



ISTEX
L'excellence documentaire pour tous

LorExplor: Semantic MediaWiki Infrastructure for exploring corpus with ISTEX

Journée doctorants 2017 Nancy





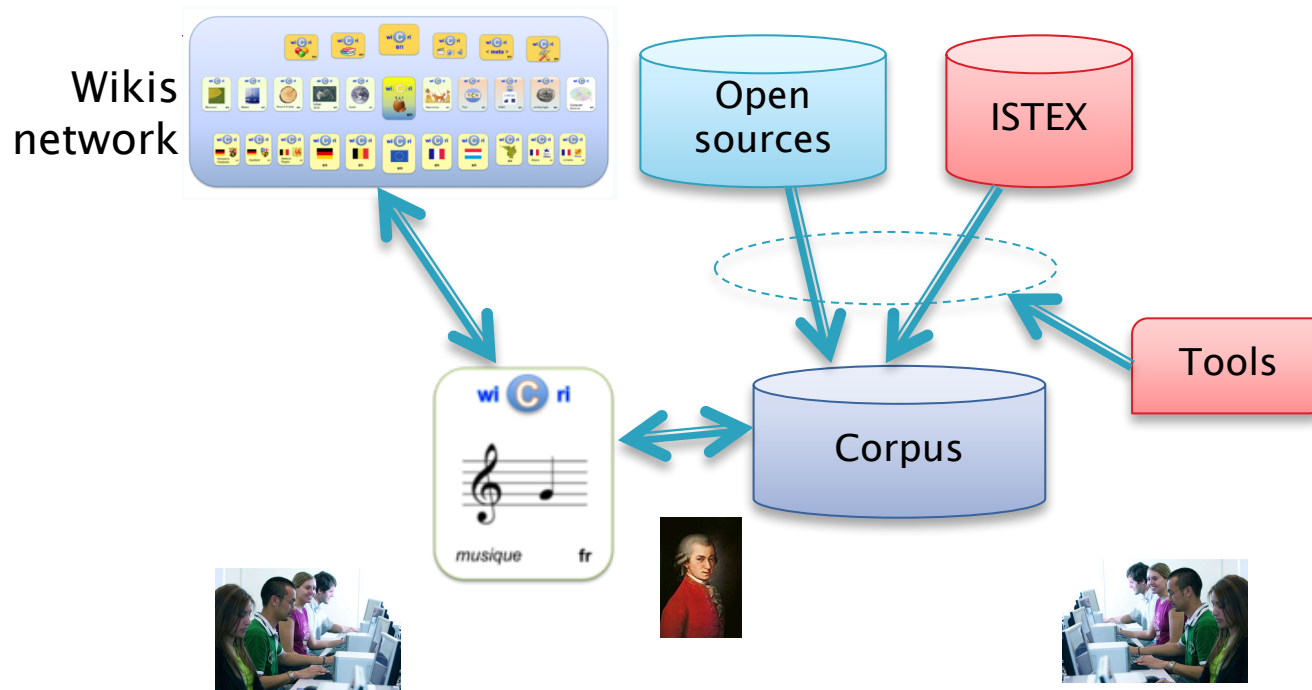
- ▶ French national initiative
 - Investments for the future
 - *Investissements d'avenir*
- ▶ Catch word
 - *Documentary excellence for everybody*
- ▶ Equipment
 - Budget : approximatively 60.000.000 €
- ▶ Archives and data
 - target: 50 000 000 documents
 - At this time: 20 000 000
 - Springer, Wiley, Elsevier, Oxford University Press...
 - Various formats: text (OCR), metadata, XML...
- ▶ A portal is available for French academic people

LorExplor focuses on:

- ▶ What can a young researcher do with 10,000 documents related to his thesis topic?

LorExplor: exploring new practices

- ▶ Text & data mining with time constraints
- ▶ Co-construction of scientific or cultural portal



Reading and writing science in an interdisciplinary world

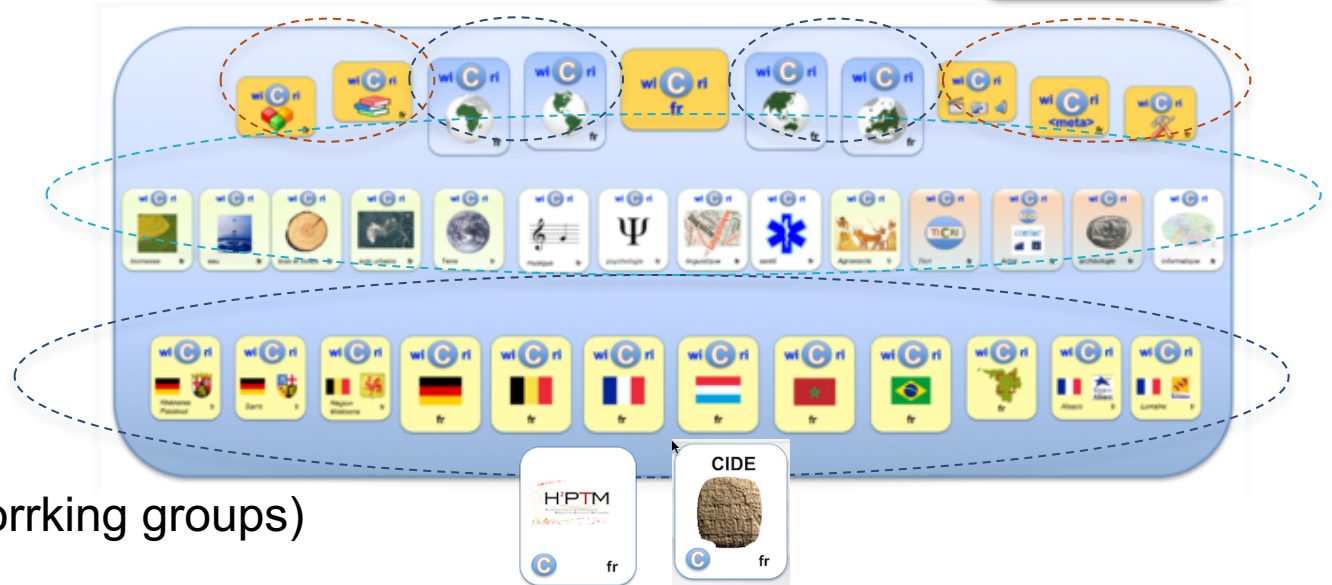


Technical wikis

Thematic wikis

Regional wikis

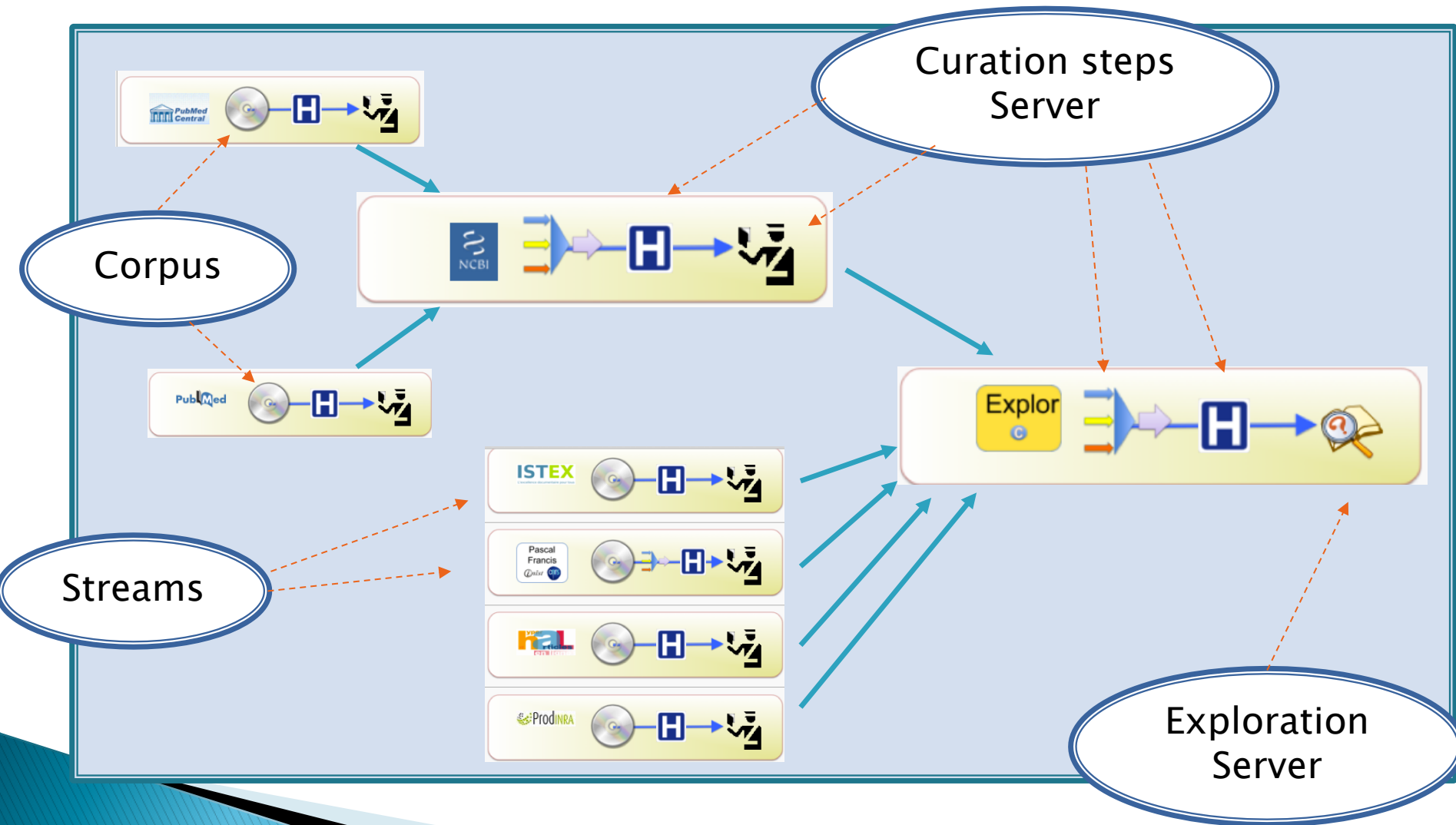
Associated wikis (working groups)



No anonymous !!!!

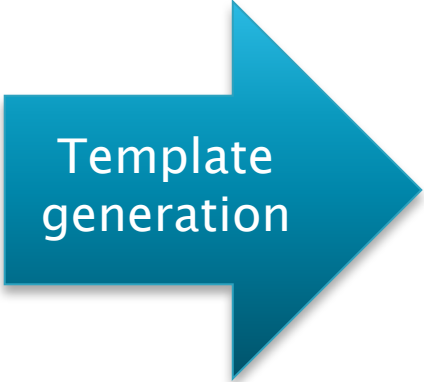
Any registered professional (researcher, practitioner, engineer ...)
can contribute to any wiki

Exploration/curation area



Exploration servers

Index Browsing with wiki summary



Template
generation

Pays

1. France (67) [🔗](#)
2. États-Unis (31) [🔗](#)
3. Royaume-Uni (14) [🔗](#)
4. Allemagne (14) [🔗](#)
5. Canada (11) [🔗](#)
6. Italie (10) [🔗](#)
7. Espagne (8) [🔗](#)
8. Suisse (6) [🔗](#)
9. Australie (6) [🔗](#)
10. Pays-Bas (5) [🔗](#)

Région

1. Californie (11) [🔗](#)
2. Île-de-France (9) [🔗](#)
3. Occitanie (région administrative) (7) [🔗](#)
4. Massachusetts (6) [🔗](#)
5. Angleterre (6) [🔗](#)
6. État de New York (5) [🔗](#)
7. Maryland (5) [🔗](#)
8. Caroline du Nord (5) [🔗](#)
9. Arizona (5) [🔗](#)
10. Washington (État) (4) [🔗](#)

Villes

1. Paris (9) [🔗](#)
2. Marseille (5) [🔗](#)
3. Montpellier (4) [🔗](#)
4. Londres (4) [🔗](#)
5. Grenoble (4) [🔗](#)
6. Berlin (4) [🔗](#)
7. Toulouse (3) [🔗](#)
8. Prague (3) [🔗](#)
9. Montréal (3) [🔗](#)
10. Zurich (2) [🔗](#)

Mots-clés anglais

- :
1. Astrophysics (3) [🔗](#)
 2. State of the art (2) [🔗](#)
 3. Software package (2) [🔗](#)
 4. Real time (2) [🔗](#)
 5. Quebec (2) [🔗](#)
 6. Perspective (2) [🔗](#)
 7. Open source software (2) [🔗](#)
 8. Measurement sensor (2) [🔗](#)
 9. Library network (2) [🔗](#)
 10. Information policy (2) [🔗](#)

Mots des titres

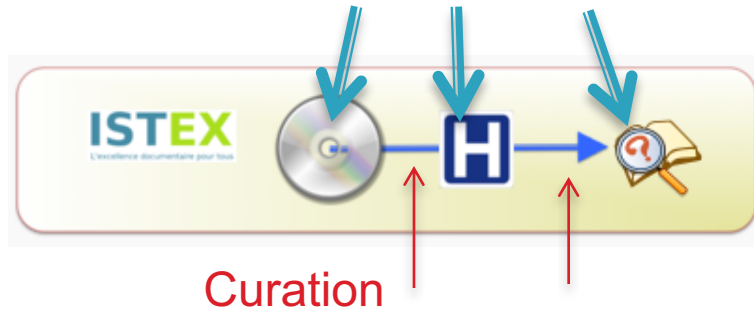
1. data (10) [🔗](#)
2. analysis (7) [🔗](#)
3. software (6) [🔗](#)
4. microbial (6) [🔗](#)
5. marine (5) [🔗](#)
6. genome (5) [🔗](#)
7. distributed (5) [🔗](#)
8. genomic (4) [🔗](#)
9. control (4) [🔗](#)
10. web (3) [🔗](#)

ISSN/revue

1. SPIE proceedings series (6) [🔗](#)
2. 1932-6203 (5) [🔗](#)
3. Lecture Notes in Computer Science (4) [🔗](#)
4. Eos Trans. AGU (3) [🔗](#)
5. 2324-9250 (3) [🔗](#)
6. 1091-6490 (3) [🔗](#)
7. 0096-3941 (3) [🔗](#)
8. 0027-8424 (3) [🔗](#)
9. 2047-217X (2) [🔗](#)
10. 1545-7885 (2) [🔗](#)

Exploration area and servers

Several servers by stream (curation steps)



http://ticri.univ-lorraine.fr/Wicri/Psycho/corpus/GrossesseFrancis/GrossesseFrancisV1/Site/fr/Main/Exp

Wikipédia (fr)

Francis

Serveur d'exploration sur la grossesse dans Francis

Attention, ce site est en cours de développement !
Attention, site généré par des moyens informatiques à partir de corpus bruts.
Les informations ne sont donc pas validées.

Eléments de l'association

Élément	Valeur
Akinrinola Bankole	5
Susheela Singh	8
Akinrinola Bankole	2
Sauf Susheela Singh	5
Akinrinola Bankole	5
Akinrinola Bankole Et	3
Susheela Singh	10
Akinrinola Bankole Ou	1292
Corpus	

List of bibliographic references

Number of relevant bibliographic references: 3.

Ident.	Authors (with country if any)	Title
Gilda Sedeh (Etats-Unis)	Akinrinola Bankole (Etats-Unis)	

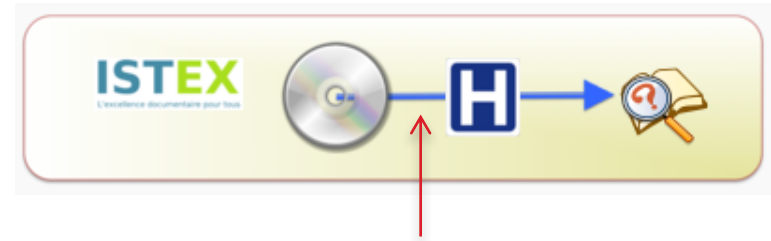


	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	H																	He
2	Li	Be											B	C	N	O	F	Ne
3	Na	Mg											Al	Si	P	S	Cl	Ar
4	K	Ca											Ga	Ge	As	Se	Br	Kr
5	Rb	Sr											In	Sn	Sb	Te	I	Xe
6	Cs	Ba	La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb		
7	Fr	Ra	Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No		

Tableau périodique des éléments chimiques

Data curation

- First samples: country names in an heterogeneous context



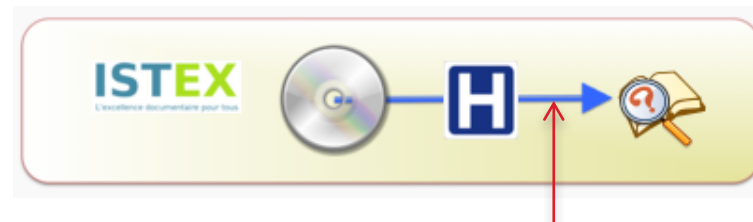
Serveur d'exploration sur la didactique – Wicri Wicri

ticri.univ-lorraine.fr/wicri.fr/index.php/Serveur_d'explor

Les plus visités | Catégorie: Palais ... | PLOS ONE: "Fres..." | Dilib, module Ex... | Getting Started

1	Pascal Francis	  	Le premier corpus est constitué de 4284 notices extraites de Pascal/Francis avec la requête « n didactique ». Une requête plus large (sans pré zone) donne environ 8000 notices.
2	PubMed	  	Le corpus PubMed est extrait avec le critère « qui sélectionne 4013 notices.
3	PubMed Central	  	Le corpus PMC est extrait avec le critère « did qui sélectionne 402 notices (la forme « didactio sélectionne environ 8000
4	Convergence NCB	  	Ce flux rassemble les 4415 notices venant de l et PubMed Central
5	HAL SHS	  	
Flux principal		  	Ce flux rassemble la totalité des 8643 notices.
Zoom	Auteurs français	  	Ce zoom propose une analyse plus fine autour travaux réalisés avec affiliations françaises (12 notices)
Zoom	Enseignement des langues	  	Ce zoom propose une analyse plus fine autour l'enseignement des langues (1127 notices)

Data curation: region level



These maps could be available on each servers

Exemples de sujets explorés

- ▶ Près de 200 expérimentations
 - dont Master Science information UL
 - Visibilité de la ville du Havre, (5303 au total/ dont 3003 ISTEEX)
 - Le cobalt au Maghreb, (4062 / 3027)
 - Un poisson : le scalaire, (1329 / 906)
 - Un arbre fruitier : l'oranger (8819 / 2923)
 - Le libre accès en Belgique, (3696 / 2961)
 - Master Paris 8
 - La Maladie de Parkinson en France, (11473 / 3727)
 - La Paléopathologie (5459 / 2469)
 - Université de Trèves (6789 / 2846)
 - La thérapie familiale en francophonie (3463 / 2817)
 - Système d'information stratégique et agriculture (3011 / 2042)
 - Informatique et sciences de l'information
 - OCR (8000), TEI (700), ...
 - Volumétrie accrue sur machine virtuelle ISTEEX/INIST
 - Le SIDA en Afrique Subsaharienne (30.000)
 - L'activité scientifique de Pittsburg (26.000)

Offres de services aux doctorants

- ▶ Analyse d'une thématique scientifique
 - Serveur associé à une thématique scientifique
 - avec un Zoom sur les sujets de thèse
 - avec un dossier wiki (sémantique)
 - Performances accrues si applications en santé, médecine ou sciences de la vie...
- ▶ R&D autour des « big data » autour de l'information scientifique
 - Élaboration, curation de corpus
 - Démonstrations d'outils de TAL ou visualisation
 - Wikis sémantiques (en réseau)