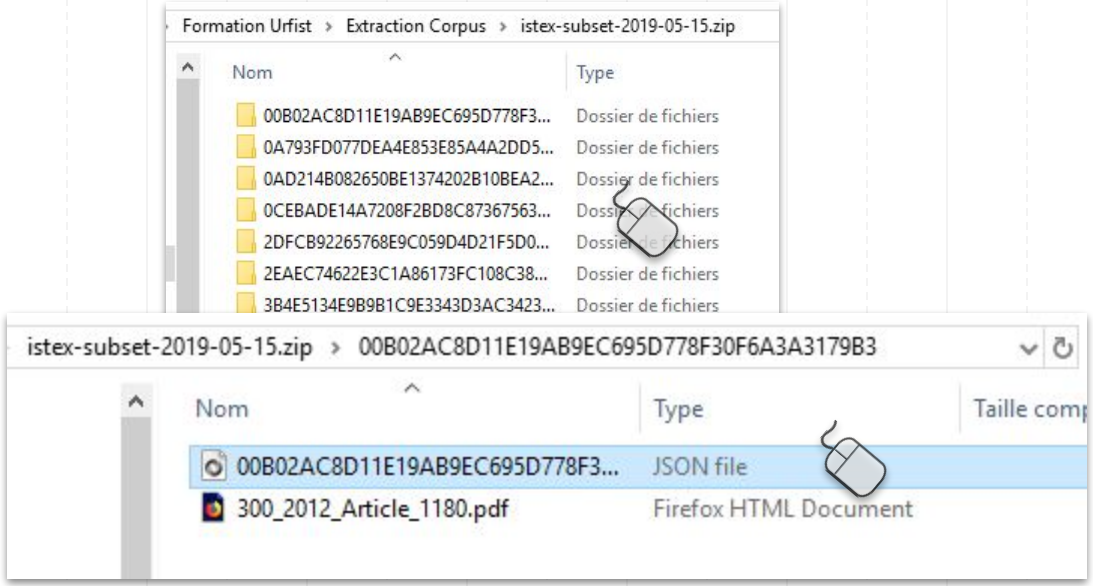


Enregistrer et attendre l'extraction complète

Ne pas oublier de fermer pour que le formulaire soit remis à zéro (les boutons partager et télécharger sont alors désactivés et l'historique mis à jour)

ISTEX-DL : étape 3

Fichier zip du corpus extrait



ISTEX-DL : étape 3

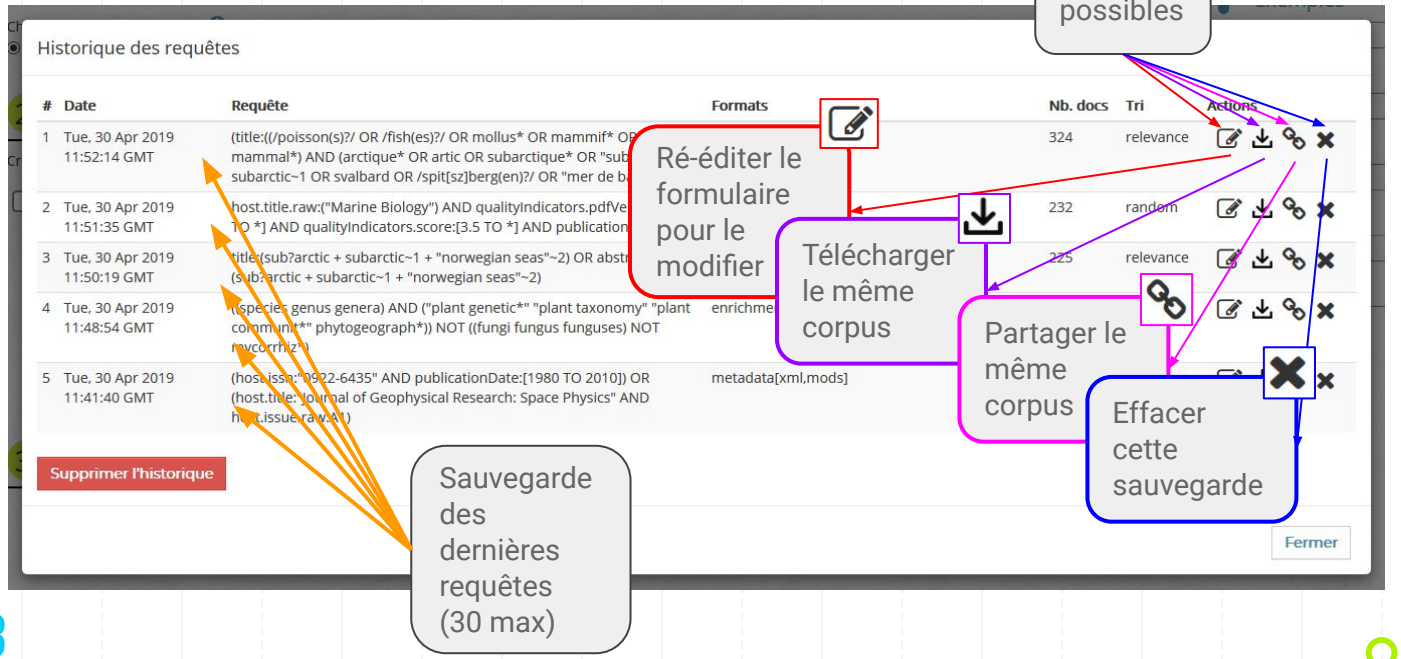
Fichier zip du corpus extrait


```
{
  "arkIstex": "ark:/67375/VQC-271B662S-H",
  "title": "Feeding ecology of dominant groundfish in the northern Bering Sea",
  "abstract": "Abstract: We investigated current diets of the six most abundant benthic fish in the
northern Bering Sea. Our objective was to explore feeding strategies and potential competition with
other top predators as ecosystem changes occur in the northern Bering Sea ecosystem. Our approach used
stomach content data collected from field sampling during spring 2006 and 2007. Calanoid copepods and
ampeliscid amphipods were important prey of Arctic cod (Boreogadus saida) but in different proportions
depending upon fish size, feeding location, and local environmental conditions. Snailfish (Liparidae)
occupied a broad niche and fed on a variety of benthic amphipods. Arctic alligatorfish (Ulcina olrikii)
and Arctic staghorn sculpin (Gymnocanthus tricuspis) consumed ampeliscid amphipods predominantly.
Shorthorn sculpin (Myoxocephalus scorpius) had a less-diverse diet, with snow crab (Chionoecetes
opilio) most important by weight. Finally, all Bering flounder (Hippoglossoides robustus) sampled had
empty stomachs. Our results indicate that ampeliscid amphipods, which have high biomass in the central
region of the northern Bering Sea, are the most important prey for the dominant groundfish in the
Chirikov Basin. Generally, all dominant benthic fish in the northern Bering Sea had narrow feeding
niches, except snailfish. High diet overlap was found among many of the fish species, including Arctic
cod and snailfish, snailfish and Arctic alligatorfish, and Arctic alligatorfish and Arctic staghorn
sculpin. These findings are consistent with a relatively short food chain for benthic fish that are for
the most part specialized feeders with narrow preferences for food and may be affected by changes in
benthic prey distributions.",
  "language": [
    "eng"
  ],
  "publicationDate": "2012",
  "copyrightDate": "2012",
  "doi": [
    "10.1007/s00300-012-1180-9"
  ],
  "ark": [
    "ark:/67375/VQC-271B662S-H"
  ],
  "articleId": [
    "1180",
    "s00300-012-1180-9"
  ],
  "genre": [
    "research-article"
  ],
  "originalGenre": [
    "OriginalPaper"
  ],
  "author": [
    {
```


ISTEX-DL : étape 3

Conserver une trace & Rejouer

Bouton “Historique”



ISTEX-DL : étape 3

Aide

Documentation :

- Application ISTEX-DL
  - adresse: https://doc.istex.fr/tdm/extraction/istex-dl.html
  - directement via le pied de page

NB : Application en évolution constante…

Vos retours sont bienvenus !

Autres Outils : Quels Avantages ?

1 - Fonction “extract” (API ISTEX) :

à ajouter et à paramétrer directement dans l’URL d’interrogation

https://api.istex.fr/document/?q=...&**extract**=...

- option **rankBy = qualityOverRelevance**
- option **rankBy = random** avec **randomSeed** (nombre unique à 13 chiffres)

NB : ISTEX-DL exploite la fonction “extract”

Documentation

- Fonction “extract” : https://doc.istex.fr/tdm/extraction/extract-feature.html

Autres Outils : Quels Avantages ?

2 - **Moissonneurs** : utilitaires à installer et à lancer en ligne de commande avec la liste des options choisies :

"istex-api-harvester" :

- extraction **au-delà de 100 000** documents (scroll à paramétrer)
- tous modes d'option rankBy

Documentation

Tous les détails pour installer et utiliser :

- moissonneur **istex-api-harvester**

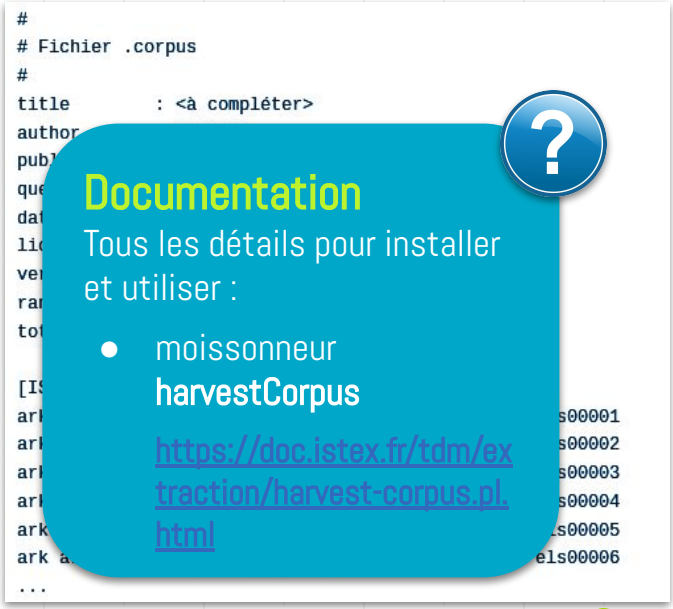
  https://doc.istex.fr/tdm/extraction/istex-api-harvester.html

Autres Outils : Quels Avantages ?

2 - **Moissonneurs** : utilitaires à installer et à lancer en ligne de commande avec la liste des options choisies :

"harvestCorpus" :

- extraction au-delà de 100 000 documents
- fichier **".corpus"** (liste des identifiants des documents)
- **statsCorpus :** outil de calcul de **statistiques** associé à cet utilitaire
- tirage aléatoire (avec "graine aléatoire" possible)
- possibilité de renommer les fichiers extraits



Télécharger un corpus ISTEX > Et après ?

Exploration du corpus dans LODEX

Import du fichier dans Lodex

Modélisation à l'aide du modèle prédéfini

Exploration des différents graphiques

Objectifs :

- Explorer le contenu du corpus pour affiner à nouveau la requête
- Exposer le corpus pour le partager avec la communauté scientifique

Exploration et valorisation du corpus

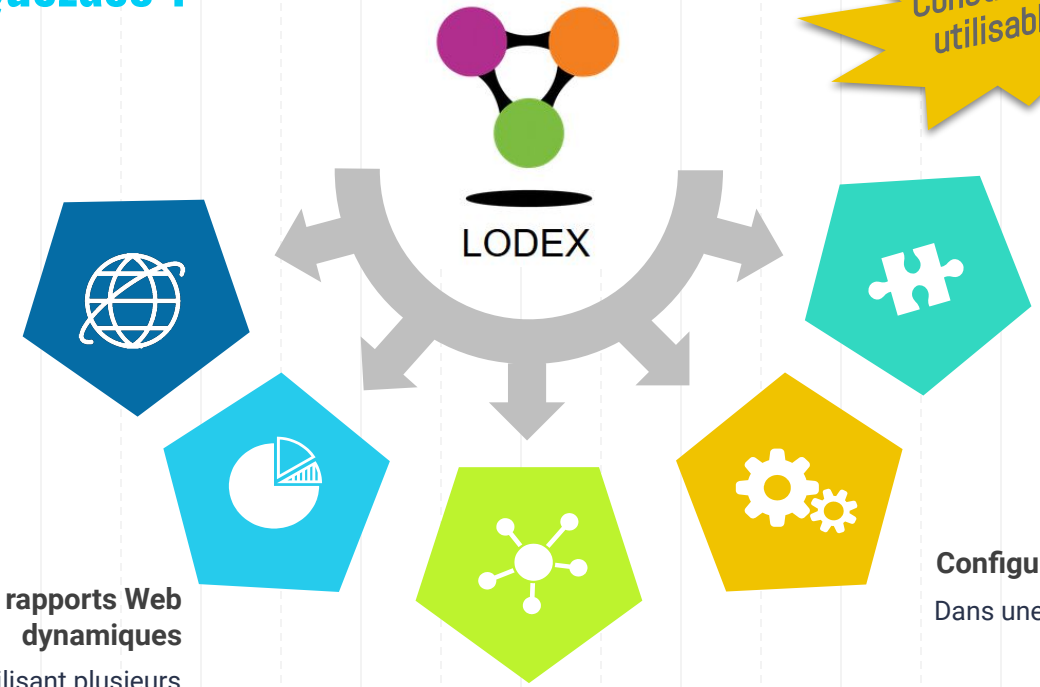
LODEX

4

LODEX : Quézaco ?

Application web open source
dédiée aux données structurées

Construit pour ISTEX, utilisable par tous !



Exposer ses données structurées sur le web

À partir de données csv, tsv, xml, json

Publier des rapports Web dynamiques

À l'aide de visualisations utilisant plusieurs graphiques (camemberts, histogrammes,….)

Exporter ses données en format du Web sémantique

JSON-LD, N-Quads

Gérer et attribuer des identifiants pérennes

ARK

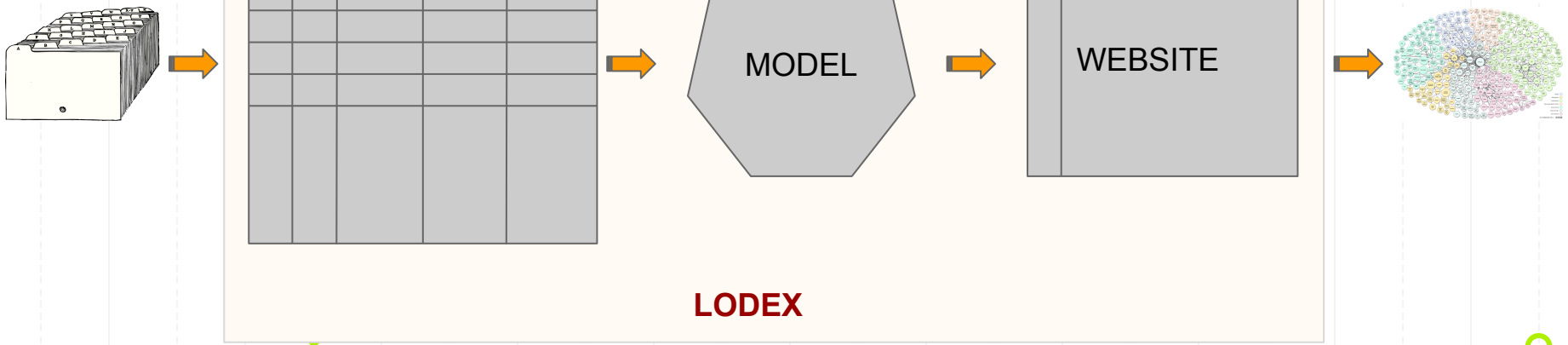
Configurer ses données

Dans une interface administrateur

https://lodex.inist.fr/

https://github.com/Inist-CNRS/lodex

LODEX > Principe

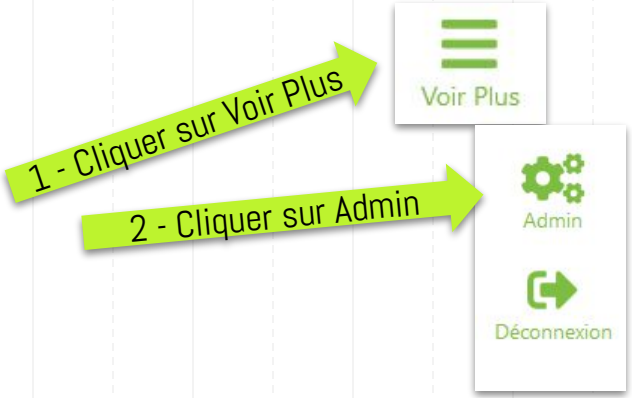


LODEX > on passe à l'action !

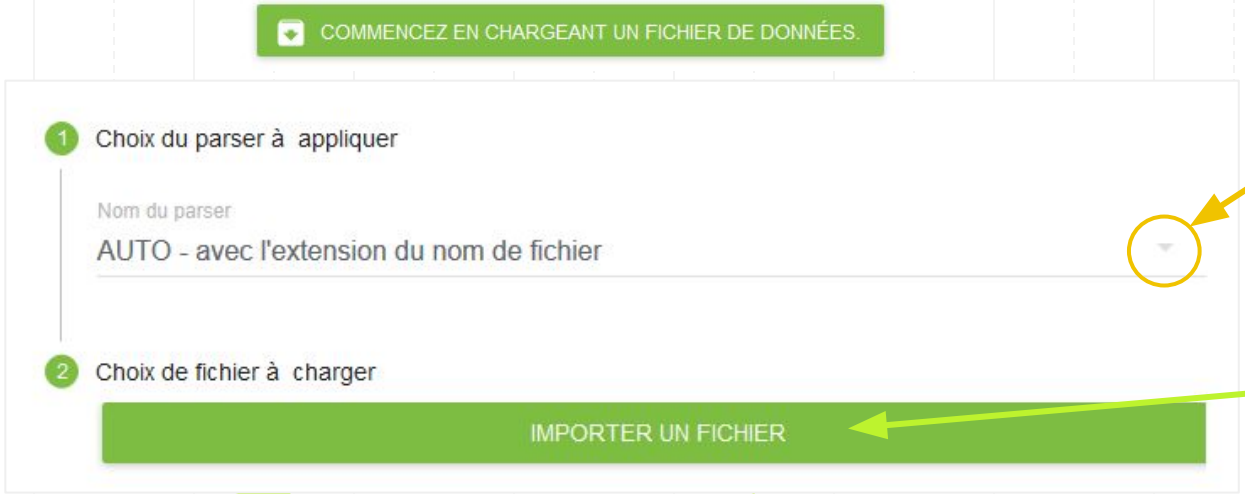
1. Accéder à une instance de l'outil
   - Adresse : **http://formation-urfistn.lodex-dev.inist.fr/**
   - Nom d'utilisateur : **admin**
   - Mot de passe : **secret**

instances numérotées de **2 à 20**

LODEX > on passe à l'action !



2. Import des données

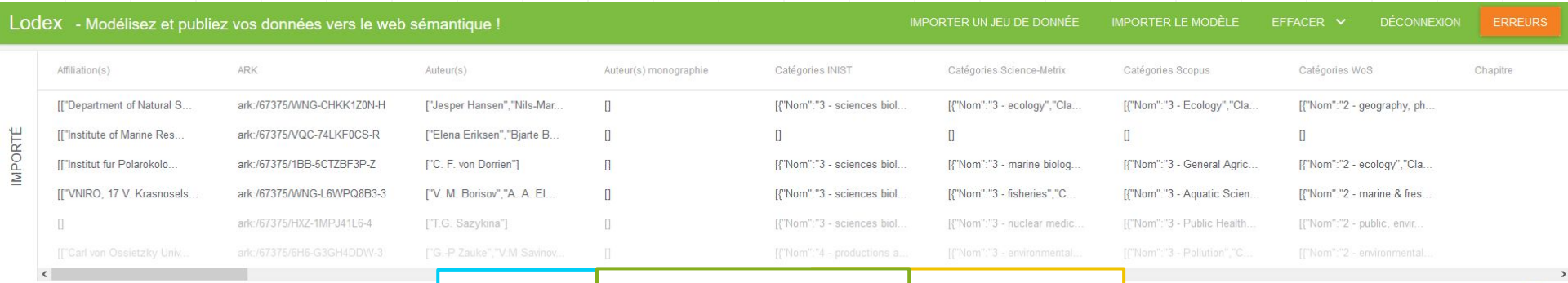


Choisir dans le menu déroulant : ZIP - résultat de dl.istex.fr

Choisir le fichier téléchargé depuis ISTEX-DL : istex-subset-YYYY-MM-DD.zip

LODEX > on passe à l'action !

Le jeu de données importé



LODEX > on passe à l'action !

1.1. Création du modèle

- Ajouter une colonne depuis le jeu de données original
- Ajouter une nouvelle colonne

1197 ressources chargées 56 colonnes chargées 62 colonnes ajoutées

uri	Nom du corpus	Nombre de documents	Titre de l'article	Lien vers le PDF	Type de corpus	Domaine(s)	Auteur(s)	Affiliation(s)	Langu
uid:/F8Z20TW8	Les poissons et les mollus...	/api/run/count-all	Late Pleistocene and Holoc...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie","Géographie"]	["Jesper Hansen","Nils-Mar...	[["Department of Natural S...	anglais
uid:/TZDP19FM	Les poissons et les mollus...	/api/run/count-all	Ecological significance of...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie","Géographie"]	["Elena Eriksen","Bjarte B...	[["Institute of Marine Res...	anglais
uid:/CQSKD1FG	Les poissons et les mollus...	/api/run/count-all	Reproduction and larval ec...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie","Géographie"]	["C. F. von Dorrien"]	[["Institut für Polarökolo...	anglais
uid:/D297MCJP	Les poissons et les mollus...	/api/run/count-all	Long-term variations of bi...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie","Géographie"]	["V. M. Borisov","A. A. El...	[["VNIRO, 17 V. Krasnosels...	anglais
uid:/MD69C5G4	Les poissons et les mollus...	/api/run/count-all	Long-Distance Radionuclide...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie","Géographie"]	["T.G. Sazykina"]	[]	anglais
uid:/T0T5PT3G	Les poissons et les mollus...	/api/run/count-all	Heavy metals in fish from ...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie","Géographie"]	["G.-P Zauke","V.M Savinov...	[["Carl von Ossietzky Univ...	anglais

LODEX > Le modèle

Permet de configurer toutes les informations à afficher sur un corpus, à partir :

- Du jeu de données
- De nouvelles colonnes

LODEX > on passe à l'action !

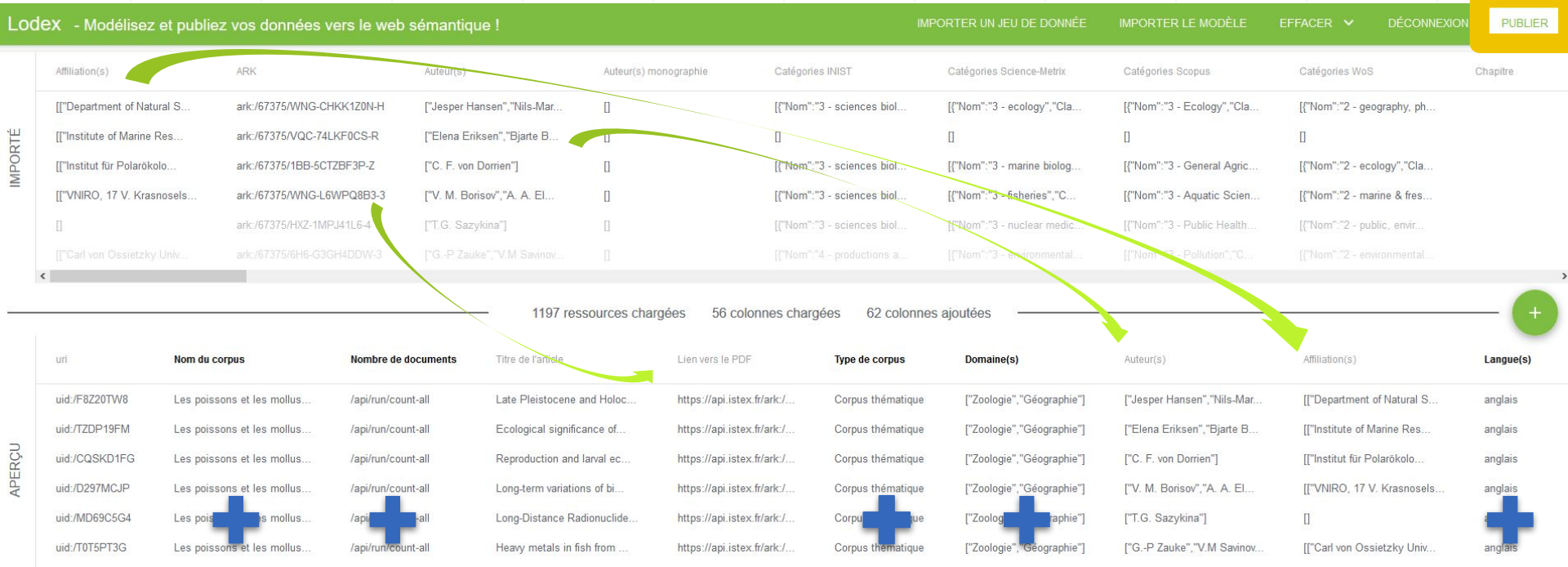
1.2. Import d'un modèle existant

Confirmer et Sélectionner le fichier modele-atelier.json

Un même modèle peut être appliqué à différents jeux de données de structure identique

LODEX > on passe à l'action !

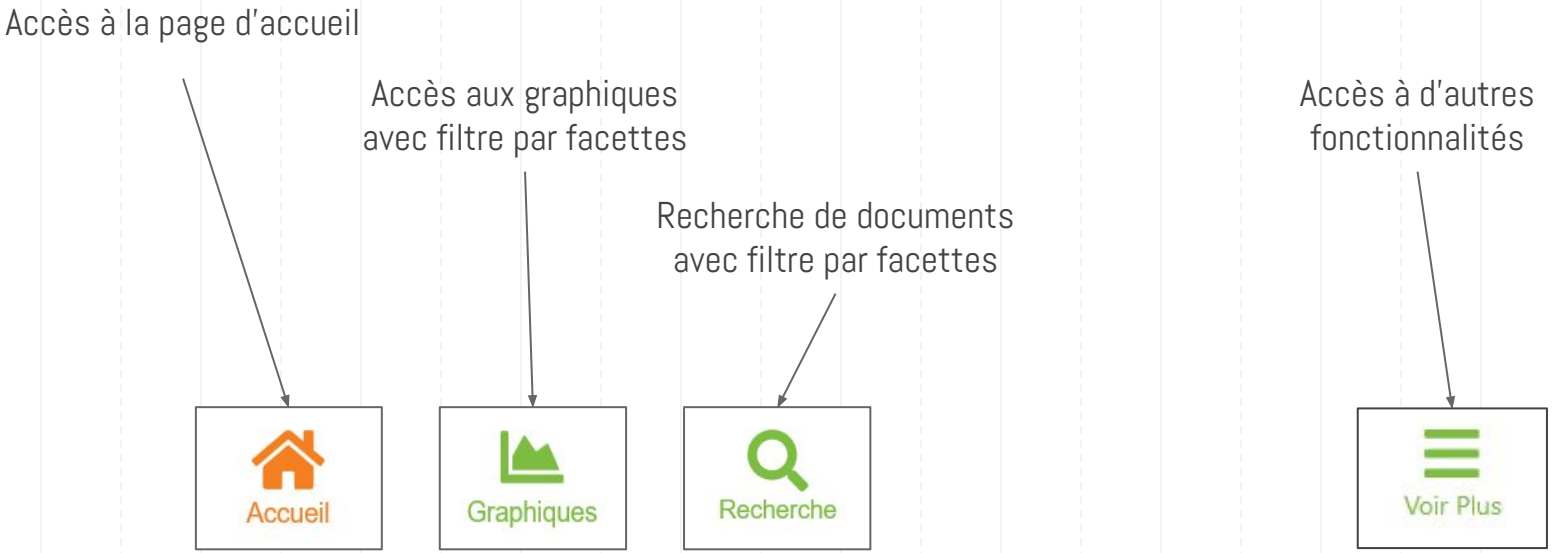
Le modèle importé



LODEX > on passe à l'action !

Les données publiées

LODEX > comment naviguer dans le corpus



LODEX > exploration des données publiées

Page d'accueil

- Présentation du corpus
  - Identification du corpus
  - Rappel du cas d'usage
  - Possibilité de rejouer la requête
  - Droits d'utilisation des documents
  - Citation du corpus

Exposition du corpus

LODEX > modification des données publiées

- 2 icônes disponibles pour chaque champ
  - Pour modifier la valeur du champ
  - Pour modifier le paramétrage du champ (format, facette, etc.)

Permet d'adapter le modèle d'un corpus à l'autre

- Exercices
  1. Modifier le titre du corpus
  2. Modifier le format du titre du corpus

LODEX > exploration des données publiées

Les graphiques

- Visualisation de la répartition des documents selon différents angles de vue
  - Dates de publication
  - Revues
  - Type de publication, type de document
  - Éditeurs
  - Mots-clés d'auteur
  - Catégories scientifiques (Science-Metrix,Scopus, WoS, Inist)
  - Indicateurs (version de PDF, score de qualité)

Analyse du corpus

LODEX > exploration des données publiées

Les graphiques

- Exercices

1. Explorer le graphique “Revues”
   - ☐ utiliser la facette “Titre de revue” pour accéder aux documents d'une revue
   - ☐ utiliser la facette “Date de publication” pour cibler les revues récentes
   - ☐ changer le nombre de revues affichées

LODEX > exploration des données publiées

Les graphiques

- Exercices

2. Explorer le graphique “Type de documents”

   utiliser la facette “Type de documents” pour identifier les documents qui ne sont **pas** de type : article, article de recherche, article de synthèse

3. Explorer le graphique “Catégories Science-Metrix”

   naviguer dans les différents niveaux de hiérarchie pour valider la qualité du corpus

LODEX > de l'exploration à la ré-interrogation d'ISTEX

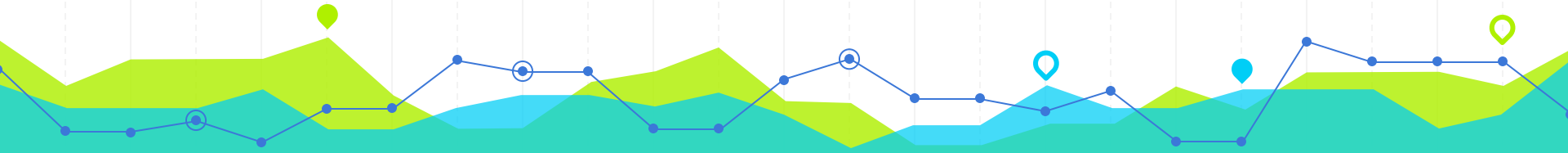
Équation affinée

```
(title:((/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR
/spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2 ))

OR abstract:((/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR
/spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2 ))

OR subject.value:((/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR
/spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2 )))

AND language:eng

AND qualityIndicators.pdfWordCount:[500 TO 10000]

NOT genre:abstract NOT categories.scienceMetrix:pediatrics
```

Résultats (03 mars 2020) : 1 197 ➢ 1 194 docs

LODEX > exemples de corpus Inist

- Corpus avec enrichissement classification issue de Topic Modeling
  - Vieillissement
- Corpus avec annotation manuelle disponible
  - Unitex
- Corpus avec enrichissement systématique animale
  - Échantillon Systématique Animale (version en développement)
- Corpus avec évolution diachronique des thématiques
  - Géosciences (version en développement)



5

Liens utiles

Adresses & Co
Documentation & Informations
Contact & Discussion

Adresses & Co

Se connecter :

- ISTEX : http://www.istex.fr
- Démonstrateur ISTEX : http://demo.istex.fr/
- Interface ISTEX-DL : https://dl.istex.fr/
- Infos Lodex : https://lodex.inist.fr/
- Données ISTEX : https://data.istex.fr/

S'authentifier :

- Vérifier ses droits d'accès : https://api.istex.fr/auth
- Vérifier son accès par fédération d'identité : https://api.istex.fr/auth?auth=fede

Documentation & Informations

Se documenter :

- Documentation “Usage TDM d’ISTEX” : https://doc.istex.fr/tdm/
- Documentation “API ISTEX” : https://doc.istex.fr/api/
- Documentation “Lodex” : https://user-doc.lodex.inist.fr/

Se tenir informé :

- Blog ISTEX : https://blog.istex.fr/
- Plateforme Twitter : @ISTEX_Platform

Contact & Discussion

Chercher de l'aide / Contribuer à l'amélioration :

- Contact :
  - Via le formulaire : https://www.istex.fr/contact/
  - Via la liste : contact@listes.istex.fr

- Liste de discussion : users@listes.istex.fr (liste publique)

L'équipe ISTEX

Annexes 6

S'aider...

Interrogation > quelques champs utiles

*Penser à utiliser les champs **"host.title" et "host.issn"** pour rechercher une revue (changement d'ISSN ou de titre au cours du temps, erreurs orthographiques, etc.)*

Titre d'article	title
Auteur	author.name
Résumé	abstract
Mots-clés d'auteur	subject.value
Date de publication	publicationDate
Langue	language
Titre de revue	host.title
ISSN revue	host.issn

*Les champs **.raw** permettent d'interroger sur la valeur exacte du champ.*

*Exemple : **q=host.title.raw:"Science"** ne ramènera que les documents de la revue Science et non ceux de "The Science of the Total Environment", etc.*

S'aider...

Interrogation > quelques champs utiles

Éditeur	corpusName
Type de publication	host.genre
Type de document	genre

Enrichissements

Disciplines scientifiques	categories.scopus, categories.inist, categories.sciencesMetrix, categories.wos
Entités nommées	namedEntities.unitex, etc.

S'aider...

Interrogation > quelques champs utiles

Indicateurs (TDM)

Score de qualité	qualityIndicators.score
Version de PDF	qualityIndicators.pdfVersion
Nombre de mots par PDF	qualityIndicators.pdfWordCount

À venir

Pdf image	*qualityIndicators.pdfText*
Nombre de mots par page	*qualityIndicators.pdfWordsPerPage*

Sauter le pas...

Interrogation > intérêts de l'API ISTEX

- Exploration du contenu du fonds ISTEX et répartition des documents en fonction d'une facette
  - Exemple : connaître toutes les langues des documents présents dans ISTEX
- Interrogation selon une ou plusieurs facettes imbriquées
  - Exemple : Pour chaque langue présente dans ISTEX, connaître le nombre de documents par siècle
- Interrogation de toutes les facettes disponibles dans l'API (nombre réduit sur le démonstrateur)

Interface de Programmation Applicative *: solution informatique qui permet à des applications de communiquer entre elles et de s'échanger mutuellement des services ou des données*

Tutos

Pour comprendre ce qu'est l'API ISTEX et apprendre à l'utiliser, reportez-vous aux tutoriels :

https://istex-tutorial.data.istex.fr/

S'inspirer...

Interrogation > Exemples de requêtes sur l'API

*Exemple **1** : lister tous les éditeurs présents dans ISTEX et connaître le nombre de documents pour chacun d'eux **sans afficher** les documents (emploi de **size=0**)*

Éditeurs	https://api.istex.fr/document/?q=*&facet=**corpusName[*]**&size=0
Types de publication	https://api.istex.fr/document/?q=*&facet=**host.genre[*]**&size=0
Types de document	https://api.istex.fr/document/?q=*&facet=**genre[*]**&size=0

S'inspirer...

Interrogation > Exemples de requêtes sur l'API

*Exemple **1** : explorer la répartition du nombre de documents Istex par période de **100 ans entre les 15e et 21e siècles***

Dates de publication	par siècle (entre 1401 et 2100)	https://api.istex.fr/document/?q=*&size=0&facet=publicationDate[**1401-2100:100**]
	par période de 10 ans (20& et début 21 siècles)	https://api.istex.fr/document/?q=*&size=0&facet=publicationDate[**1901-2020:10**]
Langues & dates	les dates pour chaque langue **(facettes imbriquées)**	https://api.istex.fr/document/?q=*&facet=**language[*]>publicationDate[1401-2100:100]**&size=0

S'inspirer...

Interrogation > Exemples de requêtes sur l'API

*Exemples **1 & 2** : savoir si les documents présents dans ISTEX **comportent ou non un résumé et afficher** ces documents*

Présence de résumé	avec résumé	https://api.istex.fr/document/?q=(**abstract:***)
	sans résumé	https://api.istex.fr/document/?q=(**NOT (abstract:*)**)
Présence de mots-clés d'auteur	avec mots-clés	https://api.istex.fr/document/?q=(**subject.value.raw:***)&size=0
	sans mots-clés	https://api.istex.fr/document/?q=(**NOT (subject.value.raw:***)**)&size=0

S'inspirer...

Interrogation > Exemples de requêtes sur l'API

*Exemple **1** : quantifier les documents appartenant à **chaque catégorie Scopus** d'un corpus de documents sur **l'Arctique***

Domaines scientifiques

catégories Scopus d'un corpus particulier

https://api.istex.fr/document/?q=**arctic**&facet=**categories.scopus[*]**&size=0

catégories Science-Metrix de l'archive ISTEX

https://api.istex.fr/document/?q=***&facet=**categories.scienceMetrix[*]**&size=0

S'inspirer...

Interrogation > Exemples de requêtes sur l'API

*Exemple **1** : obtenir la distribution du nombre de documents ISTEX en fonction de leur **score de qualité**, et ce par **pas de 1***

Indicateurs	score de qualité	https://api.istex.fr/document/?q=*&size=0&facet=qualityIndicators.**score[1-10:1]**
	version de PDF	https://api.istex.fr/document/?q=*&size=0&facet=qualityIndicators.**pdfVersion[*]**
	nombre de mots par PDF	https://api.istex.fr/document/?q=*&size=0&facet=qualityIndicators.**pdfWordCount[1-10000:1000]**

Merci !

Avez-vous des questions ?