

ISTEX

Un réservoir pour le TDM

TDM ?

Text & Data 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

Un document VS un lot de documents

Usage Documentaire



Usage TDM



Exemples d'usage TDM

Pilotage & Recherche

- ★ Détection de sujets de recherche émergents
- ★ Détection de plagiat
- ★ Recommandation d'articles

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

Exemples d'usage TDM à l'Inist

Réaliser des analyses scientifiques

- ★ Analyses syntaxiques et terminologiques
- ★ Evolution temporelle de thématiques

Évaluer la performance d'outils

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

Mettre au point ou tester des outils

- ★ Détection de noms d'espèces animales et végétales
Entity-Fishing/IRC3
- ★ Détection de définitions de concepts scientifiques
Unitex
- ★ Détection de termes et de leurs variantes
TermSuite
- ★ Constitution et analyse de corpus
TermSuite/Teef + Neurodoc
- ★ Exploration thématique
Topic Modeling, Iramuteq

ISTEX ?

“ *Construire le socle de la bibliothèque scientifique numérique nationale.* ”

- ★ 2011 - 2018, un projet créé dans le cadre des PIA
- ★ Aujourd'hui, un service pour l'ESR

Finalité ?

- ★ Acquisition massive et centralisée d'archives scientifiques
 - collections rétrospectives de la littérature scientifique
 - toutes disciplines
- ★ Mise à disposition des données

<https://www.istex.fr/>



ISTEX



1. Usage documentaire



- bouton 
- API/widgets 
- Google Scholar 
- Outils Biblio. 



Enrichissements

2. Usage TDM



data.istex.fr



dl.istex.fr



corpus.istex.fr

Gargantext, Cillex, Cortext, Iramuteq ...

En chiffres

- ★ **29** corpus éditeur
- ★ **9 307** revues
- ★ **348 056** monographies
- ★ **23 320 273** documents
- ★ **73 740 142** enrichissements | métadonnées
générees
- ★ **~50 000** PDF consultés par semaine

Mode d'accès

- ★ Réservé à l'enseignement supérieur et la recherche.
- ★ Accessible par adhésion des établissements.

Par :

- *plage d'adresse IP*
- *login:password*
- *fédération d'identité*

ISTEX

Authentification

Vous êtes sur le point de lancer l'adhésion à ISTEX, si vous voulez vous informer sur ce qu'offre l'adhésion, cliquez [ici](#).

L'identifiant et le mot de passe à utiliser sont ceux du site [licencesnationales.fr](#)

Identifiant

Mot de passe

Se connecter

Vous avez oublié votre mot de passe ?

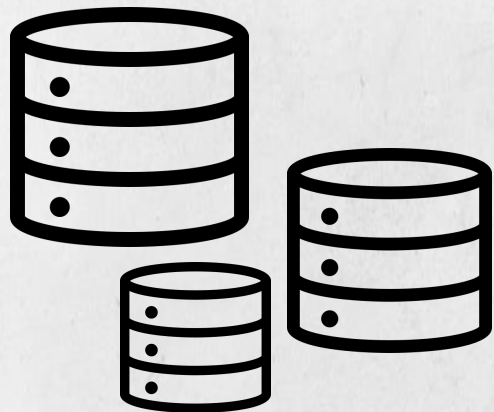
Votre établissement n'a pas encore de compte ? Vous serez dirigé sur le site [licencesnationales.fr](#) de l'ABES pour en créer un.

+ Créer un compte

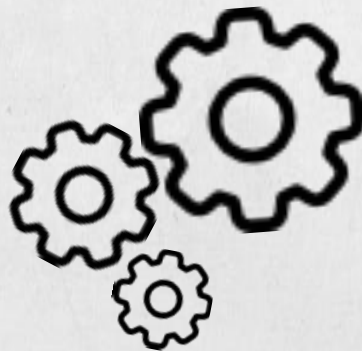
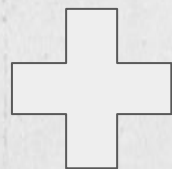
 Adhérer

ISTEX & la science ouverte

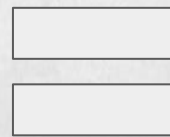
La fouilles et l'analyse



Des données,
Des textes



Des outils



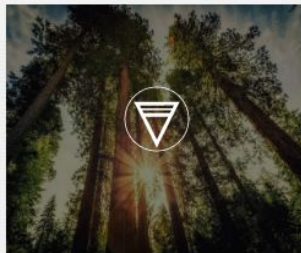
De nouvelles
perspectives



Text Analytics + Semantic Enrichment

Transform previously unusable text to data in a richly annotated, machine-readable and standardise data formats

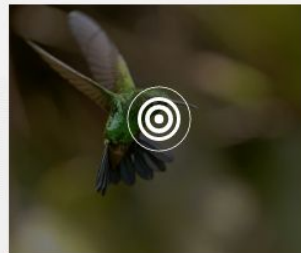
[Learn more](#)



Ontology Management

Harmonise technology through scientific ontologies adhering to public standards, whilst curating vocabularies

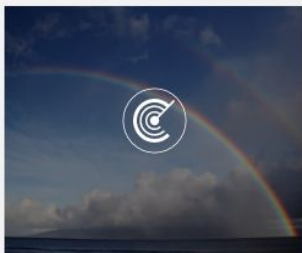
[Learn more](#)



Enterprise FAIR Data

The challenges of harmonising data to be Findable, Accessible, Interoperable and Reusable

[Learn more](#)



Semantic Search

Semantic search, visualise results, integrate into your existing platforms and automate your workflows



Artificial Intelligence

Combine deep learning and semantic algorithms to build powerful models that can exploit life science data and accelerate its use in R&D

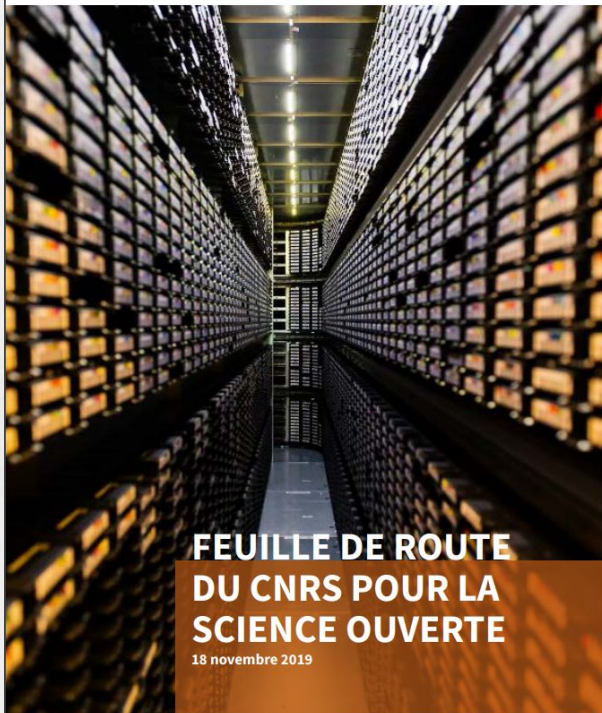


Knowledge Graphs

Semantic enrichment technology that facilitates the production of knowledge graphs



Privacy - Terms



FEUILLE DE ROUTE DU CNRS POUR LA SCIENCE OUVERTE

18 novembre 2019

3 | LA FOUILLE ET L'ANALYSE DES TEXTES ET DES DONNÉES

Objectif: **Faciliter la fouille des textes** et des données avec le développement des infrastructures, des outils et des compétences permettant l'analyse de contenus scientifiques **en toute indépendance.**

ISTEX une réponse pour

Faciliter la fouille et l'analyse de textes :

- ★ Un seul lieu pour des nombreuses sources
- ★ Permettre l'accès en masse à des données standardisées, interopérables et augmentées

ISTEX une réponse pour

La fouille de texte et l'analyse **en toute indépendance** :

- ★ Textes et métadonnées stockées par le CNRS et accessibles sans restriction aux membres de l'ESR
- ★ Un cadre juridique sécurisé : une license appropriée
- ★ Des connexions vers des outils/plateformes créés par le monde académique

Des données compatibles TDM

Des données adaptées

- ★ utilisation de formats standards

- TEI, MODS, RDF

- ★ homogénéisation des données

- standardisation (langues, type de documents)
- correction des erreurs

⇒ Gommer les différences entre les fournisseurs
OBJECTIF : Minimiser les “pré-traitements”

Des données augmentées

- ★ Des données Ré-océrisation
Pour permettre aux documents d'être "recherchable" & "analysable" par des outils TDM
- ★ Structuration du plein texte
Pour permettre à des outils de traiter ou de chercher dans différentes parties du texte
- ★ Production de nouvelles métadonnées
Pour extraire les documents spécifiques à un domaine de recherche

Enrichissement des données

du TDM dans ISTEX

Extraction des Références bibliographiques

Détecter et structurer
les références
bibliographiques des
articles en XML TEI

References

1. Lopez, P., Romary, L.: Multiple retrieval models and regression models for prior art search. In: CLEF 2009 Workshop, Technical Notes, Corfu, Greece (2009)
2. Nakov, P., Schwartz, A., Hearst, M.: Citances: Citation sentences for semantic

```

▼<biblStruct xml:id="b0" resp="#ISTEX-API" change="#refBibs-istex">
  ▼<analytic>
    ▼<title level="a" type="main">
      Multiple retrieval models and regression models for prior art search
    </title>
    ▼<author>
      ▼<persName>
        <forename type="first">P</forename>
        <surname>Lopez</surname>
      </persName>
    </author>
    ▼<author>
      ▼<persName>
        <forename type="first">L</forename>
        <surname>Romary</surname>
      </persName>
    </author>
  </analytic>
  ▼<monogr>
    <title level="m">CLEF 2009 Workshop</title>
    ▼<meeting>
      ▼<address>
        <addrLine>Corfu, Greece</addrLine>
      </address>
    </meeting>
    ▼<imprint>
      <date type="published" when="2009"/>
    </imprint>
  </monogr>
</biblStruct>
  
```


Indexation automatique

Extraire du texte les termes les plus représentatifs du contenu
quelque soit le domaine scientifique

A Retrospective Mortality Study of Workers Exposed to Arsenic in a Gold Mine and Refinery in France

L. Simonato, MD, J.J. Moulin, MD, B. Javelaud, MD, PhD, G. Ferro, BSc,
P. Wild, BSc, R. Winkelmann, MA, and R. Saraccl, MD

A historical mortality study of a cohort of employees of a gold mining and refining company was carried out in Salsigne, France. A major goal of the study was to investigate the relationship between lung cancer mortality and exposure to arsenic, radon, silica, and other contaminants of the working environment. A twofold excess of lung cancer was found both among miners and smelters, mainly concentrated among workers who had experienced exposure to past levels of arsenic, radon, and silica. The consistency of the results in the mine and the refinery are suggestive of a carcinogenic risk from both soluble and insoluble arsenic, although the potential role of other factors cannot be dismissed. © 1994 Wiley-Liss, Inc.

Key words: radon, silica, gold mining and refining, retrospective cohort, lung cancer

INTRODUCTION

An apparent high incidence of neoplasms of the respiratory system among employees in gold extraction and refining in Salsigne (Aude) was first reported in 1977 [doctoral thesis by Perisse, 1976–77] from the Department of Pneumology of the General Hospital in Carcassonne. Forty cases of lung cancer were included in the first investigation, whose results, even in the absence of a formal comparison group, appeared to indicate a large excess when considering the time period and the size of the population studied. A similar case series was subsequently reported in 1985 in another doctoral thesis written by Jammes [1985].

```
<listAnnotation type="rd-teeft">
  <annotationBlock corresp="text" xmlns="https://www.tei-c.org/ns/1.0">
    <keywords change="#istex:rd" resp="#istex:rd">
      <term>
        <term>lung cancer</term>
        <fs type="statistics">
          <cf name="frequency">
            <numeric value="17"/>
          </cf>
          <cf name="specificity">
            <numeric value="1"/>
          </cf>
        </fs>
      </term>
      <term>
        <term>radon</term>
        <fs type="statistics">
          <cf name="frequency">
            <numeric value="14"/>
          </cf>
          <cf name="specificity">
            <numeric value="0.823529411764786"/>
          </cf>
        </fs>
      </term>
    </keywords>
  </annotationBlock>
</listAnnotation>
```

Lung Cancer
Cohort
Arsenic
Miner
Refinery
Salsigne
Diesel exhaust
First exposure


Catégorisation par domaines scientifiques

⇒ 71,5%

⇒ 36,2%


★ Par appariement :

WoS, 

Scopus, 

Science Metrix 

★ Par apprentissage automatique :

Classification Pascal/Francis 



Catégorisation par appariement
WoS : Plant Sciences

Catégorisation par apprentissage
Agronomie, Sciences du sol et productions végétales

Détection des entités nommées

⇒ 69,2%

9 types d'entités :

- Personnes,
- Lieux,
- Organisations,
- Projets financés,
- Organisme financeur,
- Hébergeur de ressources,
- URL,
- Dates,
- Citations

INTRODUCTION

An apparent high incidence of neoplasms of the respiratory system among employees in gold extraction and refining → **Salsigne** (Aude) was first reported in 1977 [doctoral thesis by Perisse, 1976-77] from the **Department of Pneumology of the General Hospital in Carcassonne**. Forty cases of lung cancer were included in the first investigation, whose results, even in the absence of a formal comparison group, appeared to indicate a large excess when considering the time period and the size of the population studied. A similar case series was subsequently reported in **1985** in another doctoral thesis written by Jammes [1985].

Désambiguïisation et alignement des lieux

Alignement des entités nommées de type PlaceName avec la ressource GeoNames



The current issue and full text archive of this journal is available at www.emeraldinsight.com/0025-1747.htm

MD
47,1

Management implications of foreign players in the English Premiership League football

Nnamdi Madichie
University of East London Business School, London, UK

Position	Team	P	W	D	L	F	A	GD	Pts
1	Manchester United	38	27	6	5	80	22	58	87
2	Chelsea	38	25	10	3	65	26	38	85
3	Arsenal	38	24	11	3	74	31	43	83
4	Liverpool	38	21	13	4	67	28	39	76
5	Everton	38	19	8	11	55	33	22	65
6	Aston Villa	38	16	12	10	71	51	20	60
7	Sheff Weds	38	15	12	11	50	48	2	58
8	Portsmouth	38	16	9	13	48	40	8	57
9	Manchester City	38	15	10	13	45	53	-8	55
10	West Ham United	38	13	10	15	42	50	-8	49

Source: www.skysports.com/football/league/0,19540,11660,00.html

Table III.
Top ten Barclays Premiership teams (end of season - 11 May 2008)

U.K.

Portsmouth

U.S.A.

FEDERAL EMPLOYEES AND VOTING IN FEDERAL ELECTIONS

DALMAS H. NELSON, *San Fernando Valley State College*
and
PAUL J. HOFFMANN, *St. Paul, Minnesota*

AMONG the many uncertainties of American political behavior are the voting practices of those who work for the federal government. Though they constitute a substantial number of voters in the nation and their support is cultivated somewhat by both major political parties, little solid information has been available as to their ways at the polls. This paper examines their role in federal elections.

Speculation about federal employee voting practices has tended to take divergent directions. One view is that represented by *Business Week*, which says that federal employees have a stake in governmental activity and large expenditures which "generally makes them Democrats in national politics, even during Republican administrations."¹

The third group is made up of the three Virginia cities of Alexandria, Falls Church, and Portsmouth. They are, on the map, within the perimeters of Fairfax County in the case of Alexandria and Falls Church, and of Norfolk County in the case of Portsmouth.² But technically they are outside of the counties by virtue of city-county separation. All three are Section 16 cities, Alexandria and Falls Church due to their proximity to the capital and Portsmouth as the result of determination by the Commission that federal personnel (and their wives) constitute a majority of the voters. The three separated cities are grouped together because the electoral process in these separated cities was felt to be significantly different from that of either counties or municipalities not separated from their counties.



data.istex.fr

data.istex.fr expose le Triple Store des données ISTEX via son SPARQL Endpoint.

[← Named entity / PlaceName entity](#)

ENTITÉ NOMMÉE



Portsmouth

CETTE ENTITÉ DANS ISTEX

40409 results



Portsmouth's tax installment plan

Harold I. Baumes

1932, National Municipal Review - Journal



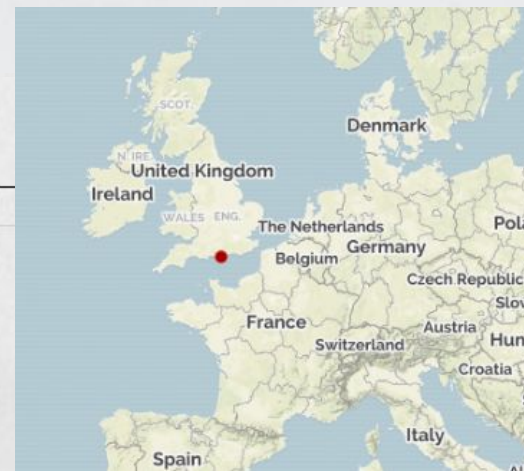
The joint effects of an evaporation duct and high k-factor conditions on a microwave transhorizon path over the sea

Xiandong Shen;Enric Vilar

1996, Annales Des Télécommunications - Journal



A COMPUTER-BASED INFORMATION SERVICE



Des terminologies

Travaux en cours:

- ★ Extraction de termes du thésaurus COVID sur le corpus ISTE_X SARS-MERS
- ★ Extraction de définitions de concept terminologie dans corpus de document ISTE_X

Thésaurus Covid-19
Langue des données : français
Chercher

Accueil / Explorer / Naviguer / Vocabulaires / Thésaurus Covid-19 Aide | English español

Liste	Hiérarchie				
<ul style="list-style-type: none"> cascade immuno inflammatoire inflammation réponse immunitaire <ul style="list-style-type: none"> anticorps <ul style="list-style-type: none"> anticorps anti-SRAS-CoV-2 anticorps neutralisant auto-anticorps anticorps antiphospholipides isotype paratope séroconversion titre d'anticorps antigène cellule immunitaire immunité protéine de la réponse immunitaire <ul style="list-style-type: none"> immunoglobuline <ul style="list-style-type: none"> anticorps <ul style="list-style-type: none"> anticorps anti-SRAS-CoV-2 anticorps neutralisant auto-anticorps anticorps antiphospholipides isotype paratope séroconversion titre d'anticorps chaîne protéique des immunoglobulines immunoglobuline membranaire inflammasome peptide anti-microbien diagnostic interaction virus-hôte maladie émergente maladie infectieuse 	<p>... > réponse immunitaire > protéine de la réponse immunitaire > immunoglobuline > anticorps > auto-anticorps</p> <p>cascade immuno inflammatoire > réponse immunitaire > anticorps > auto-anticorps</p> <p>TERME PRÉFÉRENTIEL auto-anticorps </p> <hr/> <p>DÉFINITION(S)</p> <p>Un auto-anticorps est un anticorps produit par le système immunitaire et dirigé contre une ou plusieurs protéines de l'individu lui-même. (Wikipédia)</p> <p>anticorps</p> <hr/> <p>CONCEPT(S) GÉNÉRIQUE(S)</p> <p>anticorps antiphospholipides</p> <hr/> <p>CONCEPT(S) SPÉCIFIQUE(S)</p> <p>anticorps autologue</p> <hr/> <p>SYNONYME(S)</p> <p>autoanticorps</p> <hr/> <p>TRADUCTIONS</p> <p>autoantibody anglais</p> <hr/> <p>URI</p> <p>http://data.loterre.fr/ark:/67375/C0X-TTK16HVN-Q </p> <p>Télécharger ce concept: RDF/XML Turtle JSON-LD</p> <hr/> <p>EQUIVALENT(S) EXACT(S)</p> <table style="width: 100%; border: none;"> <tr> <td style="border: none;">https://en.wikipedia.org/wiki/Autoantibody</td> <td style="border: none; text-align: right; font-size: small;">en.wikipedia.org</td> </tr> <tr> <td style="border: none;">https://fr.wikipedia.org/wiki/Auto-anticorps</td> <td style="border: none; text-align: right; font-size: small;">fr.wikipedia.org</td> </tr> </table>	https://en.wikipedia.org/wiki/Autoantibody	en.wikipedia.org	https://fr.wikipedia.org/wiki/Auto-anticorps	fr.wikipedia.org
https://en.wikipedia.org/wiki/Autoantibody	en.wikipedia.org				
https://fr.wikipedia.org/wiki/Auto-anticorps	fr.wikipedia.org				

[CONTACT](#) | [MENTIONS LÉGALES](#) | [COOKIES](#) | [SITE MAP](#)

Sémantisation

Rendre les données interopérables dans le web de données

Interrogation BNF x +

https://data.bnf.fr/sparql

```

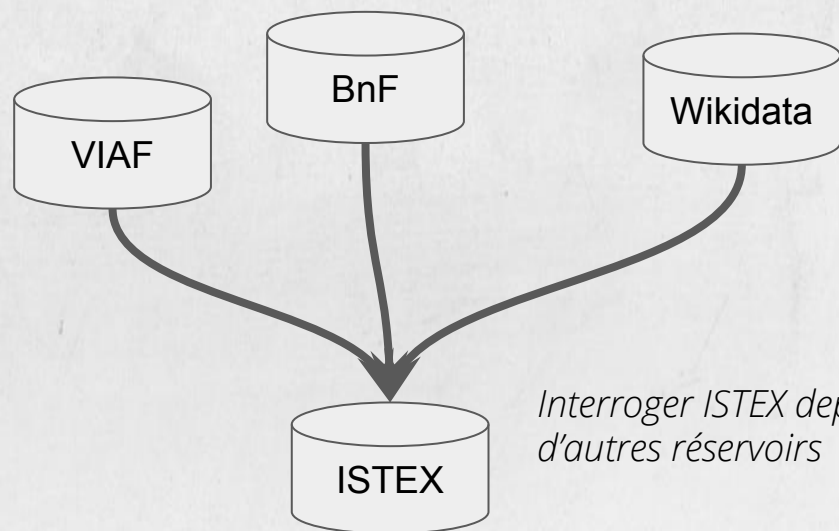
1 # Affiche des informations de la BNF sur le concept « Bioinformatique », dont la
2 # notice RAMEAU, les termes génériques et spécifiques, des alignements
3 # extérieurs... Cette requête étend la requête précédente à d'autres champs.
4
5 # Il faut interroger https://data.bnf.fr/sparql au lieu de https://data.istex.fr/sparql/
6
7+ PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
8 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
9 # Affiche l'URL de la notice RAMEAU, le libellé alternatif,
10 SELECT ?NoticeRAMEAU
11 # GROUP_CONCAT concatène des valeurs différentes d'une variable
12 # (?AutreNom) dans une autre variable (?AutresNoms).
13 # DISTINCT retire les doublons dans un ensemble de valeurs
14 # (toutes les valeurs de la variable ?AutreNom).

```

Table Response Pivotal Table Google Chart Geo

Showing 1 to 1 of 1 entries

NoticeRAMEAU	AutresNoms	TermePlusLarge	TermePlusPrecis	TermeRelie	AlignementBnf
https://catalogue.bnf.fr/ark:/12148/cb144116742	Bio-informatique, Bioinformatica, Biologie Informatique, Informatique biologique	Biologie -- Informatique	Bioinformatique structurale && Chimérogénomique && Immunoinformatique && Informatique physiologique	"Biologie systémique" @fr	http://d-nb.info/gnd/4611085-9 && http://datos.bne.es/resource/XX550402 && http://dewey.info/class/570/ && http://dewey.info/class/621/ && http://purl.org/bnfdid/38863



Interroger ISTEX depuis d'autres réservoirs

Caractérisation des données

Permettre l'interrogation sur la forme des documents

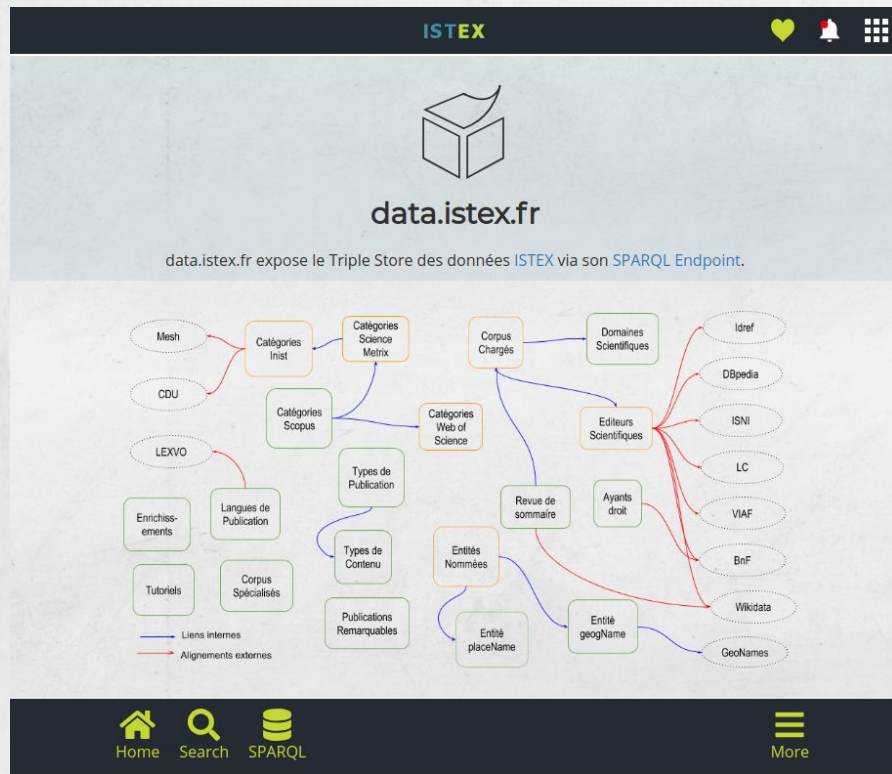
- ★ Score de qualité
- ★ Qualité des PDF
- ★ Nombre de mots
- ★ Présence d'enrichissements

Services dédiés au TDM

 data.istex.fr

Des angles vues différents sur les données

- ★ enrichissements,
- ★ catégories,
- ★ entités nommés
- ★ ...



The screenshot shows the website interface for data.istex.fr. At the top, there is a navigation bar with the ISTEX logo, a heart icon, a notification bell, and a grid icon. Below the navigation bar is a large header area with the ISTEX logo and the text "data.istex.fr". Underneath, it states "data.istex.fr expose le Triple Store des données ISTEX via son SPARQL Endpoint."

The main content area features a conceptual diagram illustrating the data structure. The diagram consists of various nodes connected by lines. Blue lines represent "Liens internes" (internal links), and red lines represent "Alignements externes" (external alignments). The nodes include:


- Mesh, CDU, LEXVO (connected to Catégories Inist)
- Enrichissements, Tutoriels, Langues de Publication, Corpus Spécialisés (connected to Catégories Inist)
- Types de Publication, Types de Contenu, Publications Remarquables (connected to Catégories Inist)
- Corpus Science Metrix (connected to Catégories Inist)
- Corpus Chargés (connected to Catégories Science Metrix)
- Domaines Scientifiques (connected to Corpus Chargés)
- Editeurs Scientifiques (connected to Corpus Chargés)
- Revue de sommaire, Ayants droit (connected to Corpus Chargés)
- Entités Nommées (connected to Corpus Chargés)
- Entité placeName, Entité geogName (connected to Entités Nommées)
- Idref, DBpedia, ISNI, LC, VIAF, BnF, Wikidata, GeoNames (connected to Editeurs Scientifiques)

At the bottom of the page, there is a navigation bar with icons for Home, Search, SPARQL, and More.



Téléchargement massif de données

- ★ prêtes à l'emploi
- ★ sur mesure
- ★ sans limite
- ★ usage local



Téléchargez un corpus ISTEX

Vous êtes membre de l'Enseignement supérieur et de la Recherche et vous souhaitez extraire un corpus de documents ISTEX?
3 étapes suffisent pour récupérer une archive zip sur votre disque dur.

- #### 1. Requête

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

Recherche classique
Recherche par ARK


Nombre de caractères restants : 67 000 ✓

Limite du nombre de documents souhaités : 3000

Choisir les documents classés : Par pertinence Aléatoirement
- #### 2. Formats et types de fichiers

Créez votre sélection en cochant ou décochant les cases ci-dessous :

<input type="checkbox"/> Texte intégral	<input type="checkbox"/> Métadonnées	<input type="checkbox"/> Enrichissements
<input type="checkbox"/> PDF	<input type="checkbox"/> XML	<input type="checkbox"/> multicat
<input type="checkbox"/> TEI	<input type="checkbox"/> MODS	<input type="checkbox"/> nb
<input type="checkbox"/> TXT	<input type="checkbox"/> Annexes	<input type="checkbox"/> refBibs
<input type="checkbox"/> OCR	<input type="checkbox"/> Couvertures	<input type="checkbox"/> teeft
<input type="checkbox"/> ZIP		<input type="checkbox"/> unitex
<input type="checkbox"/> TIFF		
- #### 3. Télécharger

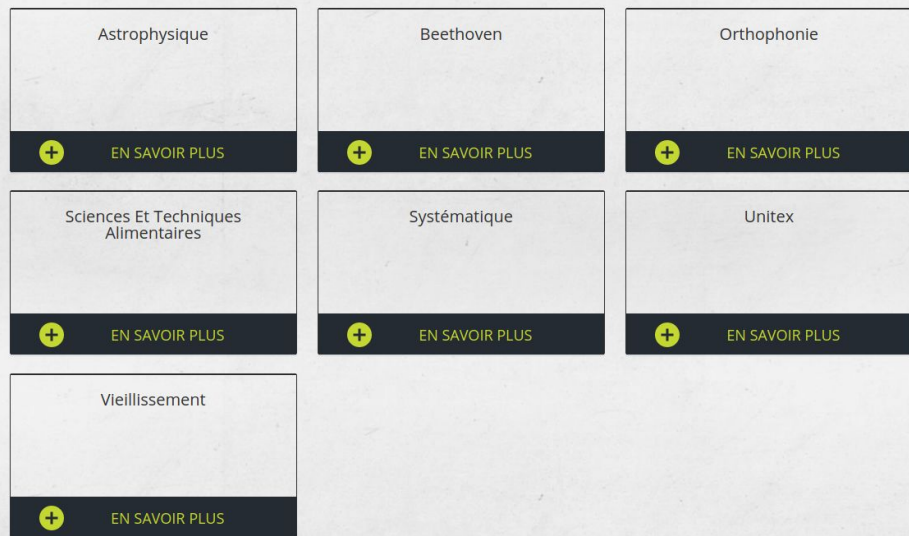


- [Exemples](#)
- [Réinitialiser](#)
- [Récupérer](#)
- [Partager](#)
- [Historique](#)

 corpus.istex.fr

Des corpus thématiques ou d'actualités réalisés à l'Inist sur des données ISTEX

- ★ [Coronavirus : SRAS MERS](#)
- ★ [Corpus Ludwig van Beethoven](#)
- ★ et bien d'autres ...

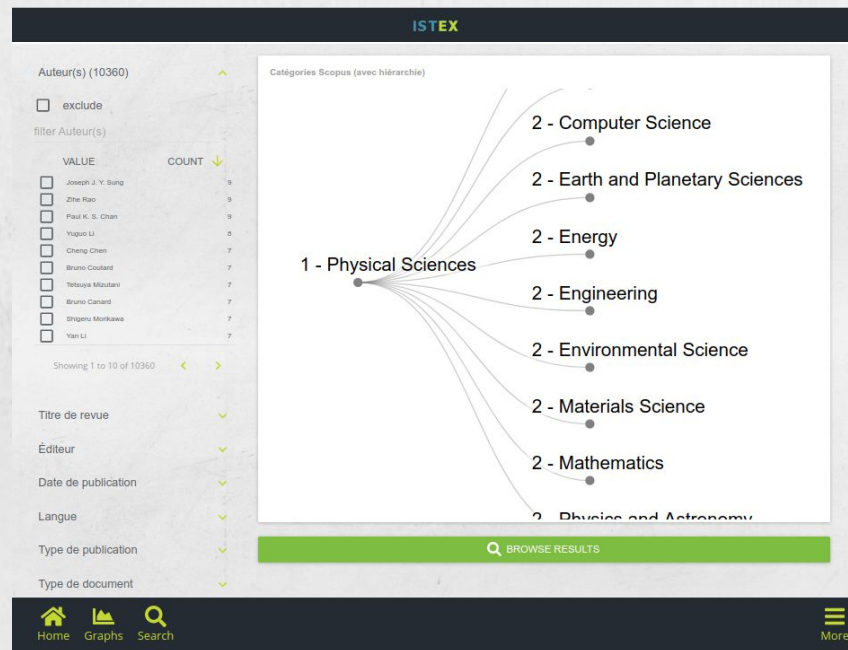




Lodex

Un outil libre d'analyse et d'exploration de données

- ★ Exploration de corpus
- ★ Visualisation de données
- ★ Sémantisation



<https://lodex.inist.fr/>

Services TDM connectés à ISTEX

Gargantext

The image displays the Gargantext software interface. On the left, a dialog box titled "Add a Corpus" is open, showing a list of databases to select from. The "ISTEX" option is highlighted with a red circle. The main window shows a network graph with nodes and edges, representing relationships between concepts. The graph is centered on "bees" and includes nodes like "field", "loss", "winter", "disease", "pesticides", "hives", "field conditions", "greenhouse", "pollination", "viruses", "mortality", "genes", "development", "mortality", "pollen", "genes", "development", "mortality", "pollen", "genes". The right sidebar shows a search result for "data / viruses / winter / bees / pollination" with a list of related terms and a snippet of text from a publication.

Add a Corpus

Type: Select a database below
Name: Select a database below
File: in fichier choisi

- Europepress
- Jstor [RIS]
- Pubmed [XML]
- Scopus [RIS]
- Web of Science [ISI]
- Zotero [RIS]
- ISTEX
- REPEC [RIS]

Network Graph

Nodes: 2 (min) 29 (max)
Edges: 8 (min) 21 (max)

Search Results

data / viruses / winter / bees / pollination

Neighbors

- honey bees
- apis mellifera
- colonies
- infection
- hives
- honeybees
- field conditions
- field
- bacteria
- plants
- adult bees
- loss
- sacbrood virus
- pollen
- abpv
- worker bees
- varroa mites
- health
- pennsylvania

Read More

Putz (5/48)

RNA viruses in hymenopteran pollinators: evidence of inter-taxa virus transmission via pollen and potential impact on non-Apis hymenopteran species...
Published in *PLoS one*...
In 2011-01-04

Systemic spread and propagation of a plant-pathogenic virus in European honeybees, *Apis mellifera*...
Published in *mbio*...
In 2014-01-22

The habitat disruption induces immune-suppression and oxidative stress in honey bees...
Published in *Ecology and evolution*.



Cortext

à venir...



∞
dashboard

queued scripts
🤖

Topic Modeling->/a...	2019-06-03 11:59:32
Data Parsing->astr...	2019-06-03 11:54:45

Create project

🔍

test2 👤 1

🤖 topic modeling->/astro2.db-1559556279018 queued 🟩 2019-06-03 11:59:32
📄 data parsing->astro2.csv-1559555982192 finished 🟩 2019-06-03 11:54:45
📄 astro2.csv finished 🟩 2019-06-03 11:54:22

projects
📁

test2 👤 1
* test 👤 1

CILLEX

The screenshot displays the CILLEX interface, which is a network-based search tool. On the left, a navigation sidebar includes links for 'cillex', 'help', 'chat with us', 'présentation', 'démonstrateur', 'graphsearch', and 'dev'. The main area features a large network graph with nodes and edges, where a central cluster is highlighted in red and labeled 'multidisciplinary'. A search bar at the bottom of the graph contains the text 'nlp graph'. On the right, a detailed view of an article is shown, titled 'High pressure effects in anaesthesia and narcosis'. The article's metadata includes its ARK ID, authors, categories, document URL, DOI, fulltext URL, genre, ID, keywords, original genre, PMID, publication date, and a BibTeX score. The abstract text is partially visible, discussing the effects of high pressure on biological functions.

Navigation Sidebar:

- cillex
- help
- chat with us
- présentation readme
- démonstrateur ISTEX
- graphsearch CILLEX
- dev

Document Details Panel:

High pressure effects in anaesthesia and narcosis
article

+10 explore neighbors 4

ark
ark:/67373/QHD-BKQ3SF2L-8

auteurs

categories

corpusName rsc-journals

document
<https://api.istex.fr/document/85F4304F58782D6E37BCA9836728E07EBB3553C>

doi
[10.1039/b517771p](https://doi.org/10.1039/b517771p)

fulltext
<https://api.istex.fr/document/85F4304F58782D6E37BCA9836728E07EBB3553C/fulltext.pdf?aid=istex-api-demo>

genre review-article

id
85F4304F58782D6E37BCA9836728E07EBB3553D

keywords

originalGenre TRV

pmid
17003895

publicationDate 2006

refBibAuteurs

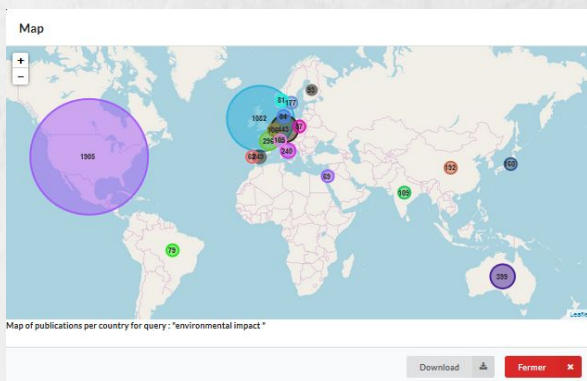
score 89.51592

shape square

text_abstract There is growing interest in determining the effects of high pressure on biological functions. Studies of brain processes under hyperbaric conditions can give a unique insight into phenomena such as nitrogen narcosis, inert gas anaesthesia, and pressure reversal of the effects of anaesthetic and narcotic agents. Such research may shed light on the action of anaesthetics, which remains poorly understood

Footer: 23de64357 7514 8746 nlp graph

Carto ISTE



environmental impact ISTEX Recherche

Environ 1 042 053 résultats (0.85 secondes)

Environmental justice and the distributional deficit in policy appraisal in...
Environmental justice: a particular set of concerns in policy process in water history: what the environmental impacts of a remedial programme or...
G.P Walker.
Journal n°2, page 1-7

Textes complet Metadonnées Enrichissements

Warning: only 1,000 results per cell!

Countries

Search:

Country	Number of publications
United States of America	1905
United Kingdom	1042
Germany	443
Australia	399
France	296

Showing 1 to 5 of 107 entries

1 2 3 4 5 - 22

8679 records(07.86%) do not contain data in the field being analyzed.
Including 5742 records that do not have affiliations.

Laboratories

Search:

Laboratory	Institution	Number of publications
OSAKI RIKKEI NATIONAL LABORATORY	OSAKI RIKKEI NATIONAL LABORATORY	53
UNIVERSITY COLLEGE LONDON	UNIVERSITY COLLEGE LONDON	34
NET PROFESSION LABORATORY	CALIFORNIA INSTITUTE OF TECHNOLOGY	11
ARCTIC CENTRE	UNIVERSITY OF LANSANE	10
LAWRENCE BERKELEY NATIONAL LABORATORY	LAWRENCE BERKELEY NATIONAL LABORATORY	10

Showing 1 to 5 of 7,600 entries

1 2 3 4 5 - 1520

9182 records(61.21%) do not contain data in the field being analyzed.
Including 5742 records that do not have affiliations.

Authors

Search:

Author	Laboratory	Institution	Number of publications
DOMINICKY BOSATO	RHODE ISLAND SCHOOL OF DESIGN		6
MAGDOLINA BAVAN ASSELI	INTERNATIONAL CENTRE FOR INTERGATIVE STUDIES	MAASTRICHT UNIVERSITY	6
E GRAHAM COOK	DEPARTMENT OF CHEMISTRY	PURDUE UNIVERSITY	5
OWEN T BUTLER	HEALTH AND SAFETY LABORATORY		4
ALLIA LASKIN	PACIFIC NORTHWEST NATIONAL LABORATORY	PACIFIC NORTHWEST NATIONAL LABORATORY	4

Showing 1 to 5 of 17,312 entries

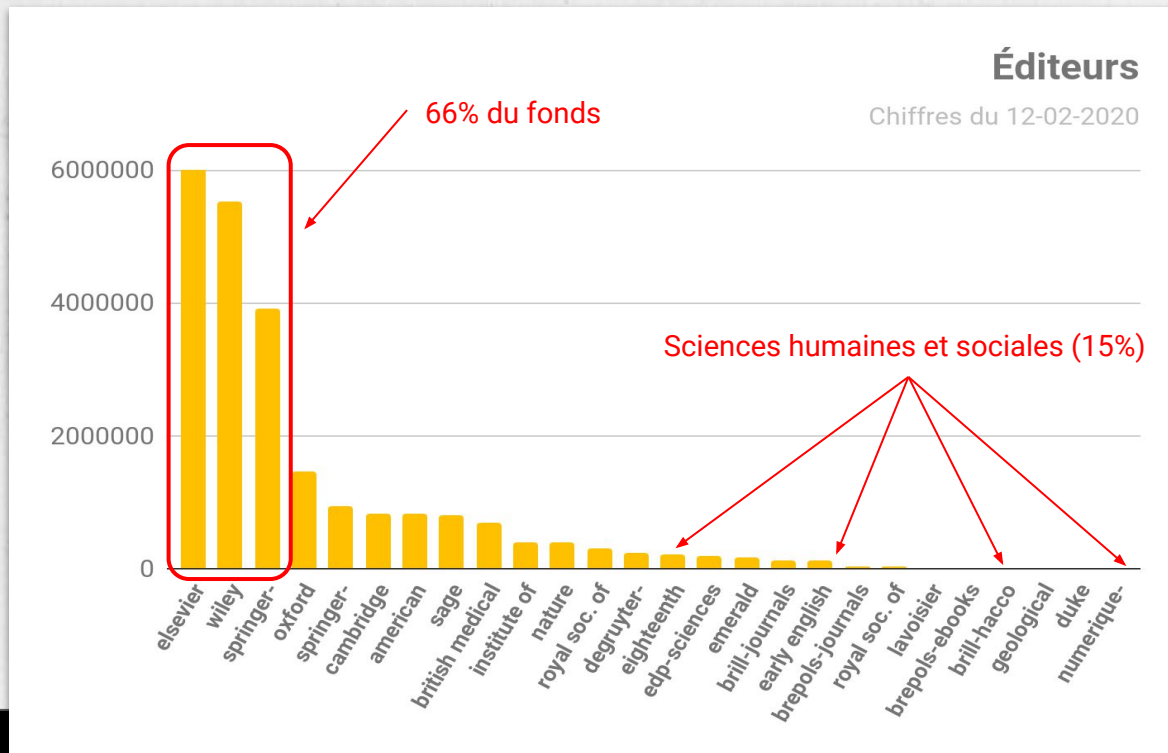
1 2 3 4 5 - 3463

9182 records(61.21%) do not contain data in the field being analyzed.

ISTEX par son contenu

Le fonds ISTEX > répartition par provenance

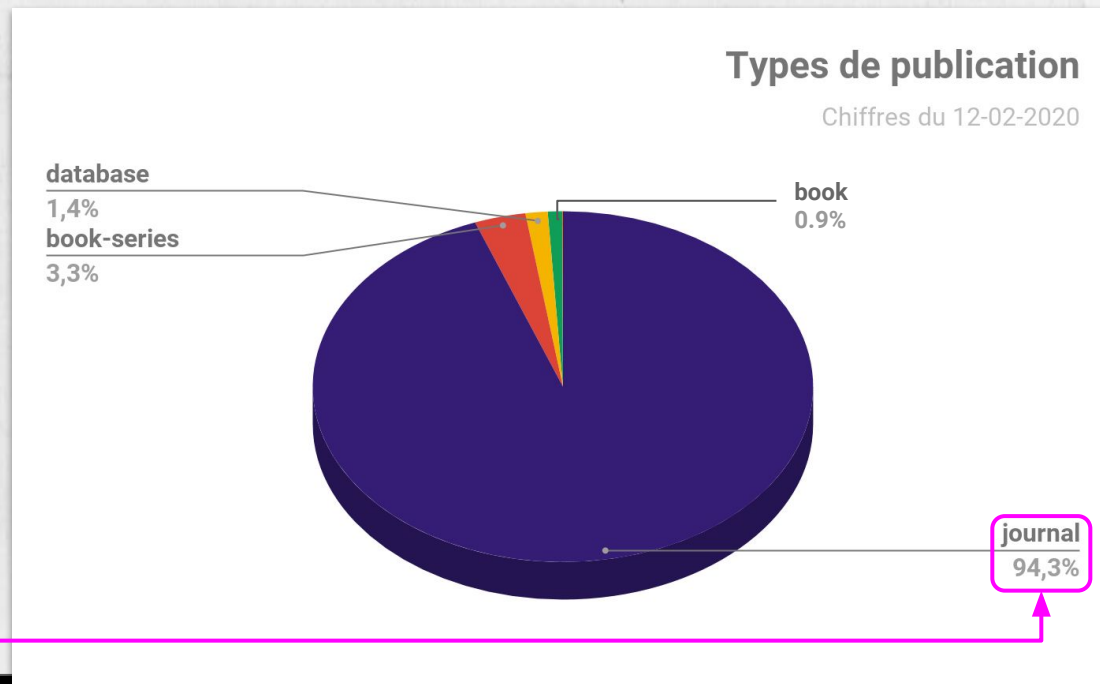
Documents provenant de 26 corpus éditeurs à la date du 13 décembre 2019, mais graphique du 20 septembre 2019



Le fonds ISTEX > répartition par types de publication

Prépondérance des documents issus de **revues**

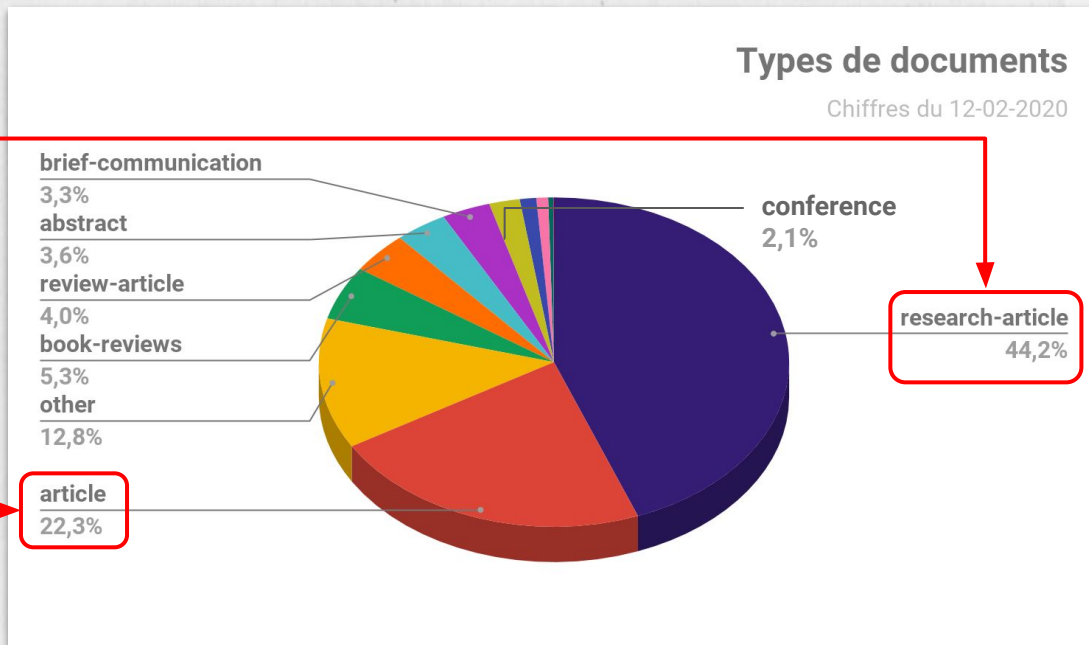
Définitions et chiffres sur data.istex.fr :
Types de publication



Le fonds ISTEX > répartition par types de contenu

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

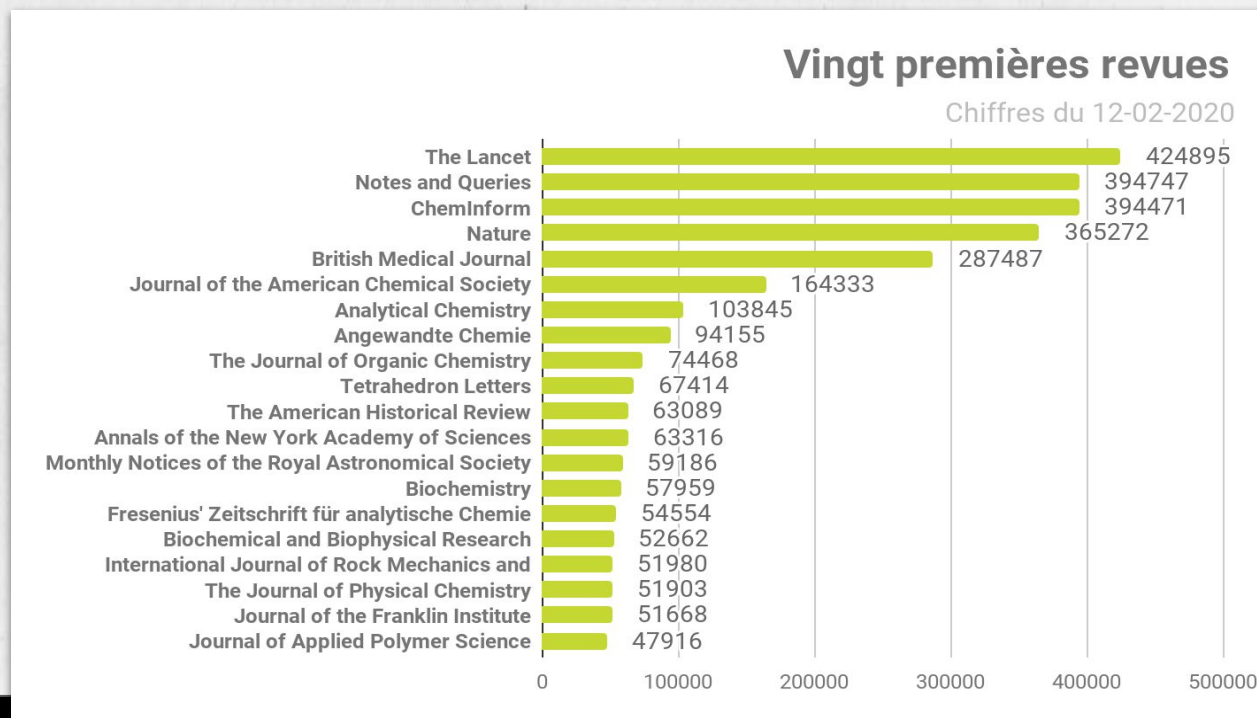
Définitions et chiffres sur data.istex.fr :
Types de contenu



Le fonds ISTEX > répartition par revues

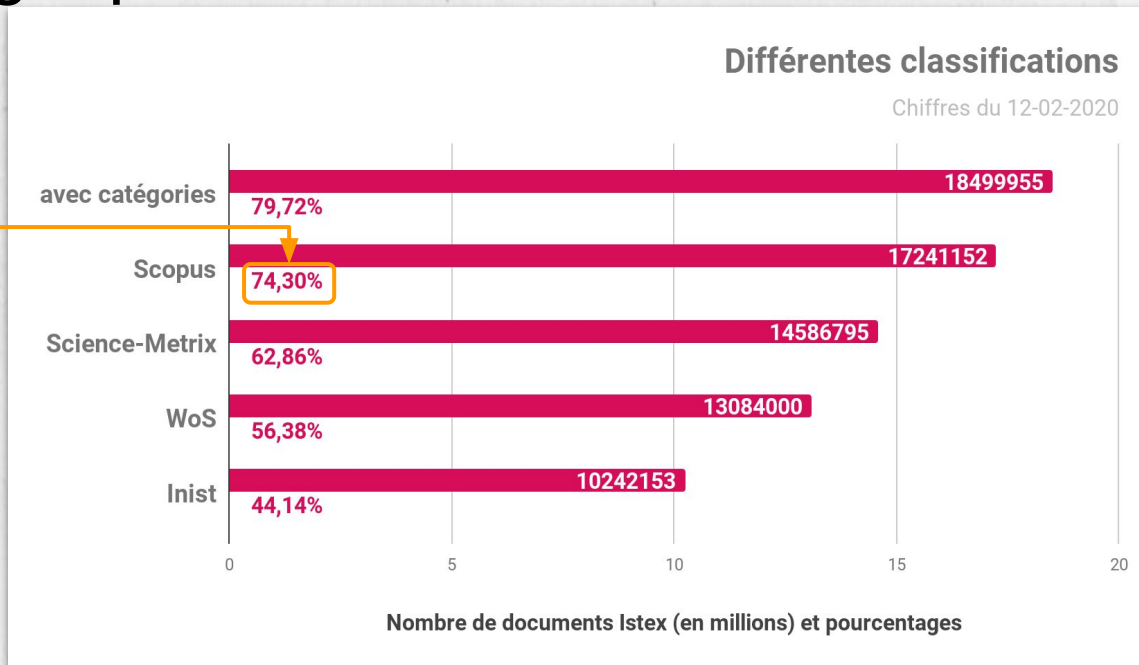
Dans le fonds de plus de **9 000** revues présentes dans ISTEX :

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



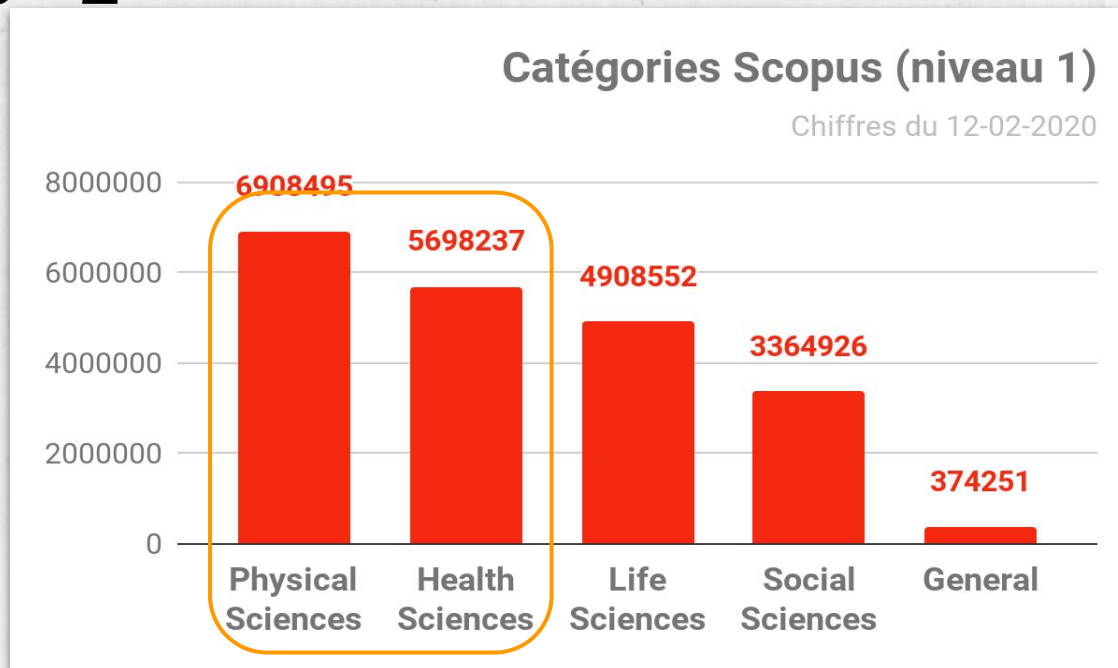
Le fonds ISTEX > répartition par domaines scientifiques - 1

Parmi les documents qui comportent au moins une catégorie scientifique (75%), **Scopus** est la classification la plus représentative



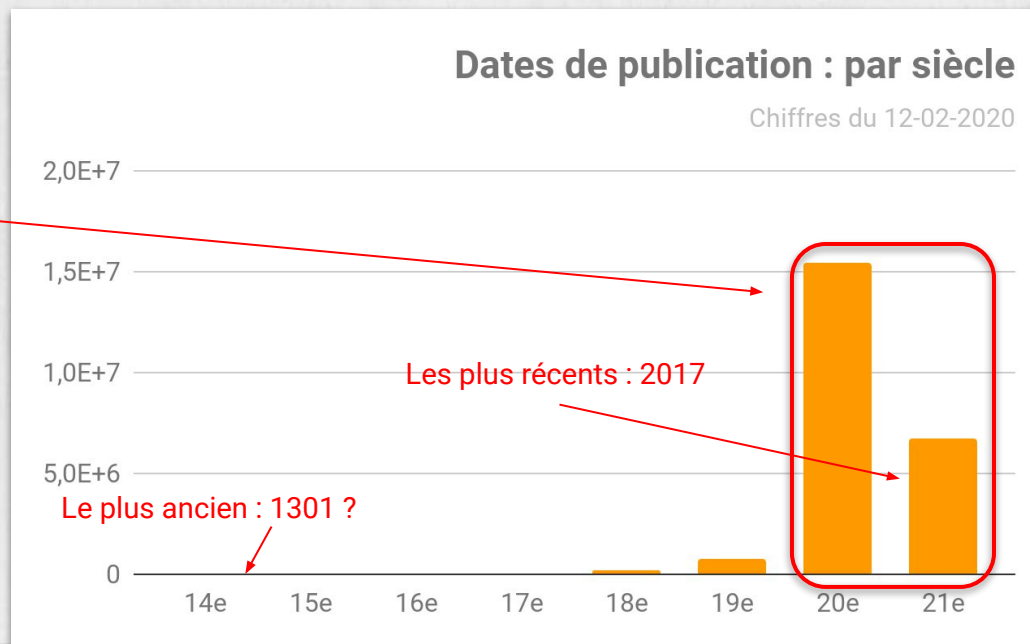
Le fonds ISTEX > répartition par domaines scientifiques - 2

72% des documents Istex catégorisés avec une catégorie Scopus font partie des 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

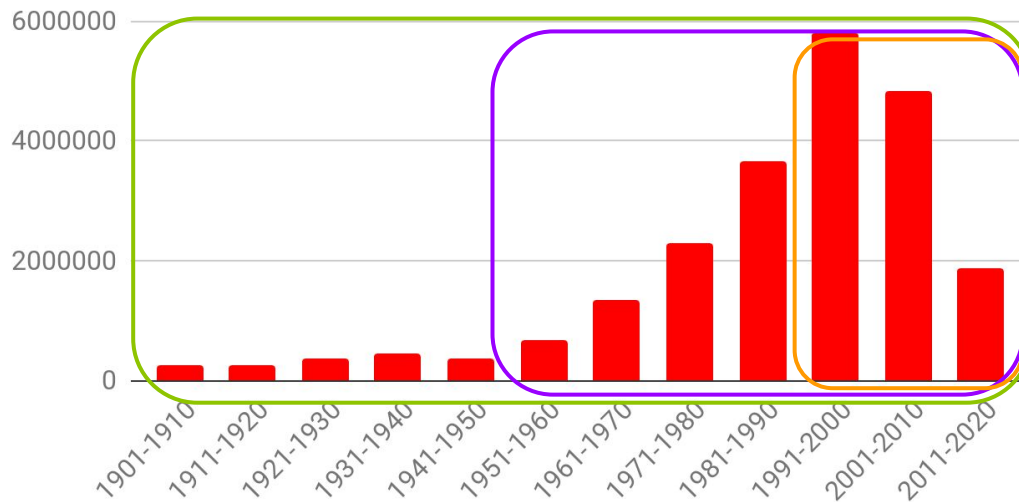
95% des documents publiés
entre 1900 et aujourd'hui
(2017)

88% des documents publiés
depuis 1950

53% des documents publiés
sur les 30 dernières années

Dates de publication : par décennie (20&21e siècles)

Chiffres du 12-02-2020

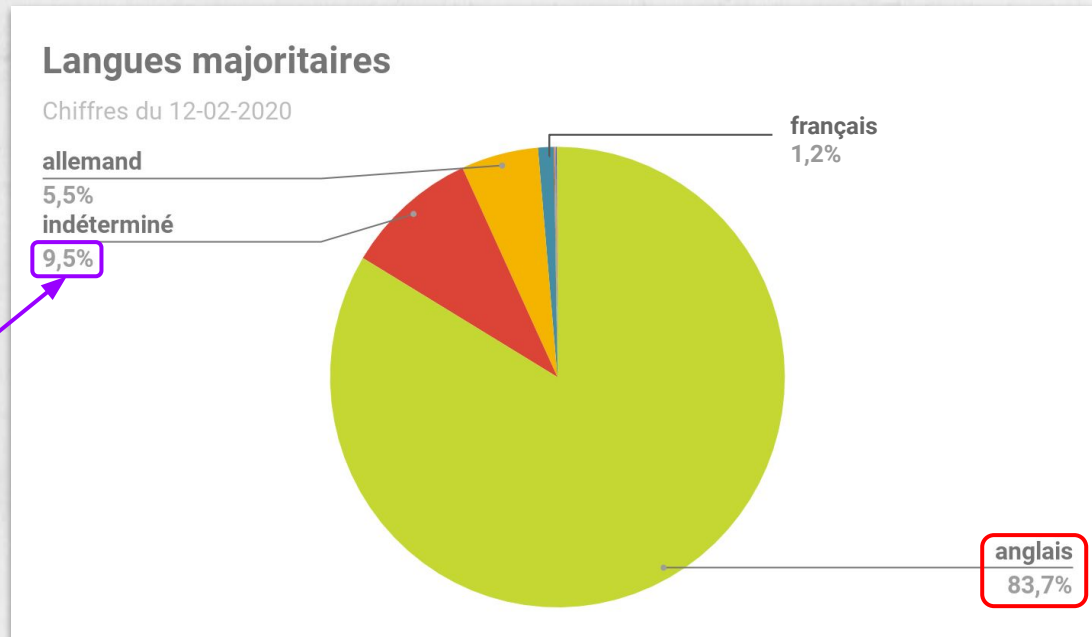


Le fonds ISTEX > répartition par langues - 1

53 langues !

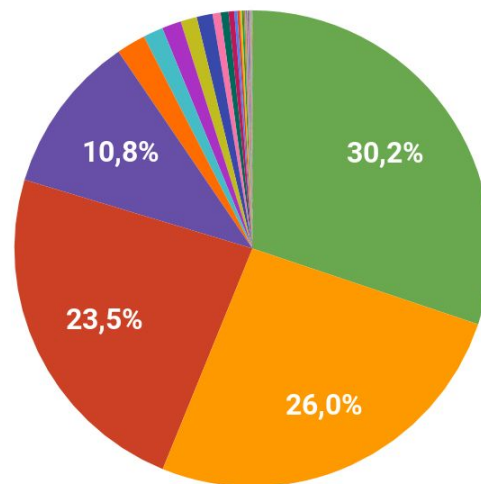
Anglais
majoritaire

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)

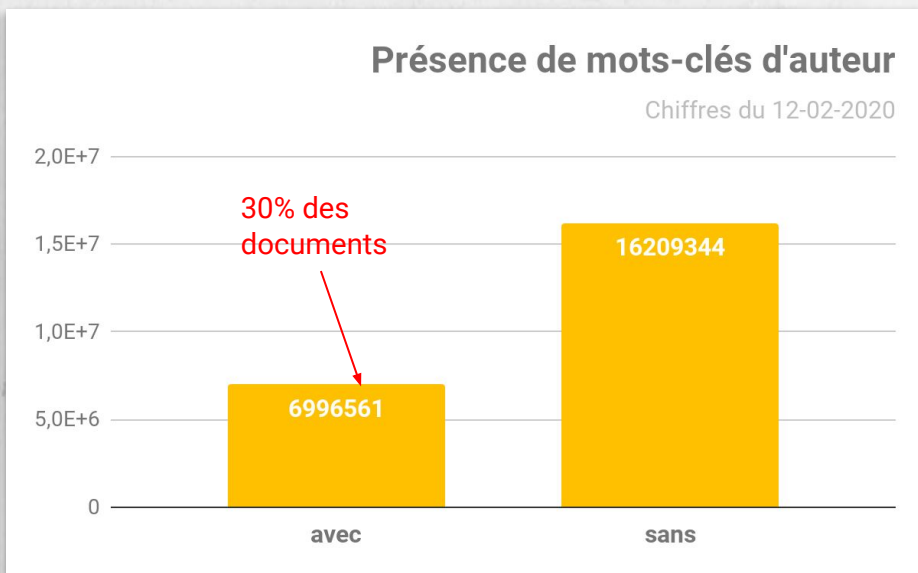
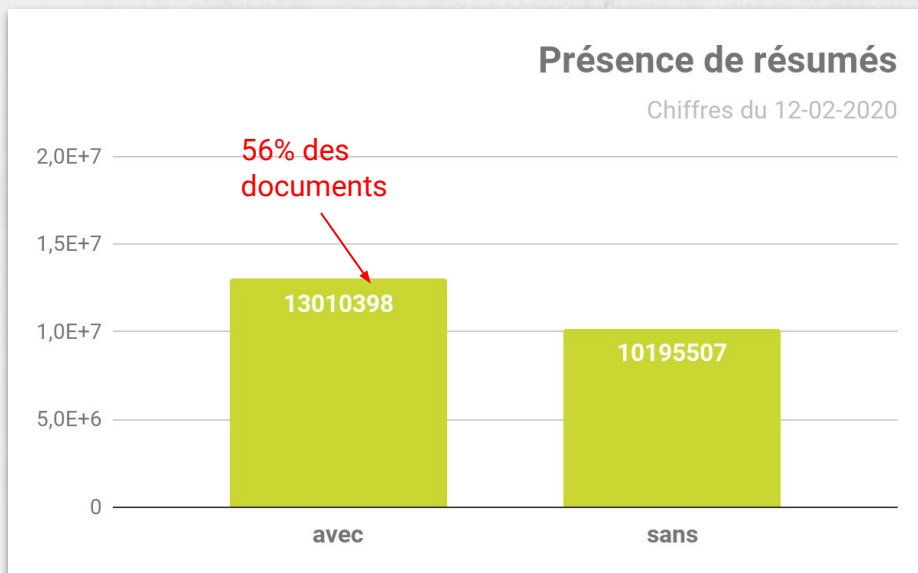


Autres langues

Chiffres du 12-02-2020

- italien
- latin
- espagnol
- néerlandais
- russe
- gallois
- romanes (langues)
- portugais
- 44 autres

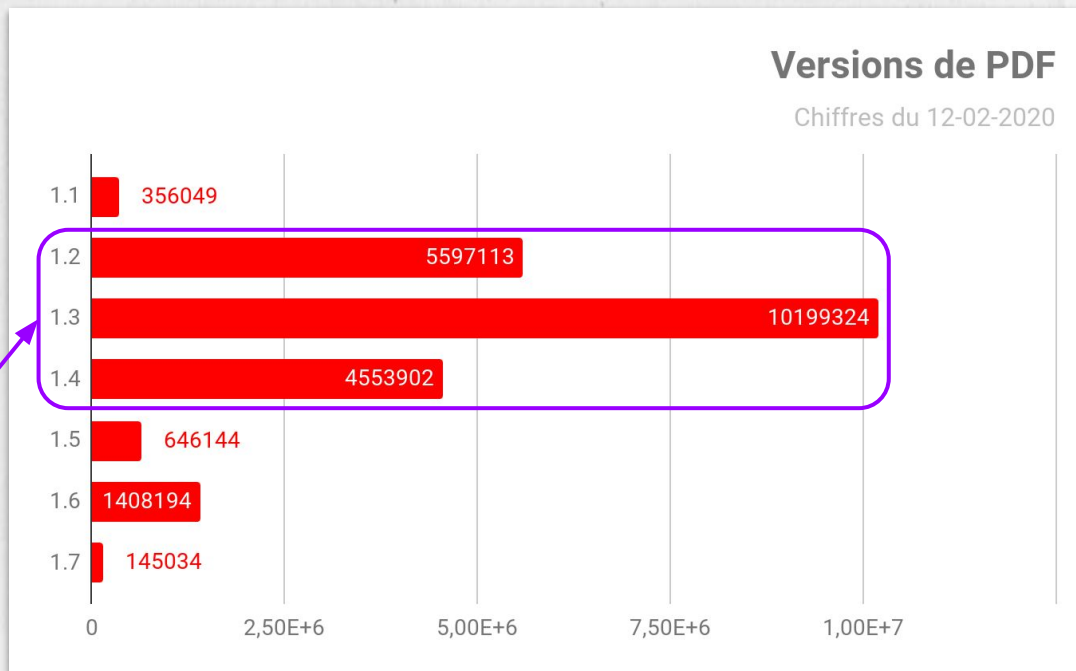
Le fonds ISTEX > répartition par éléments descriptifs du contenu



Le fonds ISTEX > répartition par indicateurs - 1

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

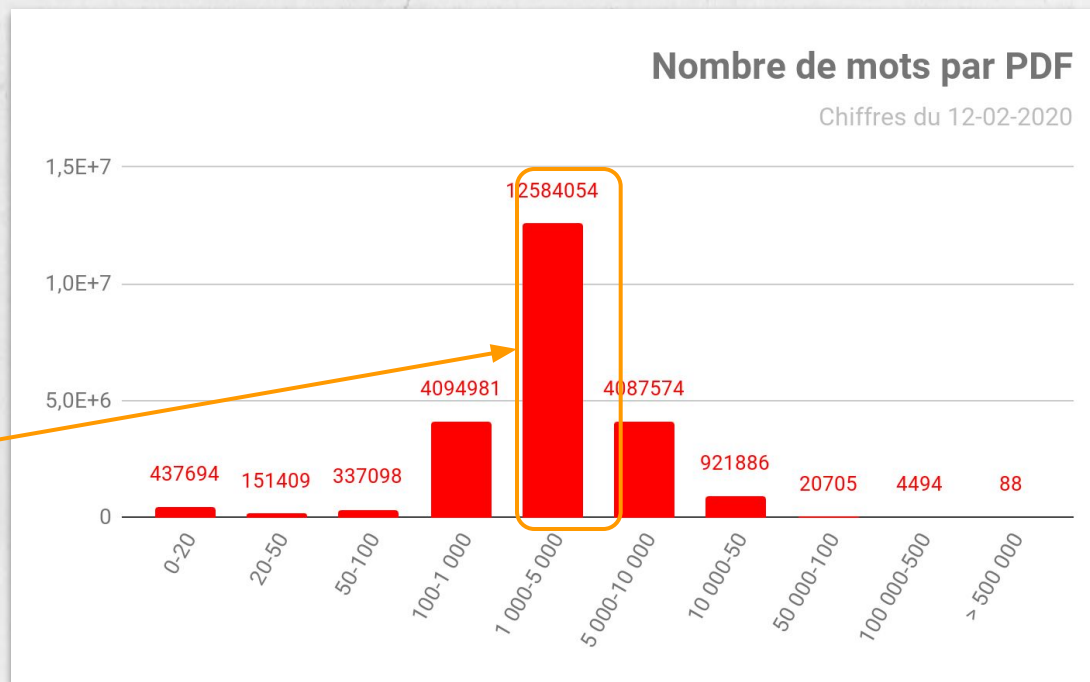
Plus de 88% des documents existent dans une version de PDF comprise entre 1.2 et 1.4



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

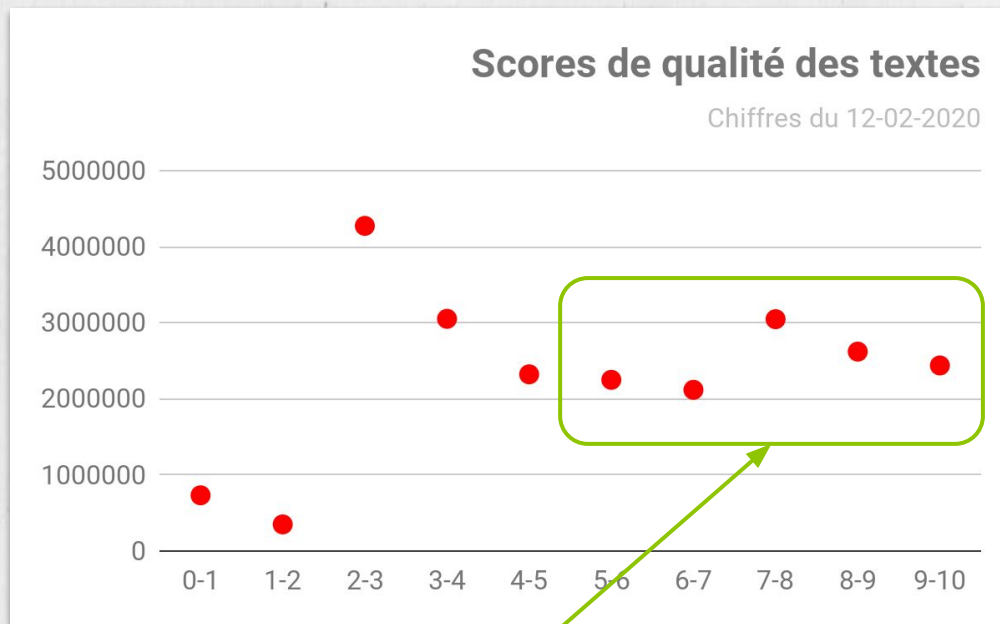


Le fonds ISTEX > répartition par indicateurs - 3

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 (12 417 689) ont un score supérieur à 5

Quelques liens

Contact & Discussion



- ★ Formulaire : <https://www.istex.fr/contact/>
- ★ Mail : contact@listes.istex.fr
- ★ Liste de discussion : users@listes.istex.fr (liste publique)

Actualités & Documentation

- ★ Blog : <https://blog.istex.fr/>
- ★ Twitter : [@ISTEX Platform](https://twitter.com/ISTEX_Platform)
- ★ Documentation : <https://doc.istex.fr>

<https://www.istex.fr>