



# **ISTEX-Usage : Projet OTELO**

**Proposition de chantier thématique sur les Terres rares, de l'exploration à l'impact environnemental**

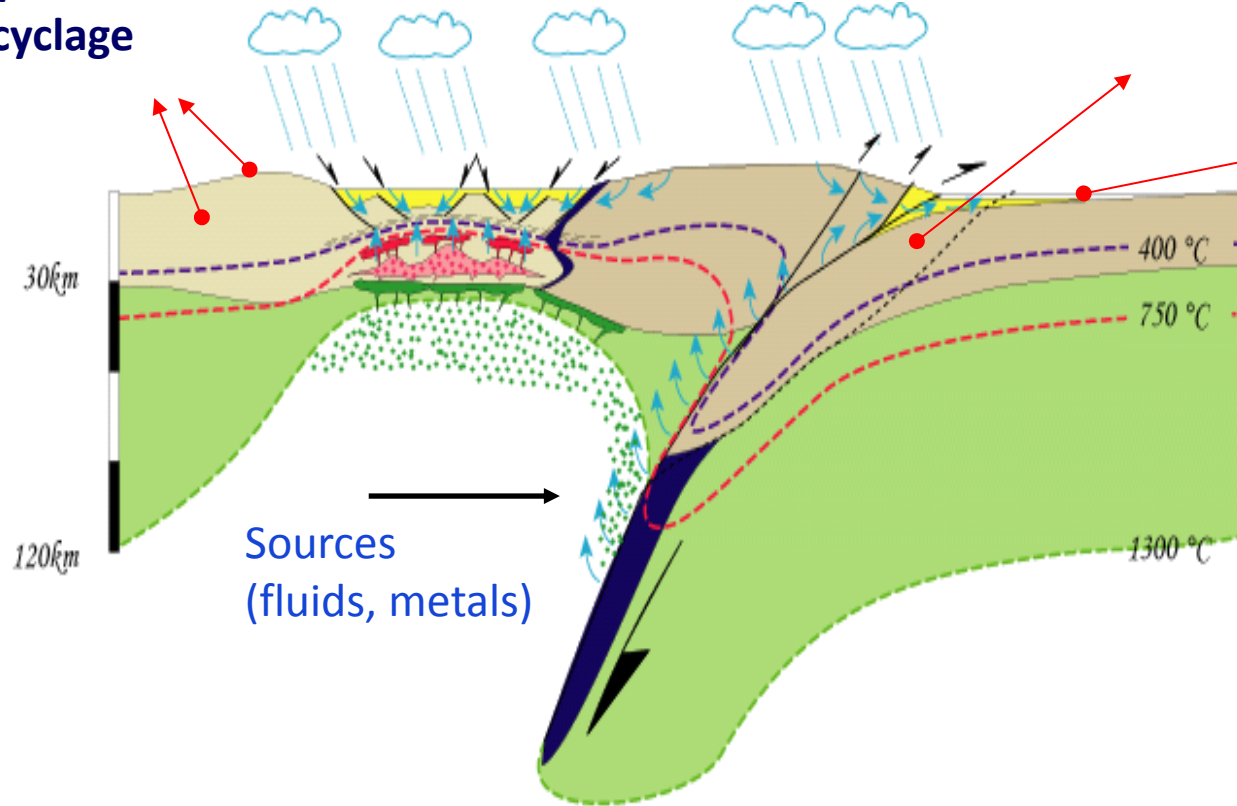
Corinne LEYVAL, Pierre-Yves ARNOULD, Anthony GUIOT



Le projet RESSOURCES21: Laboratoire d'Excellence (LabEx) :  
pour la connaissance et la gestion des métaux stratégiques  
du XXIème siècle.

Du processus concentrateur aux procédés de valorisation et de recyclage et à la protection de l'environnement

**Separation, concentration  
recyclage**



**Impact environnemental  
Capteurs**

**Geologie - Modélisation**

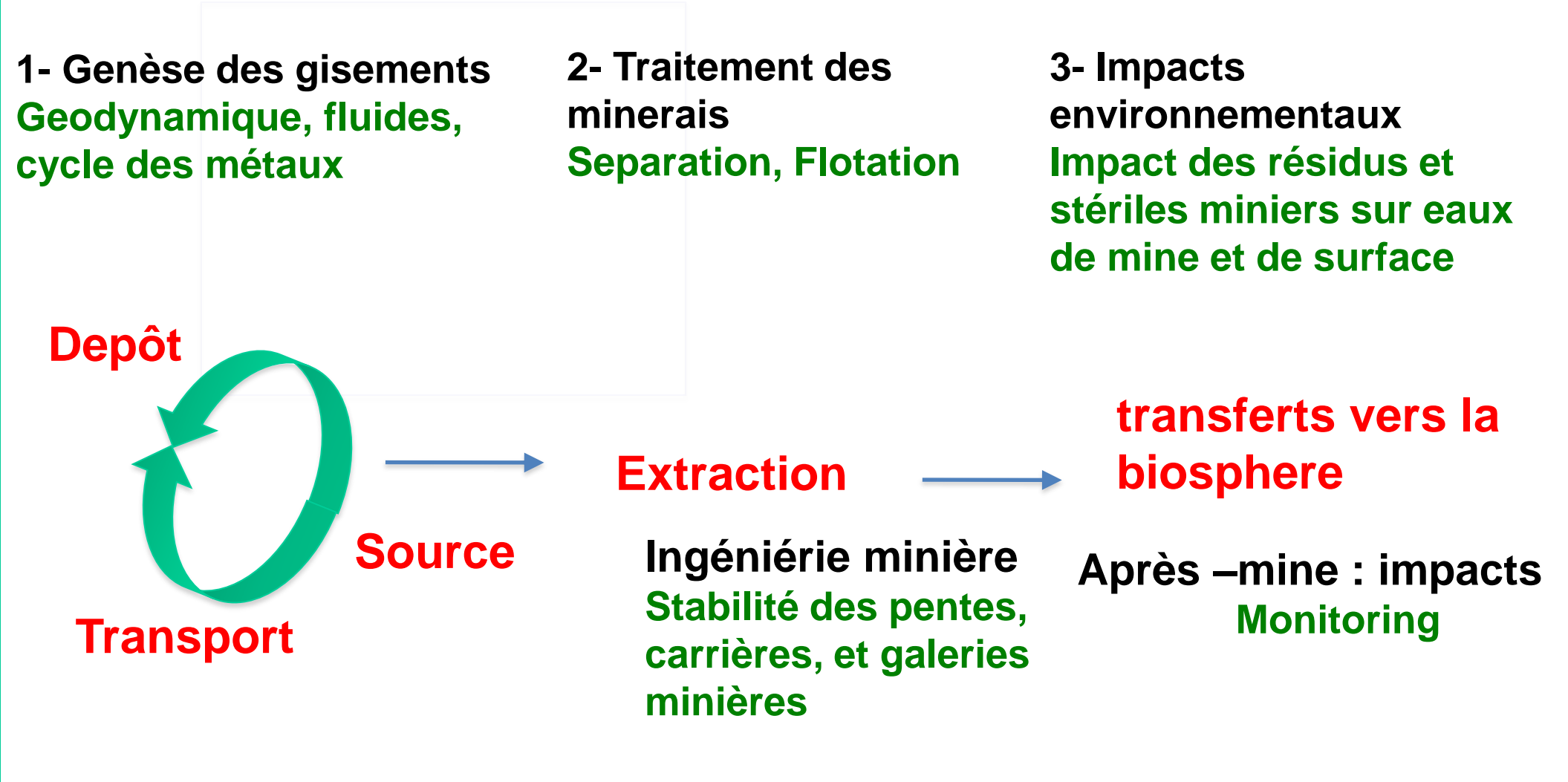
Geometrie/transport  
Analyse multi-élémentaire  
Datation

[www.ressources21.univ-lorraine.fr](http://www.ressources21.univ-lorraine.fr)

**Tutelles :** UL – CNRS – INRA

**UR :** CRPG, GéoRessources, LIEC, LSE

**Partenariat :** VOLTAIRE, BRGM, INERIS, ANDRA, ERAMET, AREVA, IMERYS, RHODIA...





# Projet OTELo :

Proposition de chantier thématique sur les Terres rares, de l'exploration à l'impact environnemental



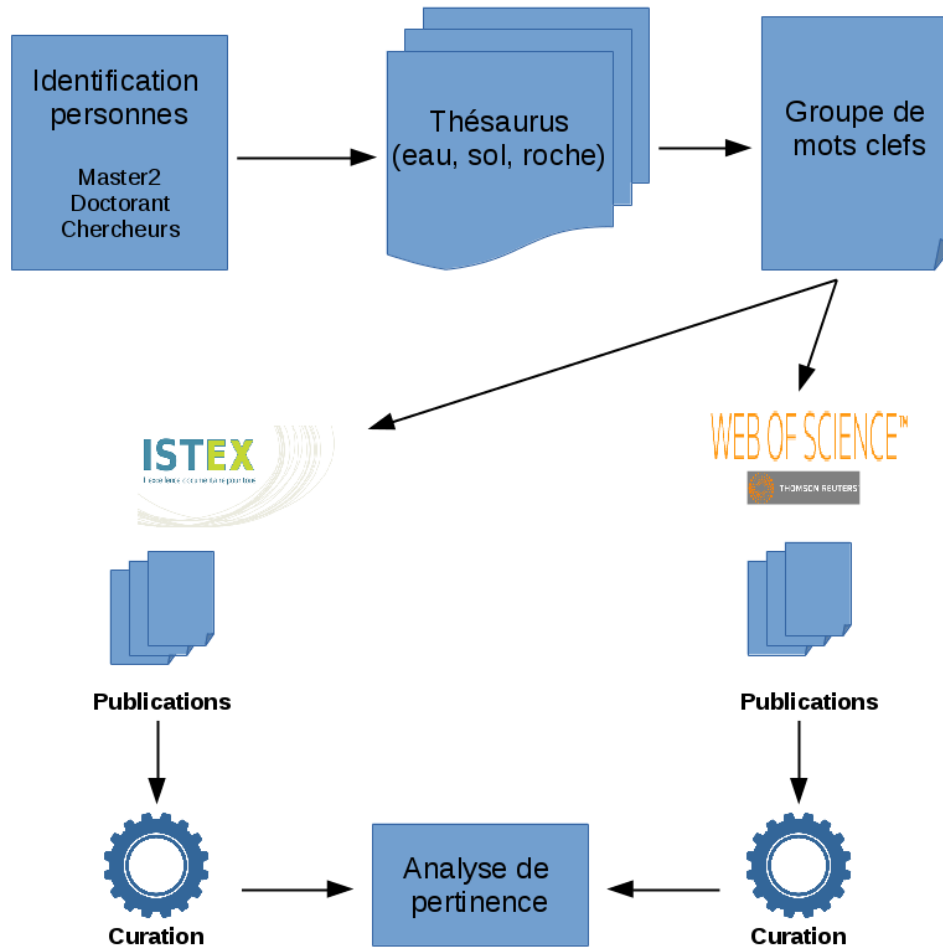


## Objectifs :

- Etudier la possibilité d'enrichir la bibliographie des chercheurs autour du thème des Terres rares par l'interrogation plein texte de ce réservoir.
- Réaliser des cartographies d'informations des pays, auteurs, laboratoires/institution publiant sur les terres rares.



# Méthodologie :





## Personnes impliquées dans le projet

Profils Expérience	Ecotoxicologue	Géologue	Systèmes d'information
<b>Chercheurs DR/Pr</b>	Corinne LEYVAL, LIEC Laure GIAMBERINI, LIEC	Anne-Sylvie ANDRE-MAYER, Georessources	
<b>Chercheurs CR/McF</b>	Damien BLAUDEZ, LIEC		
<b>Post-Doctorant</b>	Ana ROMERO-FREIRE, LIEC		
<b>Doctorant</b>	Nicolas GROSJEAN, LIEC	Francois TURLIN, Georessources	
<b>Master 2</b>		Alexandre Crepon, Georessources	
<b>ITA/BIATSS</b>			Pierre-Yves ARNOULD IE OTELO
<b>CDD ISTEEX</b>			Antony Guiot AI OTELO






## Groupes de mots clefs :

- **Réflexion sur des mots-clefs « terres rares »**
- **En date du 20/01/2017**
  - « Environmental impact » AND « rare earth elements »
  - « Rare earth elements » AND grenville AND quebec
  - « Rare earth elements deposit »
  - « Rare earth elements deposit » AND magmatic







## « Environmental impact » AND « rare earth elements »

	 	
Nb de publications	48	246
Période couverte	1994-2016	1971-2013
Nb publiés > 2013	17	
% publiés > 2013	35 %	
Nb publiés identiques	1	

Plus de résultats dans ISTEX que WOS, malgré une période couverte plus courte et quasiment pas de publications communes






## « Rare earth elements » AND grenville AND quebec

	<b>WOS</b>  THOMSON REUTERS <sup>®</sup>	
Nb de publications	16	259
Période couverte	1996-2015	1923-2013
Nb publiés > 2013	2	
% publiés > 2013	12,5 %	
Nb publiés identiques	0	

Plus de résultats dans ISTEX que WOS, malgré une période couverte plus courte et quasiment pas de publications communes






## « Rare earth elements deposit »

	 	
Nb de publications	5	6
Période couverte	2001-2015	1998-2013
Nb publiés > 2013	1	
% publiés > 2013	25 %	
Nb publiés identiques	1	

Peu de résultats dans ISTEEX et WOS, et quasiment pas de publications communes



## « Rare earth elements deposit » AND magmatic

	 	
Nb de publications	3	19
Période couverte	2004-2009	2001-2013
Nb publiés > 2013	0	
% publiés > 2013	0 %	
Nb publiés identiques	0	



# Constats

- 37 % des publications du WOS traitant du sujet des Terres Rares (d'après le groupe de mots clefs) sont postérieures à 2013

→ Sujet d'étude relativement récent

- Le nombre de publications trouvées dans ISTEEX est en moyenne 6,6 fois supérieur.

- Le taux de recouvrement entre les deux recherches est quasiment nul.





# Curations des publications

## « Environmental impact" AND "rare earth elements" »

### Recherche dans ISTEEX

Publications trouvées : 246

nombre d'articles scientifiques ou article ou « non identifié » : 185

nombre de documents ayant les mots clefs dans la référence bibliographique : 12

nombre de documents ayant "rare earth elements" dans la référence bibliographique : 6

nombre de documents ayant "Environmental impact" dans la référence bibliographique : 30

nombre de documents identifiés comme article mais étant des biblio, index : 53

→ **Publications contenant les mots clefs dans le texte hors refbib : 84**



# Curations des publications

## « Environmental impact" AND "rare earth elements »

### Recherche dans WOS :

Publications trouvées : 48

nombre d'articles accessibles en fulltext : 28

nombre de documents ayant les mots clefs dans la référence bibliographique (pour les articles scientifiques et articles) : 7

→ **Publications contenant les mots clefs dans le texte hors refbib : 21**

→ **Nombre de documents en commun : 1**  
(id: 2FC72EA042E13C09DBAE2D4477DC409A152B011B)



# Curations des publications

« rare earth elements » AND « grenville » AND « quebec »

## Recherche dans ISTEEX

Publications trouvées : 259

nombre d'articles scientifiques ou article ou « non identifié »: 194

nombre de documents ayant tous les mots clefs dans la référence bibliographique : 7

nombre de documents ayant tous "rare earth elements" dans la référence bibliographique : 26

nombre de documents ayant tous "grenville" dans la référence bibliographique : 42

nombre de documents identifiés comme article mais étant des biblio, index : 38

→ **Publications contenant les mots clefs dans le texte hors refbib : 81**



# Curations des publications

« rare earth elements » AND « grenville » AND « quebec »

Recherche dans WOS :

Publications trouvées : 16

nombre d'articles accessible en fulltext : 9

nombre de documents ayant les mots clefs dans la référence bibliographique (pour les articles scientifiques et articles) : 0

→ **Publications contenant les mots clefs dans le texte hors refbib : 9**

→ **Nombre de documents en commun : 0**



# Constats

## Taux de documents (accessibles) contenant les mots clefs cherchés hors bibliographie:

ISTEX : 43 %

- La recherche plein texte dans ISTEX amène un « bruit » important dans les réponses : les mots clefs se trouvent dans les références bibliographique
- Un certain nombre de documents ne sont pas identifiés correctement dans ISTEX (Article en lieu et place de Résumé ou index)

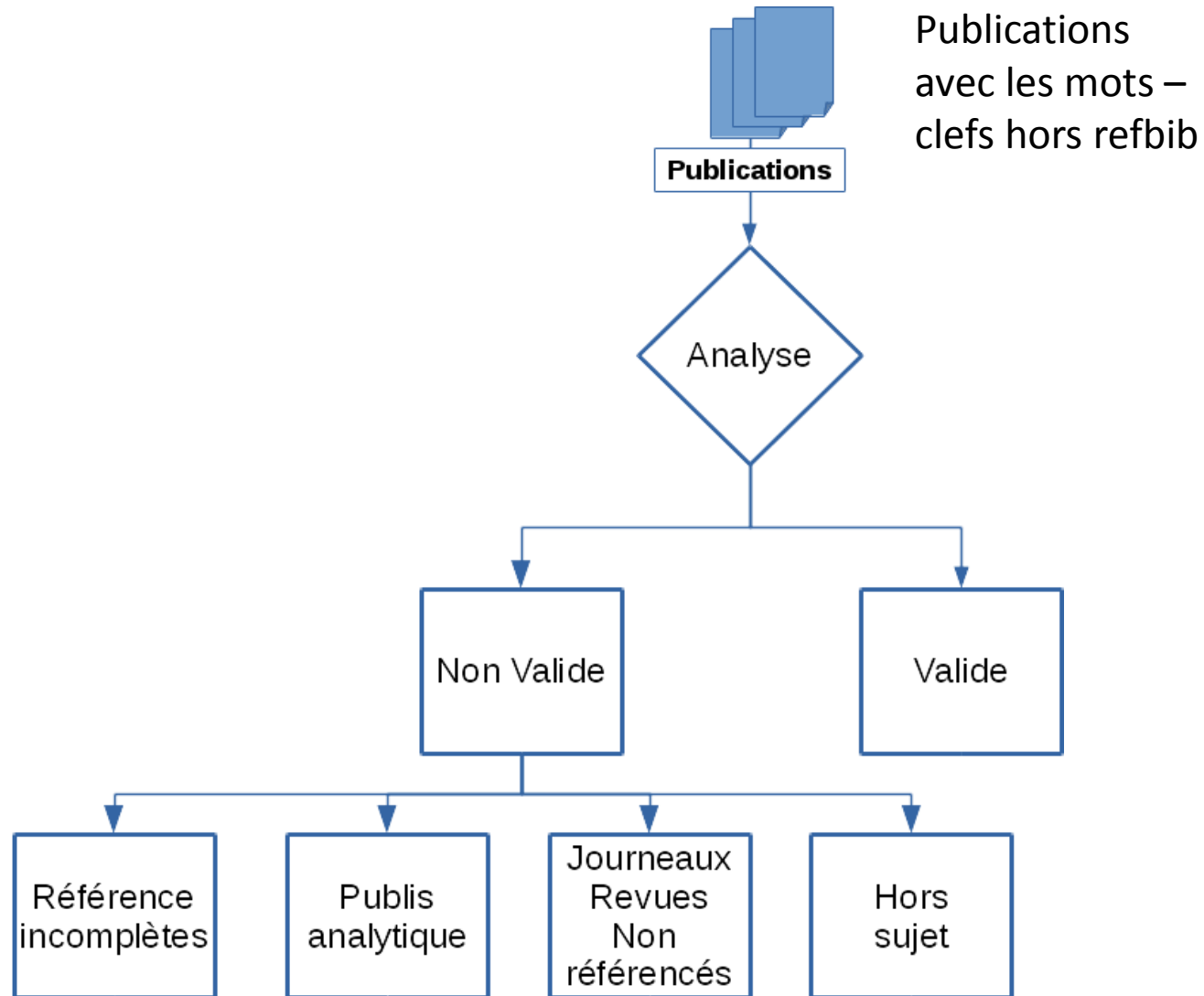
WOS : 47 %

- Le taux relativement faible est dû à l'inaccessibilité des publications en fulltext dans le WOS (URL cassées, restriction Editeur, non abonné)





# Analyse de pertinence par les chercheurs :





# Analyse de pertinence :

**« Environmental impact » AND « rare earth elements »  
(groupe de 5 chercheurs)**

*Une publication est jugée pertinente (valide) ou non pertinente (non valide) si au moins 3 chercheurs (sur 5) l'ont classée ainsi*

**« Rare earth elements » AND « grenville » AND « quebec »  
(groupe de 2 chercheurs)**

*Une publication appartient à l'une des catégories si 2 chercheurs (sur 2) l'ont classée ainsi*



# Analyse de pertinence :

		Publications Pertinentes (valides)	Publications non pertinentes				
			Référence incomplètes	Publis Analytique	Journeaux/ Revues Non référencés	Hors sujet	Total
« environmental impact » AND « rare earth elements »	ISTEX	16	6	12	4	46	84
	WOS	14	0	2	0	5	21
« rare earth elements » AND « grenville » AND « quebec »	ISTEX	7	-	-	-	-	81
	WOS	2	-	-	-	-	9

*Hors sujet : publications de type ingénierie, recyclage, analyse, matériaux spécifiques ...*



# Conclusions

- ✓ Le réservoir ISTEEX: une nouvelle ressource de publications scientifiques
- ✓ Importance de la sélection des mots clefs et leur association pour former les requêtes, et influence de l'évaluateur des résultats
- ✓ Grand nombre de résultats obtenus avec ISTEEX sur les mots-clefs choisis
- ✓ La recherche fulltext est très "large" et nécessite un travail de curation (réf. incomplètes, type de document incorrect, revues non référencées..)
- ✓ Le faible taux de recouvrement entre ISTEEX et WOS pose la question des revues couvertes par les deux portails:
  - La principale revue « rare earth elements/grenville/quebec pas dans ISTEEX.
  - Le réservoir ISTEEX couvre 67% (94 sur 139) des revues du domaine "OTELo" jusqu'à 2013.



# Conclusions - Perspectives

✓ **Sur la thématique des Terres Rares choisie l'usage du réservoir ISTEEX n'a pas permis d'enrichir significativement la bibliographie des chercheurs d'OTELo**

→ Thématique choisie relativement récente, surtout abordée dans des revues indexées WOS ?

✓ **Utilisation des deux portails pour constituer une bibliographie exhaustive**

→ Portail unique bibliographie ISTEEX, WOS (via université de lorraine) et BibCNRS ?

✓ **Plateforme d'interrogation Istex (API Istex) pas adaptée à l'usage des chercheurs**

→ Enrichir les fonctionnalités du widget Istex-SNU (sélection du périmètre de recherche, suggestion sémantique, aide à la syntaxe ...)

✓ **Curation du réservoir**

→ « Nettoyage » des publications aux références incomplètes, revues non réf, type de document mal renseigné.

✓ **Valoriser le réservoir ISTEEX par des outils d'exploration et d'exploitation : cartographie d'information.**





# Carto-ISTEX



# Objectifs

Réaliser des cartographies d'informations à partir du réservoir de publications ISTE<sup>X</sup>

Cartographies des :

- Pays
- Laboratoires/institutions
- Auteurs

Qui publie dans le domaine des Terres rares d'après une recherche dans ISTE<sup>X</sup>



# Présentation de l'application


**L'application est basée sur l'exploitation des affiliations et clef auteur contenues dans les metadonnées des publications.**




Elle est divisée en quatre modules :

- Widget ISTEEX-SNU ( recherche, résultats, facets)
- Module Pays
- Module Laboratoires
- Module Auteurs



# Présentation de l'application

 **Carto Istex**

Help

Rechercher

Environ 27 652 résultats (1.35 secondes)








1 2 3 4 5 6 7 8 9 10 >

Qualité du pdf ▼

**Meteoritic diagenesis of Upper Cretaceous and Paleocene-Eocene shallow-water...**

Résumé : In the central part of the Kruja Platform (Albania) located in the Apulian passive margin, geochemical analyses (calclmetry, Sr, REE and isotopic,  $\delta^{13}C$  and  $\delta^{18}O$ ...

Qualité du PDF: ★★★★★

Texte complet   
Métadonnées  
Enrichissements  

Publié en 2009 | Degruyter-journals | Type : Research-article

**Trace element fractionation during dehydration of eclogites from high-pressure...**

Résumé : The trace element compositions of eclogites, blueschists and mafic granulites from high-pressure terranes have been analysed to investigate element losses and f...

Qualité du PDF: ★★★★★





Publié en 2008 | Elsevier | Type : Research-article

⚠ Warning: only 15000 results parsed!

View on map
Export CSV
Improve

### Countries

Search:

Country	Number of publications
 United States of America	1632
 Japan	1196
 Germany	797
 France	710

View on map
Export CSV

### Laboratories

Search:

Laboratory	Institution	Number of publications
RESEARCH SCHOOL OF EARTH SCIENCES	AUSTRALIAN NATIONAL UNIVERSITY	55
ARGONNE NATIONAL LABORATORY	ARGONNE NATIONAL LABORATORY	45

View on map
Export CSV

### Authors

Search:

Author	Laboratory	Institution	Number of publications
J DOSTAL	DEPARTMENT OF GEOLOGY	SAINT MARY'S UNIVERSITY	28
D.W KURTZ	CENTRE FOR ASTROPHYSICS	UNIVERSITY OF CENTRAL LANCASHIRE	27

### Affiner votre recherche

Corpusname ▼

- Elsevier 13 315
- Wiley 6 132
- Springer 4 196
- Iop 1 548
- Rsc-journals 1 160
- Oup 879
- Nature 139
- Rsc-ebooks 128
- Degruyter-journals 90
- Sage 29



# Principe de fonctionnement

L'interrogation du réservoir ISTEEX se fait grâce à l'API ISTEEX via l'intégration du widget ISTEEX-SNU dans une page WEB.

L'ensemble des résultats de la requête (lot de 5000 réponses, limités à 15000 pour des raisons de performance) est envoyé aux différents modules de cartographie.

*(Les diagrammes de séquences sont disponible à l'adresse suivante :  
[https://github.com/arnouldpy/carto\\_istex/blob/master/Readme.md](https://github.com/arnouldpy/carto_istex/blob/master/Readme.md))*





# Des affiliations de qualité hétérogènes

L'analyse des affiliations a montré qu'elles peuvent être très différentes (pas de respect d'un schéma ex : TEI)

Modules Pays : utilisation du dernier champs de l'affiliation

Modules Auteur : utilisation des clefs author.\* de l'API ISTEEX

Modules Laboratoires /institutions : utilisation de dictionnaires de données sur l'affiliation.

*NB: les cas complexes de multi affiliations ne sera pas traité, seul la première sera prise en compte hormis si elle est vide ou incomplète.*



# Traitement des données (affiliations)

- Identification et uniformisation des séparateurs
- Traitement des caractères et champs spéciaux
- Utilisation des dictionnaires de données pour déterminer les laboratoires présents ainsi que les institutions.
- Tous les termes sont ensuite comptés et regroupés selon leurs similarité.



# Affichage des résultats

- Les résultats sont affichés par modules.
- Forme de visualisation disponible :
  - Tableau dynamique paginé
  - Cartographie de type bubble chart
  - Cartographie sur une map monde
- Affichage des publications ne possédant pas d'affiliations, ainsi que celles ne pouvant être traités à cause de leurs qualité.







## Autres fonctionnalités

- Possibilité de consultation des ids des publications d'un laboratoire ou d'un auteur.
- Export au format CSV des tableaux
- Export au format image des cartographies
- Possibilité d'améliorer la détection des pays (improve) en élargissant aux états/régions/départements



# Démo

 **Carto Istex**

Help

Rechercher

Environ 27 652 résultats (1.35 secondes)

1 2 3 4 5 6 7 8 9 10 >




Qualité du pdf ▾



**Meteoritic diagenesis of Upper Cretaceous and Paleocene-Eocene shallow-water...**




Résumé : In the central part of the Kruja Platform (Albania) located in the Apulian passive margin, geochemical analyses (calciometry, Sr, REE and isotopic,  $\delta^{13}C$  and  $\delta^{18}O$ ...

Qualité du PDF : ★★★★★

Publié en 2009 | Degruyter-journals | Type : Research-article

Texte complet   

Métadonnées  

Enrichissements   

**Trace element fractionation during dehydration of eclogites from high-pressure...**

Résumé : The trace element compositions of eclogites, blueschists and mafic granulites from high-pressure terranes have been analysed to investigate element losses and f...





Qualité du PDF : ★★★★★

⚠ Warning: only 15000 results parsed!

View on map
Export CSV
🔍 Improve

### Countries

Search:

Country	Number of publications
 United States of America	1632
 Japan	1196
 Germany	797
 France	740

View on map
Export CSV

### Laboratories

Search:

Laboratory	Institution	Number of publications
RESEARCH SCHOOL OF EARTH SCIENCES	AUSTRALIAN NATIONAL UNIVERSITY	55
ARGONNE NATIONAL LABORATORY	ARGONNE NATIONAL LABORATORY	45

View on map
Export CSV

### Authors

Search:

Author	Laboratory	Institution	Number of publications
J DOSTAL	DEPARTMENT OF GEOLOGY	SAINT MARY'S UNIVERSITY	28
D W KURTZ	CENTRE FOR ASTROPHYSICS	UNIVERSITY OF CENTRAL LANCASHIRE	27

<https://carto-istex.otelo.univ-lorraine.fr>





# Conclusions - Perspectives

- ✓ Rédaction du rapport final du projet
- ✓ Valorisation des résultats (LabEX, d'OTELo, ...)
- ✓ Intégration de l'API V5 istex
- ✓ Intégration de la nouvelle version Widget SNU du 31/05/17
- ✓ Déploiement de carto-istex pour la communauté