



LorExplor

ISTEX
L'excellence documentaire pour tous



Composants XML pour analyser les usages ISTEX

Jacques Ducloy, Thierry Daunois, Ali Tebbakh,
Jean-Pierre Thomesse, Jean-Charles Lamirel, Frédérique Péguiron
Université de Lorraine, Loria, Paragraphe (Paris 8)

Avant ISTEEX : Dilib, Wicri

2012 en Lorraine

- ▶ Réseau de compétences sur les besoins en IST : INIST, LORIA, UL, DRRT Lorraine
- ▶ Ancrage dans les communautés : CIDE, VSST, H2PTM, DCMI...
- ▶ Racines sur la promotion de l'informatique pour les chercheurs IUCAL (J. Legras), ANL (C. Pair)
 - Le CNRS avait un département Sciences pour l'Ingénieur...
 - Il existait une Agence pour l'Informatique avec un service dédié à la Recherche (SM90)

5 juin 1991 : Florence (Italie), Séminaire IFLA/Unesco/CCF

- ▶ N. Dusoulier : ingénierie SGML pour ISO 2709 Unimarc, CCF, USMARC



- ▶ 1992 INIST/DRPN (créé par G. Delacote) : Ilib, une première boîte de composants SGML
 - Maquette d'un parser SGML dans une approche XML
 - Boîte à outils SGML pour recherche d'information
- ▶ 1994 au LORIA/INRIA (ERCIM) :
 - Dilib, un parser XML et une boîte à outils
 - Serveur d'exploration pour analyser des corpus de métadonnées hétérogènes (MedExplore)



2008 : DRRT Lorraine + Région Lorraine + Feder, INPL, Nancy–Université

- ▶ Wicri :réseau de wikis sémantiques
 - dans des disciplines scientifiques variées (Matériaux, Sciences et Génie de l'Environnement, SHS)
 - pour des pratiques diversifiés (édition, terminologie, observatoire...
- ▶ Premières réflexions sur la synergie wiki – serveur dans un cadre pédagogique

2012 LorExplor



ISTEX
L'excellence documentaire pour tous

- ▶ **Proposition Lorraine vers ISTEX : analyser les usages ouverts par ISTEX pour les communautés scientifiques,**
 - exploration de corpus,
 - par des chercheurs
 - à des fins scientifiques (vers les applications)
 - Utilisation de Wicri comme
 - outil de prise de contact,
 - infrastructure de travail.

- ▶ **Soutien ISTEX accordé pour adapter la boîte à outils XML**
 - Un ingénieur IE à mi-temps + 2 jours d'expert par mois + stagiaires...
 - Analyse des usages réalisée sur la base du volontariat
 - Très peu de soutien logistique institutionnel
 - Parfois de l'hostilité institutionnelle (résistance au changement)
 - Wicri : LORIA - 2010 -> INPL - 2018-> INIST (grâce à ISTEX)
 - Très nombreux soutiens individuels
 - Des soutiens institutionnels extérieurs à la Lorraine (Paris 8, INRA, IHEST...)

Bilan LorExplor 2018 : 200 explorations de corpus

▶ Dont : 50 dans des séances de travaux dirigés

- à partir de sujets réels, choisis par les étudiants
- par groupe de 2 à 5
- 10 séances de 3 heures (5 à 10 heures de travail personnel)



Avec une restitution notée !

▶ 2017 – 2018 : **Master Documentation Numérique** (Univ. Lorraine)

- Relations France – Australie (25.000 doc., dont 5.000 ISTEEX et 11.000 pascal)
- Pittsburgh (26.000 doc., dont 6.000 ISTEEX et 7.000 PubMed)
- Le Patient édenté complet (30.000 doc., dont 10.000 ISTEEX)
- 2016 – 2017: **Masters Paris 8, 10 sujets dont :**
 - La Maladie de Parkinson en France, (11.473 Total/ 3.727 ISTEEX)
 - La paléo pathologie (5459 / 2469)
 - Le nickel au Maghreb (3337/ 2500)
 - Université de Trèves (6789 / 2846)
 - Un poisson : l'esturgeon (4057/2398)
 - Thérapie familiale en francophonie (3463 /2817)
 - ...
- 2012 – première séance (Master Intelligence économique Nancy 2)
 - Recherche forestière en France : 1300 métadonnées Pascal...

Première conclusion et plan

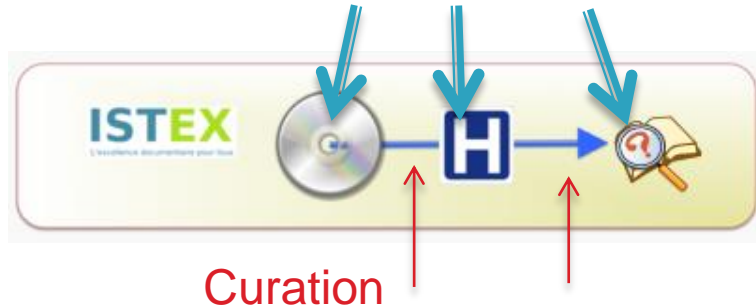
- ▶ ISTEEX démocratise l'exploration de Corpus
 - Exemple le patient édenté...
 - 30.000 documents dont 10.000 ISTEEX (30 € par document)
 - Coût avant ISTEEX pour un binôme : 300.000 €

 - ▶ Comment déployer ISTEEX ?
 - Stratégie pour « Couperin » de type « *learning centers* » ?
 - 500.000 sorties LMD par an !
 - Stratégie pour le « CNRS » : 1 équipe formée par laboratoire ?
1. Présentation du démonstrateur LorExplor
 2. Mission financée par ISTEEX : Une boîte à outil XML
 3. Retours d'expériences
 4. Conclusion prospective : infrastructures, formation

Serveur d'exploration

génééré par DILIB ISTEEX -> Serveur -> Wiki

Système d'information orienté exploration



Pays	Région	Villes
1. France (67) ↗	1. Californie (11) ↗	1. Paris (9) ↗
2. États-Unis (31) ↗	2. Île-de-France (9) ↗	2. Marseille (5) ↗
3. Royaume-Uni (14) ↗	3. Occitanie (région administrative) (7) ↗	3. Montpellier (4) ↗
4. Allemagne (14) ↗	4. Massachusetts (6) ↗	4. Londres (4) ↗
5. Canada (11) ↗	5. Angleterre (6) ↗	5. Grenoble (4) ↗
6. Italie (10) ↗	6. État de New York (5) ↗	6. Berlin (4) ↗
7. Espagne (8) ↗	7. Maryland (5) ↗	7. Toulouse (3) ↗
8. Suisse (6) ↗	8. Caroline du Nord (5) ↗	8. Prague (3) ↗
9. Australie (6) ↗	9. Arizona (5) ↗	9. Montréal (2) ↗
10. Pays-Bas (5) ↗	10. Washington (État) (4) ↗	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) ↗

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.

Éléments de l'association

Akintola Bankole	5
Sushiea Singh	8
Sushiea Singh	2
Sushiea Singh	5
Akintola Bankole	5
Akintola Bankole	3
Sushiea Singh	3
Akintola Bankole	10
Sushiea Singh	10

List of bibliographic references

Number of relevant bibliographic references: 3.
Authors (with country if any)
Ident. | Title



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										Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr		
5	Rb	Sr										Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe		
6	Cs	Ba										Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
7	Fr	Ra										Fr	Ra	Lr	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Cn	Uut	Uuq	Uup	Uuh	Uus	Uuo

↓

Tableau périodique des éléments chimiques

Dilib, une boîte à outils Sxml

- ▶ **SXML : XML lite (mais JSON+)**
 - Compatible avec les outils Unix
 - Un document = Une ligne Unix
- ▶ **Origine**
 - 1990 : Ilib : ISO 2709 (MARC, Pascal...)
 - Un LEGO pour les corpus
 - 2000 : Dilib : métadonnées hétérogènes
- ▶ **2018 : LorExplor**
 - traiter du corpus volumineux,
 - Textuel, multi-dtd
 - Réseau MediaWiki
 - Générations de modèles wiki
 - Robots



```
<index>
  <kw>Requiem</kw>
  <list>
    <item>004321</item>
    <item>012345</item>
  </list>
  <f>2</f>
</index>
```

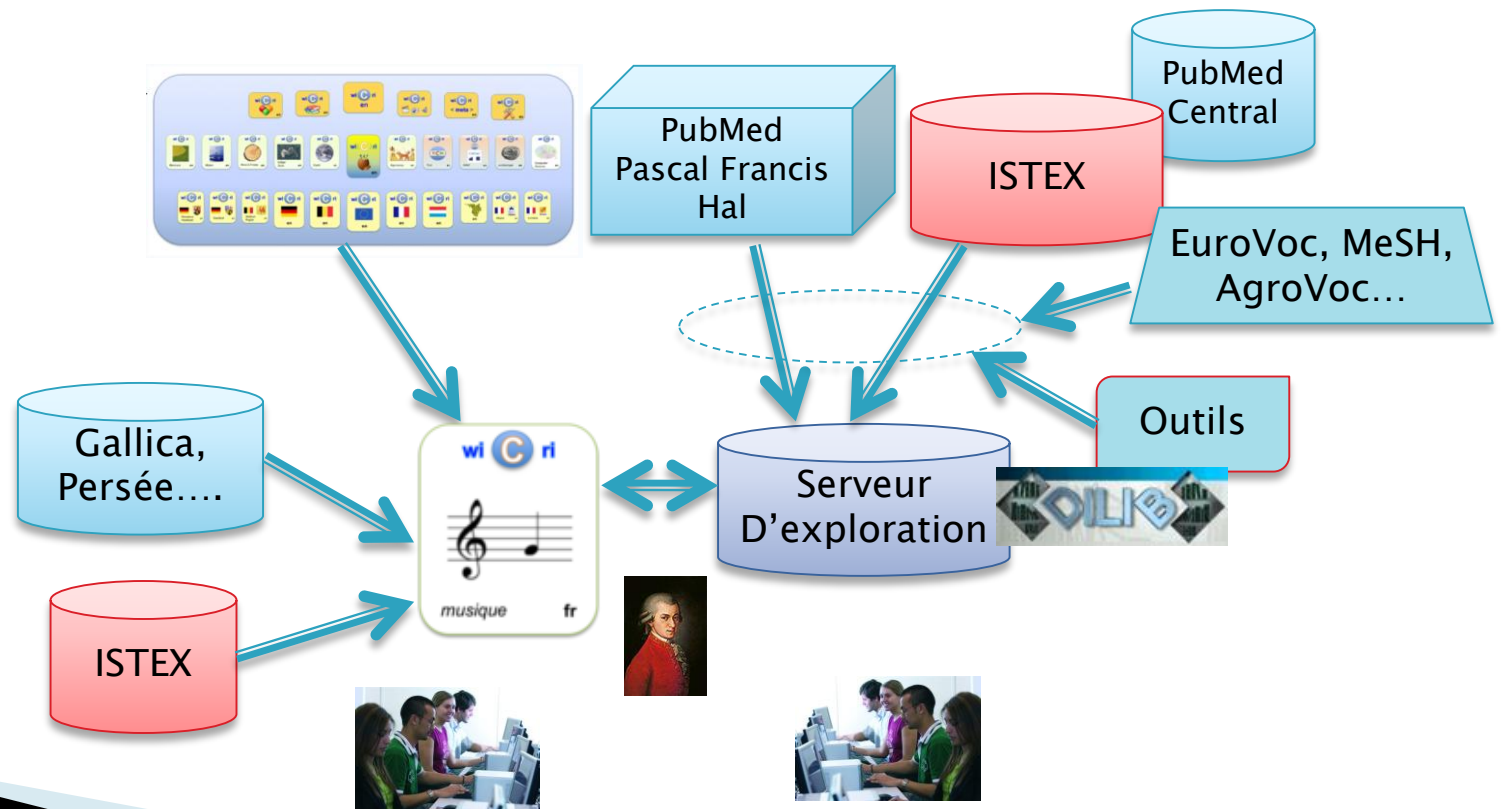

Retours d'expériences

Analyse des acteurs internationaux

- ▶ Bibliographie d'un thésard
 - Exemple : *dispositifs haptiques* : 8000 documents ISTEEX
- ▶ Pilotage / valorisation / coopérations internationales
 - Exemple : *repérer la position de la Lorraine parmi les coopérations entre la France et l'Australie* :
 - Pour une vue d'ensemble : 5.000 documents ISTEEX
 - Pour détecter des signaux faibles : 300.000 ISTEEX
- ▶ Problèmes :
 - couverture trop faible sur l'actualité
 - Élargir avec d'autres réservoirs (Hal, PubMed Central...)
 - Indexation peu précise
 - *Idem* (Pascal, Francis PubMed)
 - Contrainte de temps (réponse à appel d'offre)
 - Infrastructure informationnelle prête à l'emploi (Wicri, Serveurs)
 - Travail de type accompagnement
 - Implique un espace collaboratif (wiki sur machine virtuelle)

Infrastructure collaborative, multidisciplinaire et interoperable

- ▶ Recherches exploratoires avec contraintes de temps
- ▶ Co-construction de connaissance scientifique ou culturelle
 - Rééditions, données de la recherche, articles originaux, encyclopédiques,



Curation des données

Les pays dans un contexte hétérogène



Forme française sur Wicri	Forme anglaise sur Wicri	Forme courantes
Afrique du Sud	South Africa	South Africa ; Republic of South Africa
Arabie saoudite	Saudi Arabia	Saudi Arabia
Allemagne	Germany	Germany ; Deutschland ; Federal Republic of Germany ; Bundesrepublik Deutschland ; FRG ; DDR ; West Germany ; W. Germany ; Fed. Rep. Germany ; GDR ; German Democratic Republic ; Deutsche Demokratische Republik
Argentine	Argentina	Argentina
Australie	Australia	
Autriche	Austria	

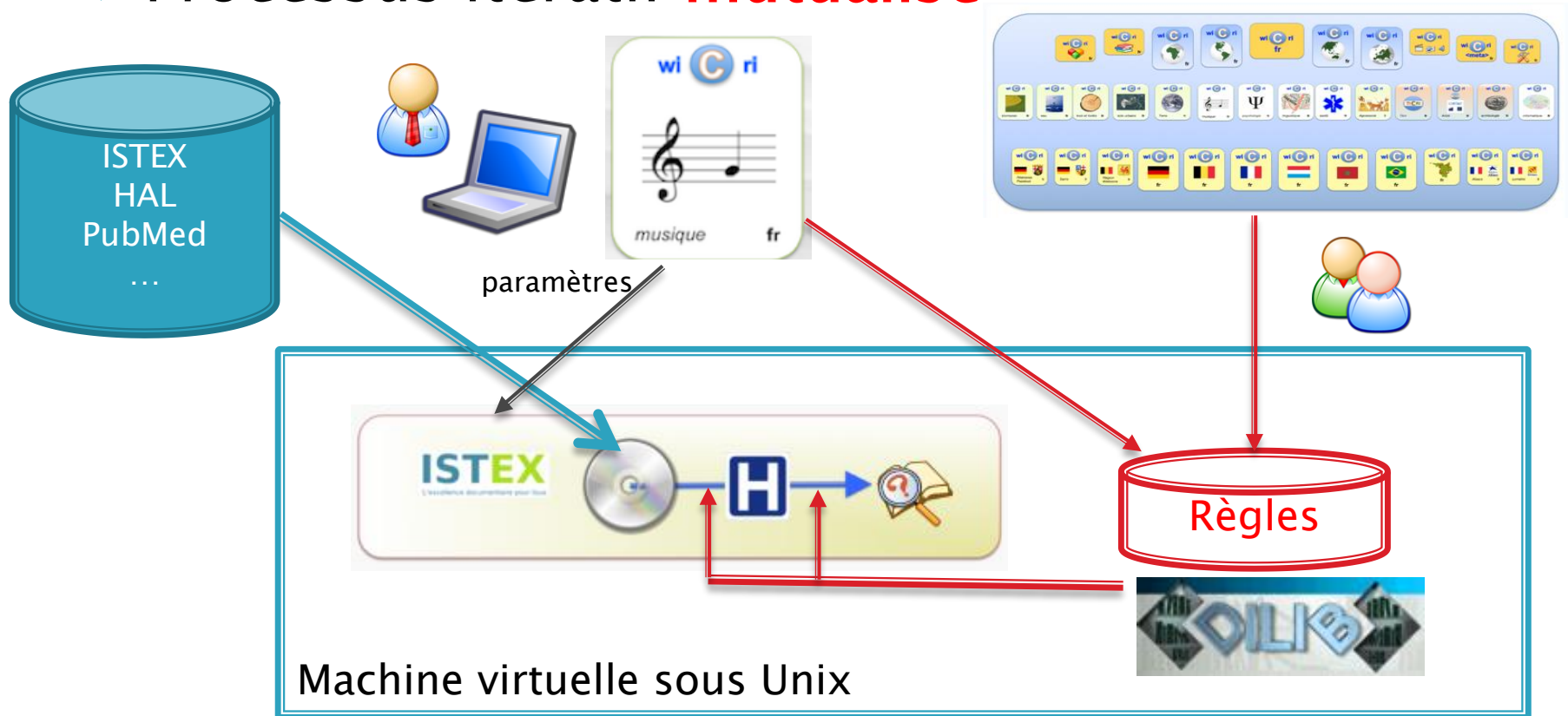
Retours sur 200 expériences

Corpus : méfiance / curation

- ▶ Exemple : *la méthode Scrum*
 - Apparemment : 9.000 documents
 - En fait 90% de bruit du à l'OCR (sérum -> scrum)
- ▶ Exemple : *le libre accès en Belgique*
 - Apparemment : 4000 documents
 - En fait : 100 à 200 sont pertinents
 - *Title: The EADGENE Microarray Data Analysis Workshop (Open Access publication)*
- ▶ Exemple : *Coopérations France Australie* (échantillon)
 - 5000 documents ISTEEX / 6000 PubMed
 - 38 ont plus de 150 auteurs / 128 dans PubMed
 - 6 en ont plus de 1000 / 115 dans PubMed ont plus de 2000
 - 3 en ont plus de 3000 / 5 dans PubMed
 - Record : 3199 / 3048

Génération de plateformes de curation et d'exploration

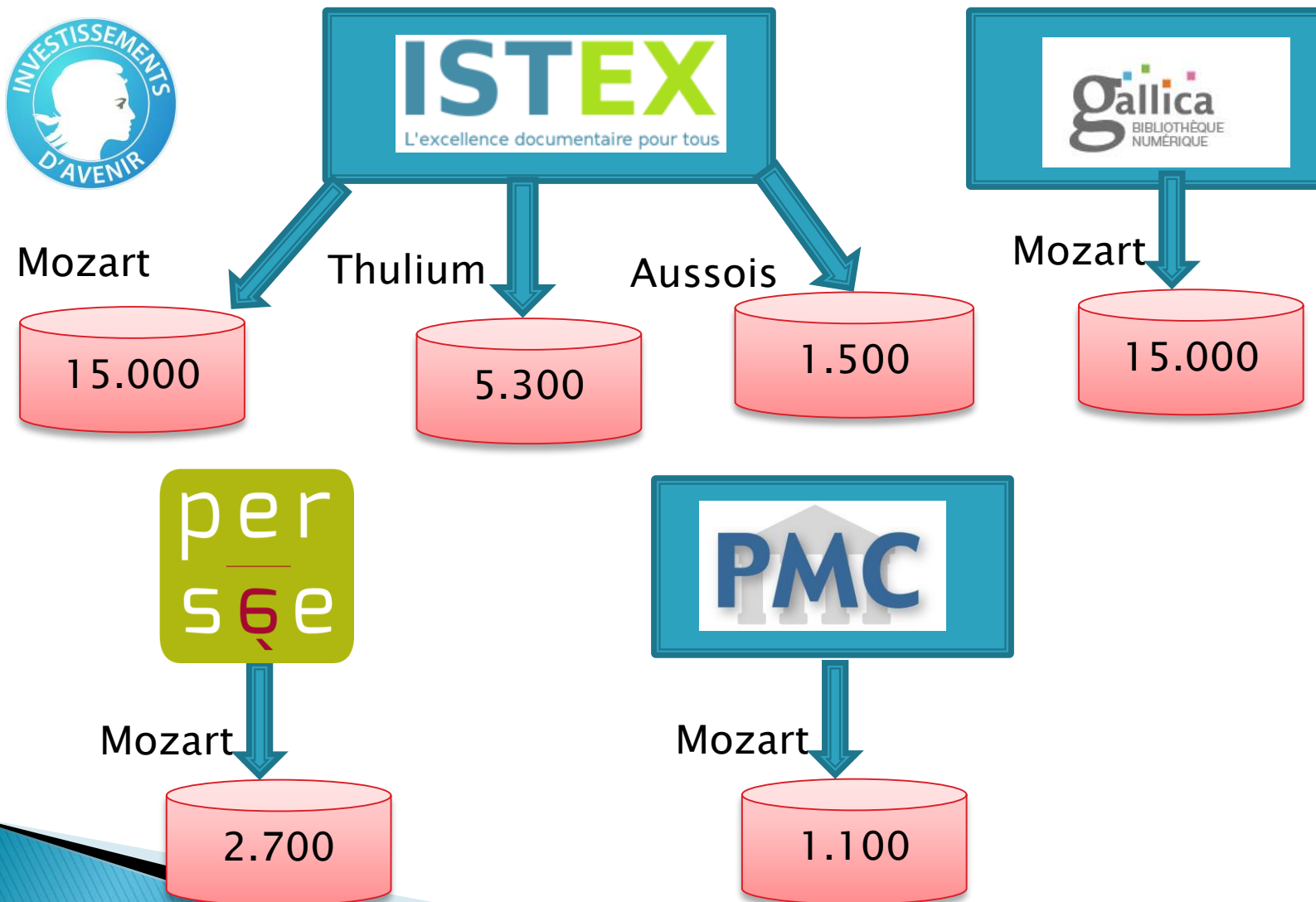
- ▶ Processus itératif **mutualisé**



Retours d'expériences sciences humaines et expérimentales

- ▶ Enjeux d'ISTEX en médecine et santé
 - Demandes réelles : *Parkinson, Lymphœdème, SIDA*
 - Détecter des connaissances inconnues,
 - Exploration des contenus
 - (et pas seulement les métadonnées)
- ▶ Enjeux dans les sciences expérimentales
 - *trouver de nouvelles espèces de poissons domesticables*
 - 150.000 documents (au moins)
 - Noms binomiaux, formules chimiques, dimensions
- ▶ Enjeux dans les humanités
 - Exemple : *philologies, dispositifs haptiques, chanson de Roland, chartes lorraines, Renaissance en Lorraine*
 - Combiner bibliographie (ISTEX) et sources (Gallica..)

Humanités : Gallica, Persée...



Corpus : méfiance / curation

▶ Exemple : Mozart

- 15.000 documents (Musique + médecine)
- Quelques problèmes de type « avenue Mozart »
- Plus sérieux :
 - Musique : peu de signalement d'affiliations
 - Médecine : forte politique d'affiliations
- Les statistiques se focalisent sur la médecine...

▶ Exemple : Parkinson en France

- Parkinson : 90.000 documents
- Extrait de 4000 documents :
 - peu de bruit
- Parkinson en France :
 - beaucoup de bruit.



▶ Quelle formation donner à un bibliothécaire pour accompagner un chercheur dans une démarche de curation?

Enrichissements

- ▶ Essais avec Teeft (ISTEX-R)
 - Difficultés techniques (explosion d'algorithmes)
 - Granularité (document) trop large
 - Souhaits :
 - Granularité paragraphe ou annotations dans le texte
 - Utilisation en aval
 - avec la curation
 - **Dans la diversité des thématiques et des besoins des utilisateurs**



Filtrage du texte « pour tous » Quels outils ? Heuristiques => Quelle formation ?

- ▶ Quelles sont les œuvres de Mozart les plus citées dans un corpus ?
 - Idée générale : utiliser le catalogue Köchel
 - Résultat : Sonate KV. 448

```
HfdCat Data/Main/Exploration/biblio.hfd \
| SxmlFindText -r "[K][Vv]*[ \.]*[0-9][0-9]* » \
| SxmlSelect -p @5 -p @1 | sort | IndexBuildRec
```

- ▶ Quelles sont les applications de « *dance therapy* » avec une dimension artistique ?
 - Recherche de présence de chorégraphes (nom-prénom) en utilisant un filtre créé pour les noms binomiaux
- ▶ Quelles sont les publications traitant des ressources liées au cobalt dans le Maghreb ?
 - Pondérations pour définir le cœur d'un corpus

Conclusion : déploiement ISTEEX quels outils ? quelle formation ?

- ▶ Déploiement type 3 phases :
 1. Formateurs et acteurs individuels motivés,
 2. Ensemble des acteurs dans des unités motivées,
 3. Déploiement général
- ▶ Réseau de wikis en cohérence sémantique
 1. Actuellement : 1 site / 38 familles communes, / 130 wikis / 200 serveurs
 2. Pas de verrous technologique pour passer à : 3 / 100 / 300 / 1000
 3. Un projet intéressant :
 - 10.000 wikis cohérents sur 100 sites !
- ▶ Boîte à outils XML (et autres outil type TAL)
 1. Utilisable avec génération de serveur « en prestation »
 2. Premier niveau d'autonomie avec paramétrage prédéfini
 3. Un projet intéressant :
 - ensemble boîte à outil + outils d'enrichissement utilisables par un chercheur de base !
- ▶ Formation :
 - expérience ateliers learning center avec la BU Lettre de l'Université de Lorraine

ISTEX vers la société

- ▶ Publier pour la société et le citoyen
 - Puissant moteur pour aller vers l'excellence
 - Rédiger des connaissances implique un traitement spécifique sur le contenu de corpus ISTEX
- ▶ Du monolithe au réseau
 - Monolithe
 - Chercheur = producteur / consommateur de PDF
 - Peu implication technique en dehors de l'équipe centrale et d'une poignée d'acteurs
 - Des grands réservoirs + un réseau
 - Les chercheurs deviennent acteurs du numérique,
 - Les communautés sont totalement impliqués,
 - On peut impliquer des acteurs de la société...

Merci

- ▶ A ISTEEX pour cette aventure technologique, scientifique et humaine,
- ▶ Aux collègues et ami(e)s qui m'accompagnent depuis 50 ans sur l'appropriation de l'informatique, puis du numérique, par les communautés de la recherche...
- ▶ Pour votre attention,
- ▶ Pour pour vos questions...
- ▶ Annexes : compléments wiki sémantiques, curation, détail déploiement

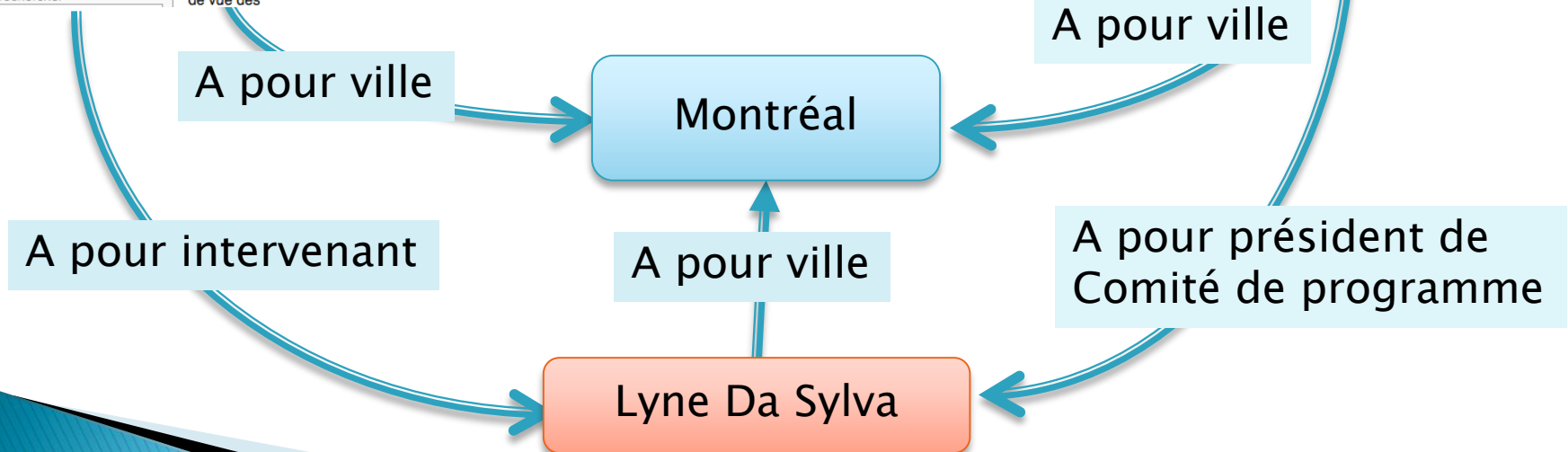
Les wikis sémantiques pour les systèmes de recherche

The screenshot shows a Wikidata page for 'CIDE 2009 Montréal'. The page title is 'CIDE 2009 Montréal'. Below the title, there is a navigation menu with options like 'page', 'discussion', 'modifier', 'historique', and 'supprimer'. The main content area is titled 'Présentation générale' and includes a description: 'Depuis 1998, CIDE propose un cycle de manifestations scientifiques sur le thème du document électronique, avec pour objectif de confronter les points de vue des'. To the right, there is a table with the following data:

CIDE 2009 Montréal	
Cycle :	CIDE
Début :	Oct 21 2009
Fin :	Oct 23 2009
Site :	europa.org/CIDE12
Ville :	Montréal
Pays :	Canada

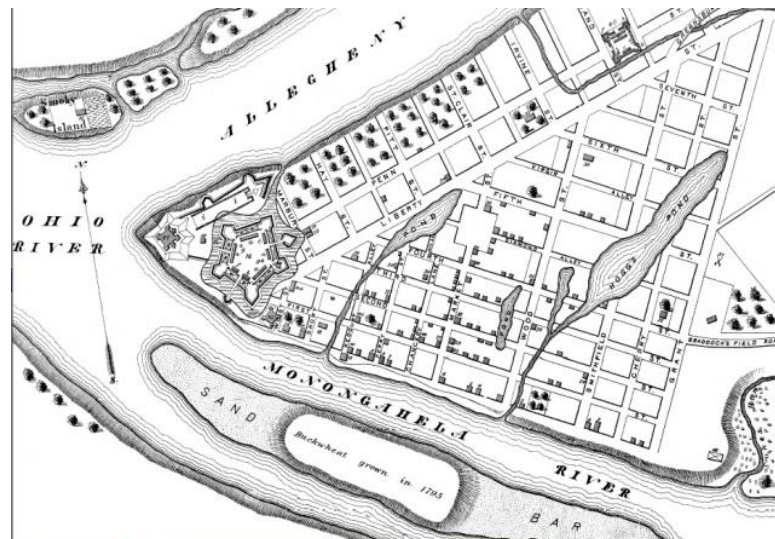
The screenshot shows a Wikidata page for 'Acfas 2017 Montréal - Analyser la science'. The page title is 'Acfas 2017 Montréal - Analyser la science'. Below the title, there is a navigation menu with options like 'page', 'discussion', 'modifier', 'historique', 'supprimer', 'renommer', and 'pr'. The main content area is titled 'Sommaire [masquer]' and includes a list of sections: '1 Introduction', '2 Organisation', and '3 Programme'. To the right, there is a table with the following data:

Acfas 2017 Montréal - Analyser la science	
Début :	8 mai 2017
Fin :	9 mai 2017
Ville :	Montréal
Pays :	Canada



Liens sémantiques

Pittsburgh est située au confluent des rivières Allegheny et Monongahela qui forment l'Ohio, un affluent du Mississippi .



Pittsburgh est située au confluent des rivières
[[sur le cours d'eau::Allegheny (rivière)|Allegheny]]
et [[sur le cours d'eau::Monongahela]] qui forment
l'[[sur le cours d'eau::Ohio (rivière)|Ohio]], un affluent du
[[Mississippi (fleuve)|Mississippi]] .

Faits relatifs à Pittsburgh ⓘ — Recherche de pages similaires avec + 🔍 .

Voir comme RDF 🔗

Sur le cours d'eau Allegheny (rivière) + 🔍 , Monongahela + 🔍 et Ohio (rivière) + 🔍

L'Ohio

==Les affluents de l'Ohio==

("liste calculée")

```
{{#ask:[[est un affluent::{{PAGENAME}}]]  
| format=ul  
| sep=,_  
| intro=Rivières citées sur Wicri Eau :_ }}
```

==Les villes traversées par l'Ohio==

("liste calculée")

```
{{#ask:[[sur le cours  
d'eau::{{PAGENAME}}]]  
| format=ul  
| sep=,_  
| intro=Villes citées sur Wicri Eau :_ }}
```



navigation

- Accueil
- Communauté
- Actualités
- Modifications récentes
- Index alphabétique
- Index thématique
- Une page au hasard
- Aide

rechercher

boîte à outils

- Pages liées
- Suivi des liens
- Importer un fichier
- Pages spéciales
- Version imprimable
- Lien historique
- Chercher les propriétés

Ohio (rivière)

 Pour les articles homonymes, voir [Ohio \(homonymie\)](#).

L'**Ohio** est l'un des principaux affluents du [Mississippi](#) ; il coule dans la partie Est des [États-Unis](#).

Les affluents de l'Ohio

[modifier]

(liste calculée)

Rivières citées sur Wicri Eau :

- [Allegheny \(rivière\)](#)
- [Monongahela](#)

Les villes traversées par l'Ohio

[modifier]

(liste calculée)

Villes citées sur Wicri Eau :

- [Pittsburgh](#)

Liens interwikis

[modifier]

La rivière Ohio sur [Wikipédia](#).

Catégories : [Cours d'eau de l'Ohio](#) | [Cours d'eau de l'Illinois](#) | [Cours d'eau de l'Indiana](#) | [Cours d'eau du Kentucky](#) | [Cours d'eau de Pennsylvanie](#)

Faits relatifs à [Ohio \(rivière\)](#) ⓘ — Recherche de pages similaires avec + 🔍 .

[Voir comme RDF](#) 

Est un affluent [Mississippi \(fleuve\)](#) + 🔍



Le bassin de l'Ohio



Curation des données – pays

► Codes ISO (exemple Pascal)

numé- rique	alpha -3	alpha -2	Nom français usuel	Nom ISO du pays ou territoire
004	AFG	AF	Afghanistan	AFGHANISTAN
710	ZAF	ZA	Afrique du Sud	AFRIQUE DU SUD
248	ALA	AX	Åland	Modèle:Tri1ÅLAND, ÎLES
008	ALB	AL	Albanie	ALBANIE
012	DZA	DZ	Algérie	Modèle:Tri1ALGÉ
276	DEU	DE	Allemagne	ALLEMAGNE
020	AND	AD	Andorre	ANDORRE
024	AGO	AO	Angola	ANGOLA
660	AIA	AI	Anguilla	ANGUILLA

pA A01 01 1 @0 0302-9743
 A05 @2 1375
 A08 01 1 ENG @1 Hyperbook data modeling
 A09 01 1 ENG @1 Electronic publishing, artistic imaging, and digital typography : Saint Malo, 30 March - 3 April 1998
 A11 01 1 @1 FRÖHLICH (P.)
 A11 02 1 @1 HENZE (N.)
 A11 03 1 @1 NEJDL (W.)
 A12 01 1 @1 HERSCH (Roger D.) @9 ed.
 A12 02 1 @1 ANDRE (Jacques) @9 ed.
 A12 03 1 @1 BROWN (Heather) @9 ed.
 A14 01 @1 Institut für Rechnergestützte Wissensverarbeitung, Universität Hannover, Lange Laube 3 @2 30159 Hannover @3 DEU @Z 1 aut. @Z 2 aut. @Z 3 aut.

Curation des pays – Adresses

Adresses postales
(Springer, PubMed)

Forme française sur Wicri	Forme anglaise sur Wicri	Forme courantes
Afrique du Sud	South Africa	South Africa ; Republic of South Africa
Arabie saoudite	Saudi Arabia	Saudi Arabia
Allemagne	Germany	Germany ; Deutschland ; Federal Republic of Germany ; Bundesrepublik Deutschland ; FRG ; DDR ; West Germany ; W. Germany ; Fed. Rep. Germany ; GDR ; German Democratic Republic ; Deutsche Demokratische Republik
Argentine	Argentina	Argentina
Australie	Australia	Australia

```
<titleInfo lang="eng">
  <title>Graph Access Pattern Diagrams (GAP-D): Towards a
  Unified Approach for Modeling Navigation over
  Hierarchical, Linear and Networked Structures</title>
</titleInfo>
<name type="personal">
  <namePart type="given">Matthias</namePart>
  <namePart type="family">Keller</namePart>
  <role>
    <roleTerm type="text">author</roleTerm>
  </role>
  <description>Matthias.keller@kit.edu</description>
  <affiliation>Steinbuch Centre for Computing (SCC),
  Karlsruhe Institute of Technology (KIT), D-76128,
  Karlsruhe, Germany</affiliation>
</name>
```

Curation des régions



Curation des régions

ville	code 4 chiffres	code 5 chiffres	formes courantes	district/land
Aix-la-Chapelle	W-5100	52056-52080	Aachen	region @type=land @nuts=1 : Rhénanie-du-Nord-Westphalie ; region @type=district @nuts=2 : District de Cologne
Augsbourg	W-8900	86000-86199	Augsbourg	region @type=land @nuts=1 : Bavière ; region @type=district @nuts=2 : District de Souabe
Bayreuth	W-8580	95444-95448	Bayreuth	region @type=land @nuts=1 : Bavière ; region @type=district @nuts=2 : District de Haute-Franconie
Berlin	W-1000	10115		
Bonn	W-5300	53111		

```
<r>
  <c1>
    <p>
      <k>Aix-la-Chapelle</k>
      <t>Aix-la-Chapelle</t>
    </p>
  </c1>
  <c2>
    <l>W-5100</l>
  </c2>
  <c3>
    <i>52056-52080</i>
  </c3>
  <c4>
    <l>Aachen</l>
  </c4>
  <c5>
    <region type="land" nuts="1">Rhénanie-du-Nord-Westphalie</region>
    <region type="district" nuts="2">District de Cologne</region>
  </c5>
  <c6>
    <l>
      </l>
    </c6>
  </r>
```

Curation des régions



navigation

- [Accueil](#)
- [Communauté](#)
- [Actualités](#)
- [Modifications récentes](#)
- [Index alphabétique](#)
- [Index thématique](#)
- [Page au hasard](#)
- [Aide](#)

rechercher

[Jacques Ducloy](#) [page de discussion](#) [préférences](#) [liste de suivi](#) [contributions](#) [déconnexion](#)

[wicri](#) [discussion](#) [modifier](#) [historique](#) [supprimer](#) [renommer](#) [protéger](#) [suivre](#) [réactualiser](#)

Wicri:Liste de grandes universités allemandes

Cette page introduit une liste destinée à mettre au point des mécanismes d'identification géographiques à partir d'une mention d'université. Elle fait partie d'un réseau de pages de même type dont la tête est sur [Wicri/Métadonnées](#).

Elle fait également partie des réseaux de listes propres à l'Allemagne, voir [Wicri:Liste de listes relatives à l'Allemagne](#).

Liste des universités

[\[modifier\]](#)

Université technique de Berlin	Technische Universität Berlin	country : Allemagne ; region @type=capital : Berlin ; settlement @type=city : Berlin
Université de Cologne	Universität zu Köln	country : Allemagne ; region @type=land @nuts=1 : Rhénanie-du-Nord-Westphalie ; region @type=district @nuts=2 : District de Cologne ; settlement @type=city : Cologne

Perspectives techniques : boîte à outil XML

- ▶ Objectif : développer des services d'exploration de connaissance centrés sur les besoins des utilisateurs
- ▶ Court terme (peu de moyens) : prototype déjà puissant
 - utilisable en mode « serveur -> clients »
 - Référence 1997 : Ilib pour MIRIAD (Pascal/Francis) à l'INIST
 - En 2018 : des serveurs adaptés à la complexité des disciplines scientifiques.
 - Exemples, par région, par thématique
 - services accompagnés sur machine(s) virtuelle(s) partagée par
 - Des informaticiens
 - au moins 1 à temps partiel pour la génération de serveur
 - Des praticiens maîtrisant l'algorithmique (exemple : outils thématiques)
 - Des bibliothécaires (accompagnement learning center)
 - Des praticiens, des chercheurs, des enseignants...
- ▶ Moyen terme : une bibliothèque XML « pour tous »
 - diffusable sur ordinateur personnel (génération de serveurs)
 - Projet de développement sur 3 ans
 - avec retombées sur 12 mois
- ▶ Cette remarque s'applique à tous les outils utilisés pour des enrichissements en amont

Perspectives techniques :

Réseau de wikis sémantiques

- ▶ Etat des lieux : Wicri/LorExplor, 10 ans, 1 site :
 - 35 familles de wikis de « communautés scientifiques »
 - 130 wikis au total (multilinguisme, applications)
 - 200 serveurs d'exploration sur les 35 familles
 - Problème : aucune garantie de pérennité (ni CNRS, ni UL)
- ▶ Soutien logistique du démonstrateur actuel
 - Aucun problème technique pour passer par exemple à
 - 50 familles, 200 wikis, 500 serveurs sur 3 sites
 - Seul besoin :
 - gérer les usagers LDAP (10 jours)
 - Pérennité...
- ▶ Déploiement ISTEK au niveau national
 - 10 sites physiques avec 100 wikis / 100 serveurs
 - Boite à outil XML => Robots (cohérence sémantique et administration)

Formation « learning center »

- ▶ Expérimentation (en démarrage) à la BU lettres de Nancy : ateliers de formation
 - exploration ISTEEX -> édition -> société, citoyens
- ▶ Exemples de progression
 - Démarrage
 - *A partir du dictionnaire des lorrains illustres écrit par Dom Calmet en 1711 et numérisé sur Gallica, rédiger une page biographique qui sera liée à un auteur inconnu trouvé dans un ouvrage écrit en 1850...*
 - En phase intermédiaire
 - Explorer un serveur ISTEEX pour repérer une université experte dans un domaine et compléter sa page
 - Même exercice avec création d'une page université avec son ancrage géographique et son insertion dans le réseau de wikis...
 - Paramétrer des actions de curation
 - Puis :
 - Paramétrer la création d'un nouveau serveur d'exploration
 - Utilisation d'outils style grep pour filtrer des paragraphes d'un ouvrage de 1000 pages sur Gallica
 - Ecrire un filtre en Lex...

