



# Le projet TECTONIQ pour la valorisation du patrimoine numérique de l'industrie textile : application au territoire du Nord-Pas-de-Calais



Eric Kergosien

ENADOC, journée *données patrimoniales*, Juin 2016

Gérico

STL  
avoirs  
langage  
extes

Image  
Ville  
Environnement

TETIS

eric

Un projet national pluri-disciplinaire

cnrs  
dépasser les frontières

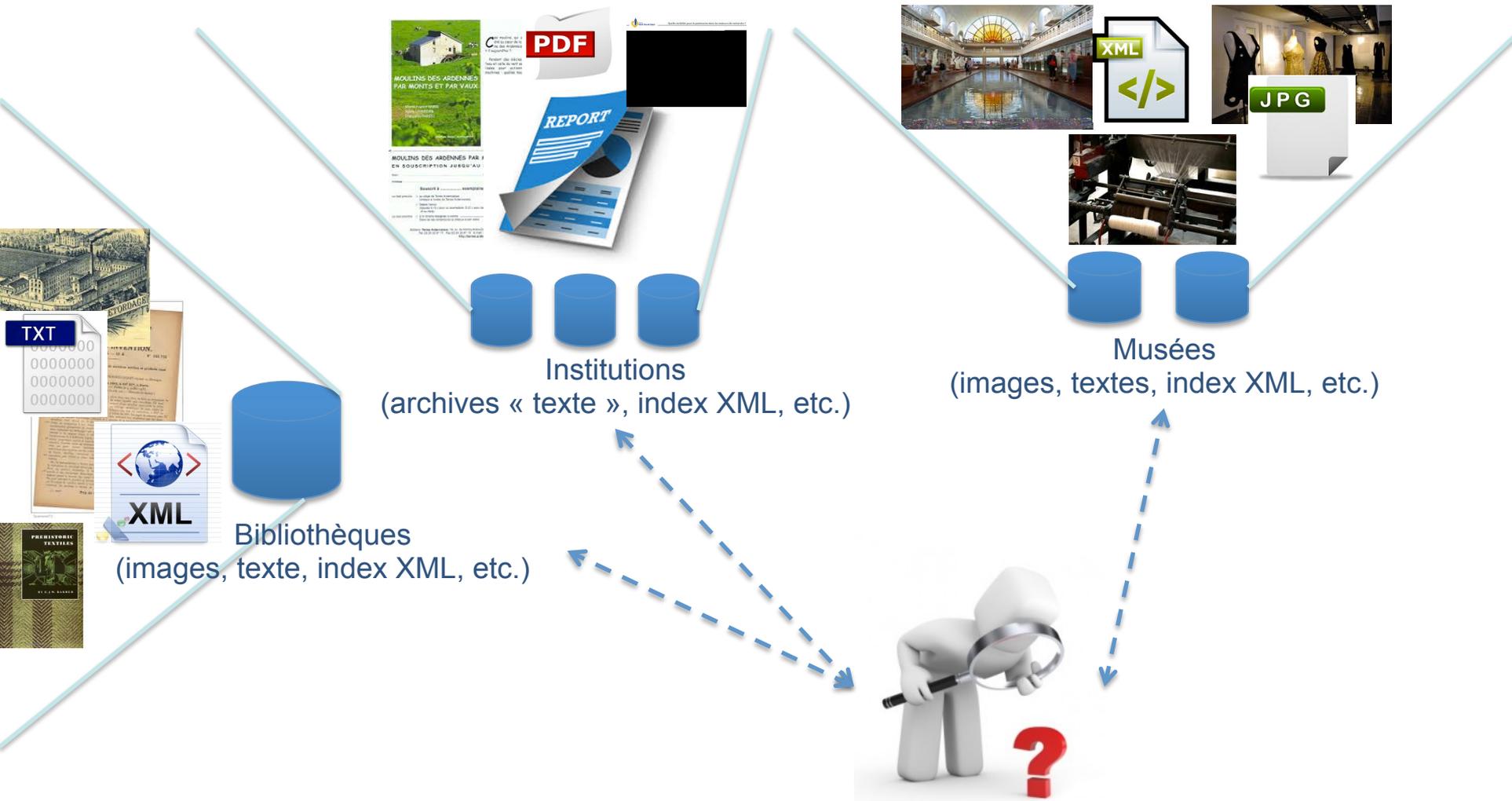
MESHs  
Lille Nord de France

Maison européenne  
des sciences de l'homme  
et de la société

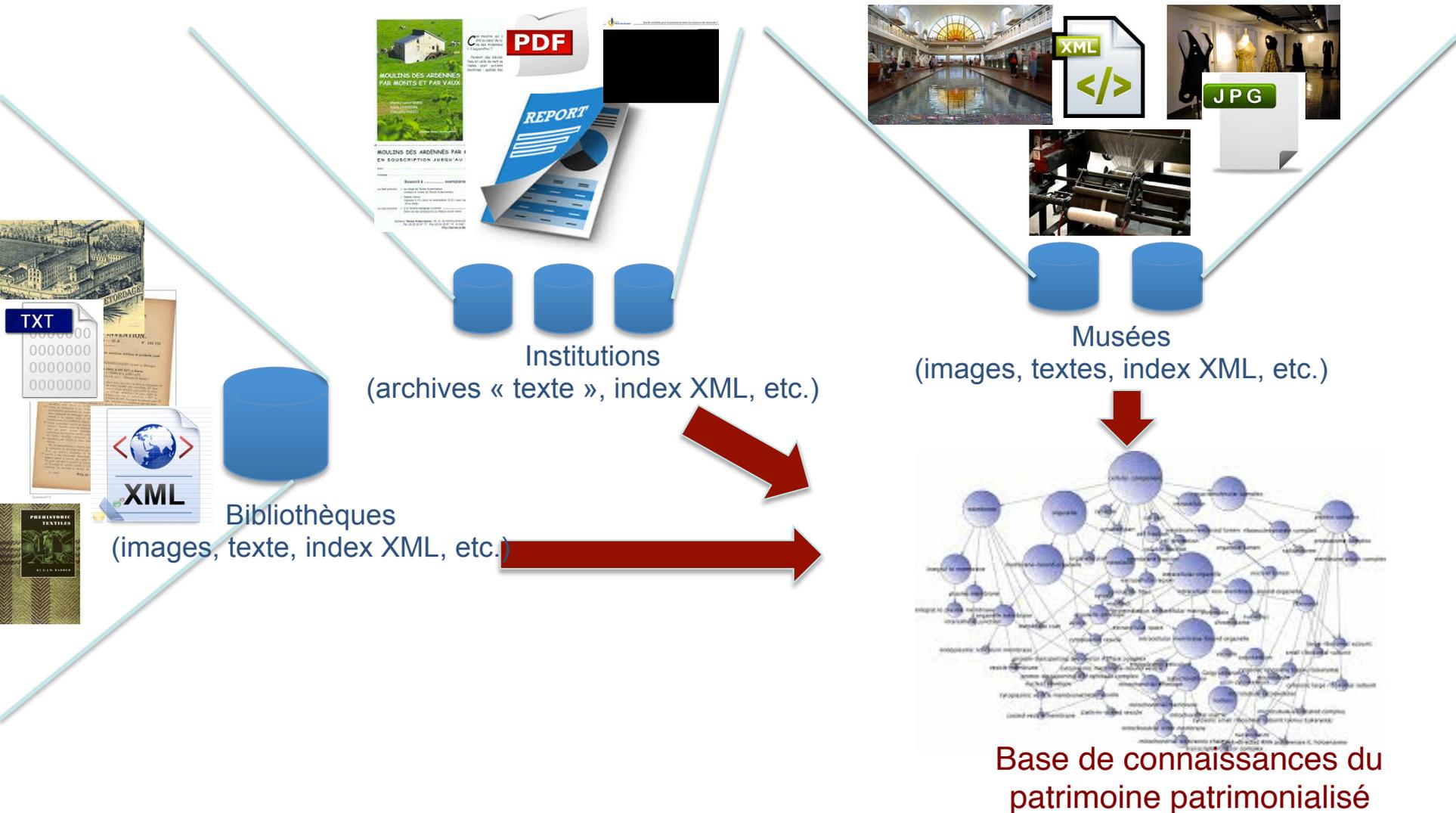
Université  
de Lille  
3 SCIENCES HUMAINES  
ET SOCIALES

Construire une base de connaissances portant sur un domaine spécifique à partir de sources de données électroniques hétérogènes

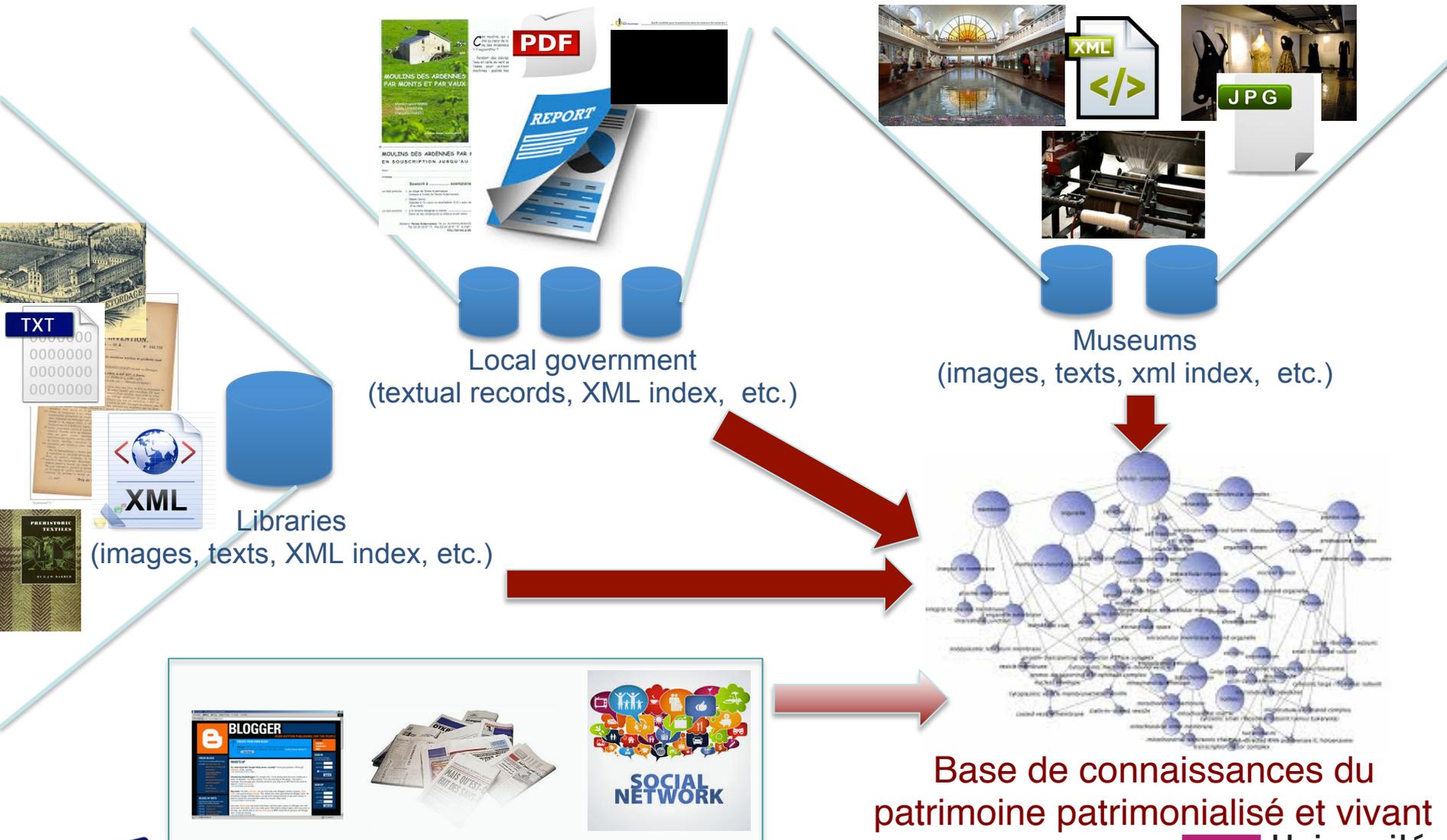
- Domaine visé : le patrimoine industriel textile
  - Périmètre géographique : la Région Nord – Pas-de-Calais
  - Durée du projet : 3 à 5 ans (fin de la 1ère année)
  - Financement : PEPS CNRS, Ministère de la culture, Université Lille 3
- 
- **Principaux buts, public-cible ?**
    - aider les experts dans la production de données numériques (documents textuels, index, etc.)
    - fournir une représentation d'ensemble synthétique du domaine visé dans le périmètre de référence pour le grand public



## Principaux défis



## Principaux défis



Construire une base de connaissances portant sur un domaine spécifique à partir de sources de données électroniques hétérogènes

- Domaine : patrimoine industriel textile
- Le territoire étudié : Région Nord Pas de Calais

Phases de travail

- a. Phase de cadrage : définitions
- b. Identification des réseaux d'acteurs du patrimoine industriel textile (patrimoine culturel, patrimoine vivant, médias, etc.)
- c. Collecte des données hétérogènes
- d. Extraction semi-automatique des informations (infos spatiales, temporelles, thématiques, acteurs, organismes, objets patrimoniaux, etc.)
- e. Normalisation et structuration des éléments d'information extraits

## a. Phase de cadrage : définitions

---

- 1982 : Déclaration de Mexico sur les politiques culturelles, Conférence mondiale sur les politiques culturelles (UNESCO) : «Le patrimoine culturel d'un peuple s'étend aux oeuvres de ses artistes, de ses architectes, de ses musiciens, de ses écrivains, de ses savants, aussi bien qu'aux créations anonymes, surgies de l'âme populaire, et à l'ensemble des valeurs qui donnent un sens à la vie. Il comprend les **oeuvres matérielles et non matérielles** qui expriment la créativité de ce peuple : **langue, rites, croyances, lieux et monuments historiques, littérature, oeuvres d'art, archives et bibliothèques,**
- 2011 : Principes conjoints ICOMOS TICCIH pour la conservation des sites, constructions, aires et paysages du patrimoine industriel : « Le patrimoine industriel comprend les **sites, les constructions, les complexes, les territoires et les paysages** ainsi que les **équipements, les objets ou les documents** qui témoignent des procédés industriels anciens ou courants de production par l'extraction et la transformation des matières premières ainsi que des infrastructures énergétiques ou de transport qui y sont associées. Il exprime une relation étroite entre l'environnement culturel et naturel puisque les procédés industriels – anciens ou modernes – dépendent de ressources naturelles, d'énergie et de voies de communication pour produire et distribuer des biens sur les marchés. Ce patrimoine comporte des dimensions immatérielles comme les savoir-faire techniques, l'organisation du travail et des travailleurs ou un héritage complexe de pratiques sociales et culturelles résultant de l'influence de l'industrie sur la vie des communautés et sur la mutation des sociétés et du monde en général. »

L'industrie textile génère du patrimoine matériel et immatériel :

- patrimoine matériel : **bâtiments, sites, infrastructures, complexes, domaines, usines, entrepôts, centres de production, machines, équipements, employés, organismes, etc.**
- patrimoine immatériel : **opinions, événements, langages, rites, savoir-faire, etc.**

Territoire d'études :



## b. Identifier les réseaux des acteurs

Identifier les acteurs en lien avec le patrimoine culturel



Liste d'acteurs



9 entretiens semi-directifs (collectivités locales)



60 acteurs

Construction d'une base de données (infos sur les acteurs et les données)



Création d'un index pour chaque acteur



Identifier les acteurs en lien avec le patrimoine vivant (web)

Sourcing web :

- identification par mots-clefs
- identification de types de sources
- construction de requêtes approfondies dans un moteur de recherche
- évaluation des résultats



58 acteurs

Crawl du site web de chaque acteur pour identifier les hyperliens

160 acteurs



hyphe

Cartographie du Web

169 acteurs



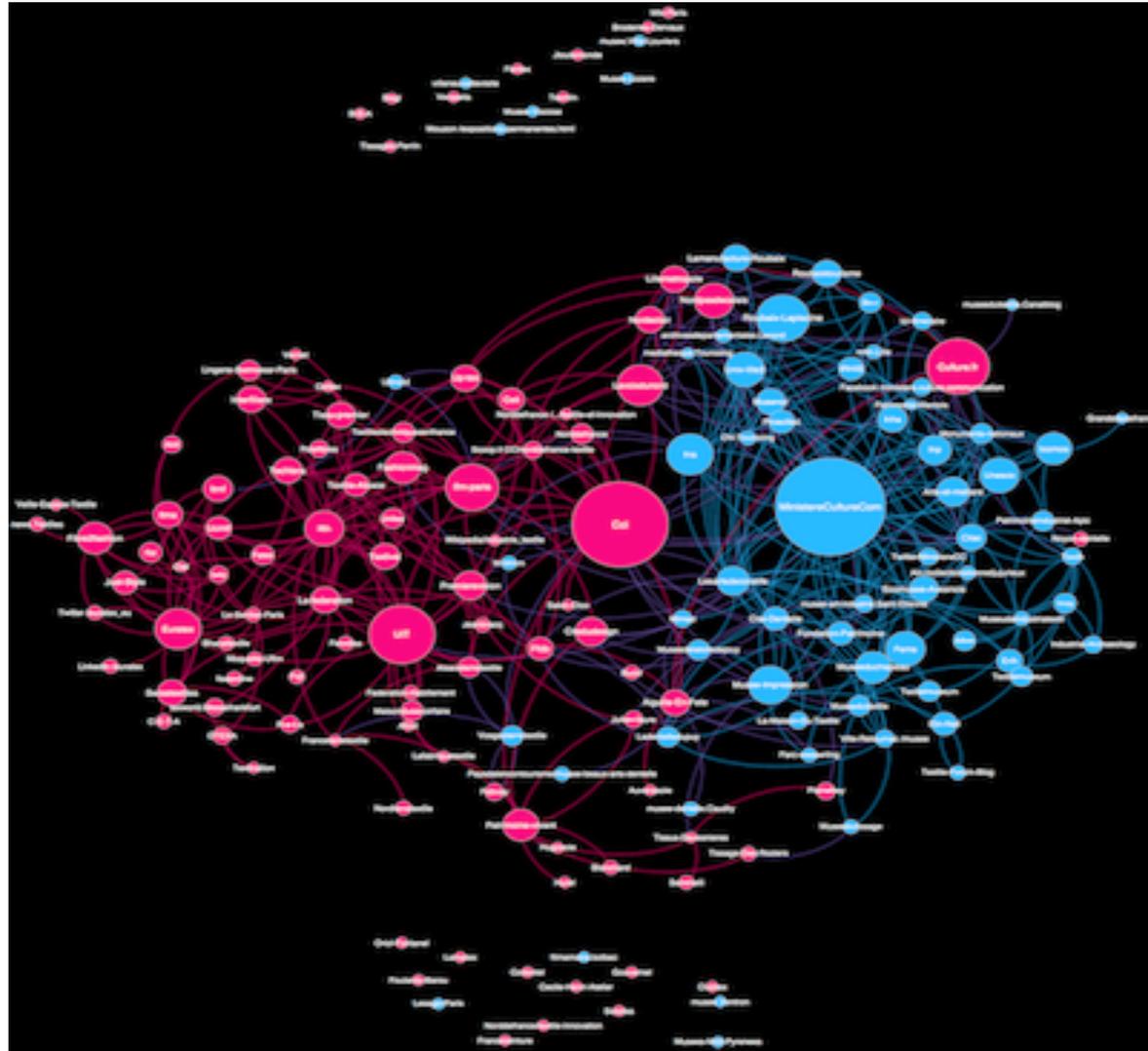
Gephi

## b. Identifier les réseaux des acteurs

Cartographie des 169 acteurs identifiés sur le web, organisés par type :

- patrimoine historique et culturel : 40,83 %
- patrimoine vivant (associations, sociétés) : 59,17 %.

✓ Plus un acteur est mentionné, plus les autres acteurs s'accordent sur son utilité, sa pertinence et son importance (Severo et Venturini, 2015)



Graphe généré avec Gephi, algorithme force-atlas 2.

## b. Identifier le réseau des acteurs

✓ Les principaux acteurs connectés :

→ patrimoine historique et culturel ::

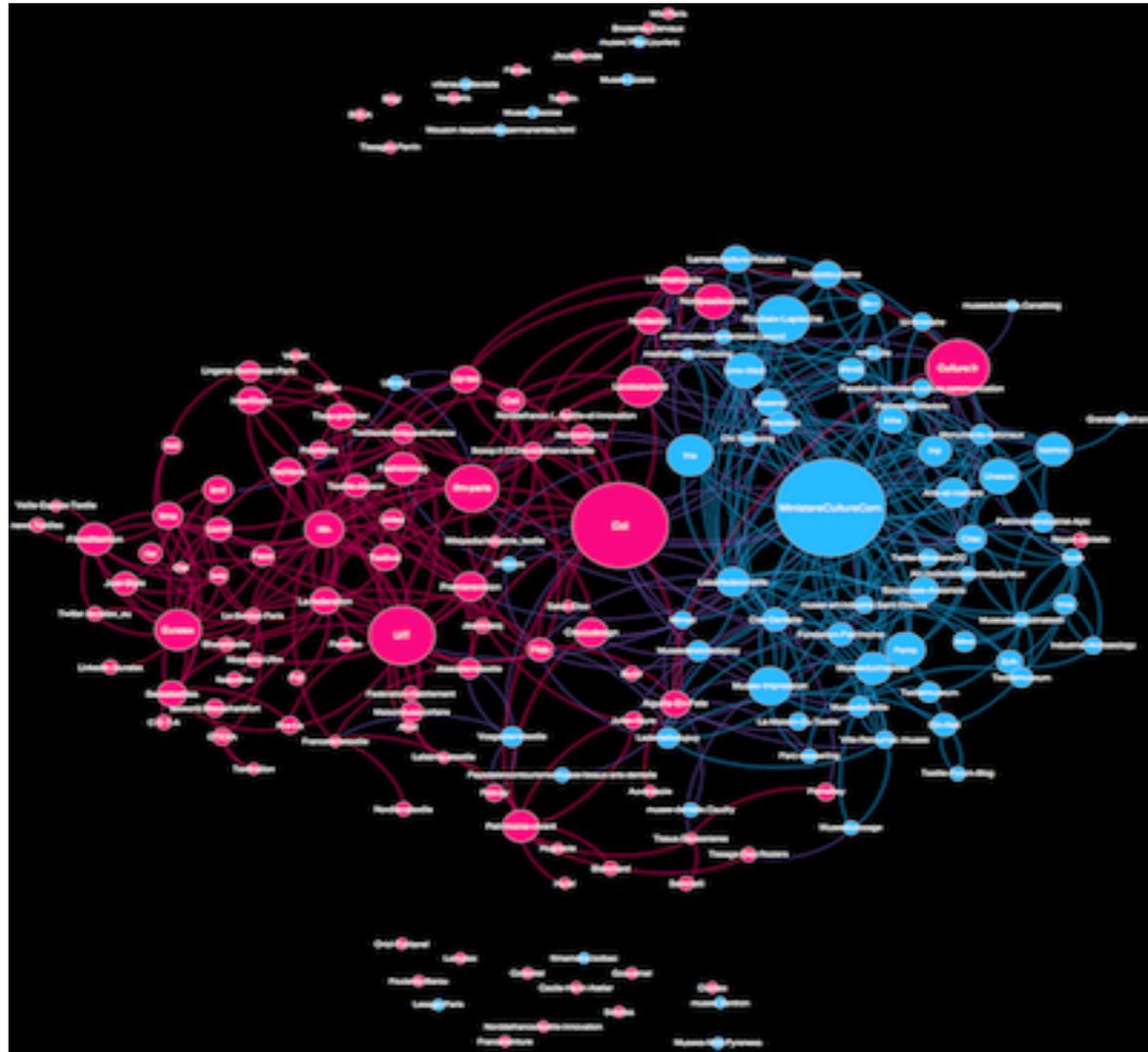
- portail de la Région NPDC
- Ministère de la Culture et de la Communication
- musée de la Piscine de Roubaix
- CILAC (asso patrimoine industriel)
- ...

→ Patrimoine vivant :

- Chambre de Commerce et d'Industrie
- L'Union des Industries Textiles
- PROSCITEC (asso mémoire des métiers)

✓ Répartition géographique :

- 29% internationale
- 30,7% nationale
- 40,3% régionale



## b. Identifier le réseau des acteurs

Localisation des acteurs agissant en France (métropolitaine)



Nord-Pas-de-Calais	(37,76 %)
Rhone-Alpes	(23,47 %)
Ile-de-France	(8,16 %)
Alsace	(5,1 %)
Lorraine	(4,08 %)
Pays-de-la-Loire	(4,08 %)
Champagne-Ardenne	(3,06 %)
Basse-Normandie	(3,06 %)
Languedoc-Roussillon	(2,04 %)
Auvergne	(2,04 %)
Midi-Pyrenees	(2,04 %)
Centre	(1,02 %)
Picardie	(1,02 %)
PACA	(1,02 %)
Aquitaine	(1,02 %)
Haute-Normandie	(1,02 %)

Patrimoine  
patrimonialisé  
(48,65 %)

Patrimoine vivant  
(51,35%)

Travaux complets : <http://www.ekergosien.net/tectoniq/>



- Collecte des données : en collaboration avec plusieurs acteurs
- Extraction des éléments d'information à partir d'index et de données Textuelles
  - approches de type TAL pour extraire les éléments d'information (infos spatiales et temporelles, thématiques, organismes)
  - utilisation du visualiseur SentiAnnotator pour le processus d'extraction et de validation par des experts
- Structuration des informations dans une ontologie de domaine
  - exemple construit manuellement dans le modèle conceptuel CIDOC CRM (standard officiel du Conseil International de Musées)

- Hard data (sources « expertes ») :
  - données provenant du Ministère de la Culture : bases de données sur le patrimoine bâti national (format Excel)
  - Métropole Européenne de Lille
    - » département Culture : données géographique (GIS) de l'agglomération
    - » département Numérique : plan local d'urbanisme (données géographiques et patrimoniales) : 50 documents pdfs décrivant le bâti.
  - Musées : données de La Piscine et du LAM
  - UNESCO : les données ouvertes
  - Service de l'Inventaire, DRAC Nord – Pas-de-Calais : plusieurs bases de données : 100 rapports pdf thématiques
  - bibliothèques : 3 corpus annotés (IRHIS , SCD Lille 3, etc.) : 7 000 documents numérisés et OCRisés
- Soft data (sources médiatiques et entreprises)
  - Presse de la voix du Nord collectée par l'association La Tossée des anciens du textile : 3000 articles entre 2000 et 2016



Le démonstrateur SentiAnnotator facilite le processus d'extraction et la validation par les experts (ISWC, octobre 2015)

- ✓ Importation de grands corpus (français, anglais)
- ✓ Outils TAL pour l'extraction des éléments d'information
- ✓ Les experts peuvent confirmer/informer les données identifiées
- ✓ Possibilité d'exporter les résultats



The screenshot displays the SentiAnnotator web interface. The browser address bar shows the URL: <http://172.16.10.226:34000/?Organization=True&SF=True&Opinion=True>. The interface is divided into several sections:

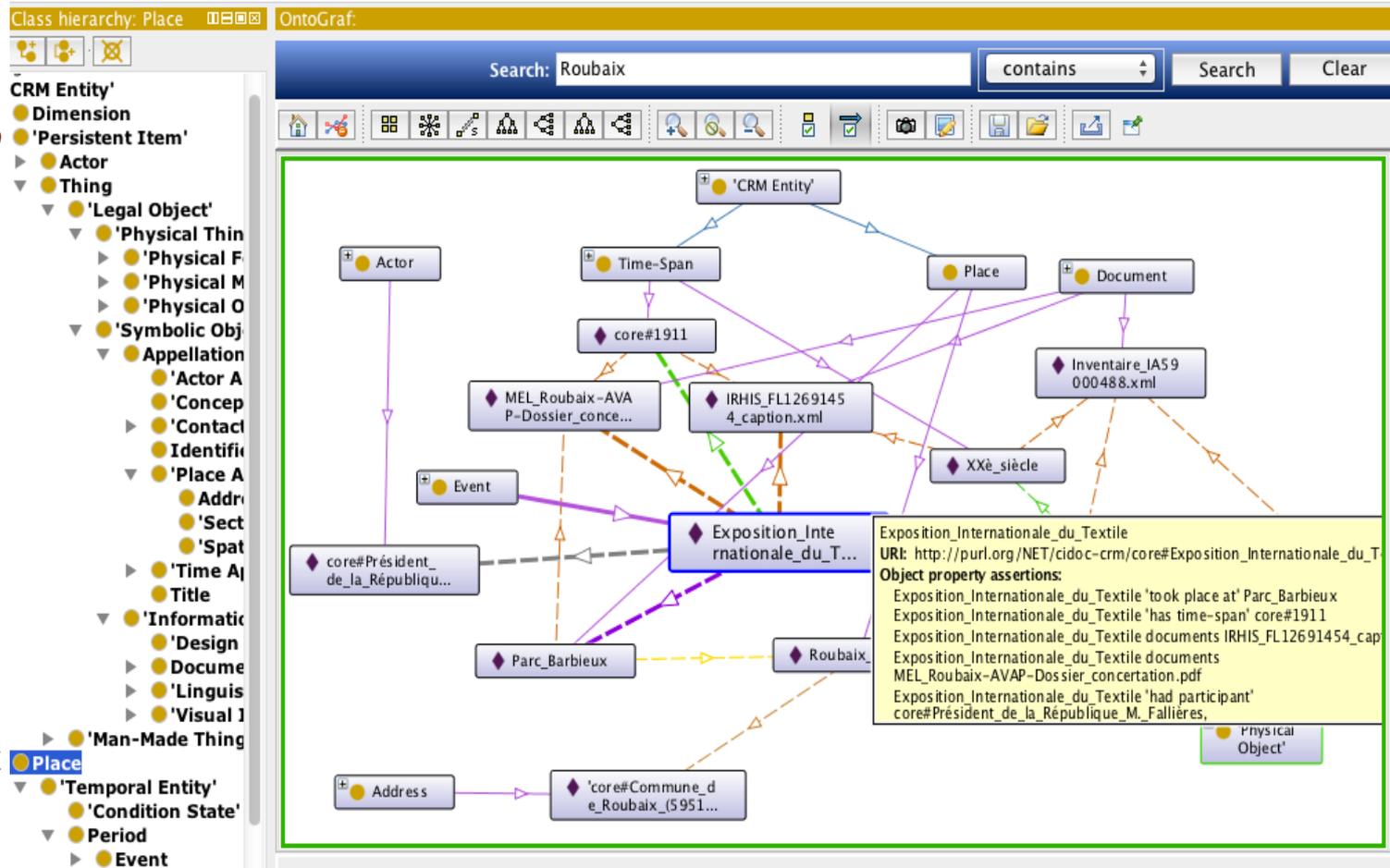
- SENTERRITOIRE VIEW**: Contains a 'DISPLAYED INFORMATION' panel with checkboxes for Spatial Features, Organization, Opinions, Other, and Topic. Below it is a 'CORPUS AND DOCUMENTS' tree view showing a hierarchy of documents like '1\_47\_2\_docs\_With\_NERs' and '2\_48\_7\_docs\_With\_NERs'.
- DOCUMENTS**: The main text area showing two document excerpts. The first excerpt discusses David Bowie's music, with highlighted entities like 'David Bowie', 'Heathen', 'Tony Visconti', 'Matt Chamberlain', and 'Nirvana'. The second excerpt discusses the CCI (Chambre de Commerce et d'Industrie) and the Région, with highlighted entities like 'CCI', 'Région', 'Frêche', 'Sète', and 'Montpellier'.
- MARKED INFORMATION**: A sidebar on the right containing three lists of extracted information:
  - Spatial Features (17)**: A list of locations including 'à Cougnenc', 'à Frêche', 'à la Région', 'à Port-La Nouvelle', and 'à Sète'.
  - Organizations (25)**: A list of organizations including 'la Région de payer la CCI', 'la Région était', 'la Région sur', 'la Région verse', 'la Région', 'les CCI', 'nationale', 'port de commerce', and 'port'.
  - Opinions (18)**: A list of opinion terms including 'avant', 'comme', 'Comme', 'exception', and 'secret'.
- UPLOAD NEW CORPUS**: A section at the bottom left for uploading new documents, with a form and a pipeline selection dropdown.
- Footer**: Shows processing statistics: 'Just Gate Process time (mm:ss) => 0:23, Total Process time (mm:ss) => 0:24' and 'Processed by Sentritoire Web services - 2014'.

<http://siso.teledetection.fr/>

Extrait de l'ontologie de domaine basée sur quatre documents hétérogènes (XML, PDF, texte)

Exemples :

- document XML *IRHIS\_FL1269145.xml* traite de l'implication du Président dans l'Exposition Internationale du Textile en 1911
- document pdf *MEL\_Roubaix\_AVA.pdf* fait état du déroulement de l'Exposition Internationale du Textile dans le Parc Barbieux à Roubaix
- ...



Logiciel Protégé (Musen et al., 1995)

- **Premiers résultats**

- Phase de cadrage : définitions issues de l'UNESCO et de TICCIH
- Proposition d'une méthode hybride qui combine des entretiens semi-dirigés et la cartographie web qualitative et quantitative exploitant les outils logiciels Navicrawler et Hyphe pour identifier de « nouveaux » acteurs
- Démonstrateur web pour extraire semi-automatiquement des éléments d'information portant sur le patrimoine industriel
- Premier exemple d'ontologie construite manuellement en OWL selon le modèle conceptuel CIDOC CRM à partir de l'information issue de quatre documents hétérogènes distincts, désormais reliés par leur contenu

- **Travaux en cours**

- Construction semi-automatique d'une base de connaissances à partir de notre corpus actuel de 1600 documents, en exploitant le modèle CIDOC CRM. Nous pouvons extraire des textes les éléments d'informations en lien avec le patrimoine industriel de ces types : lieux, événements, thématiques, entités temporelles, acteurs. Les outils et méthodes automatiques développés dans le laboratoire TETIS (Montpellier) sont exploités sur ces 1600 documents.

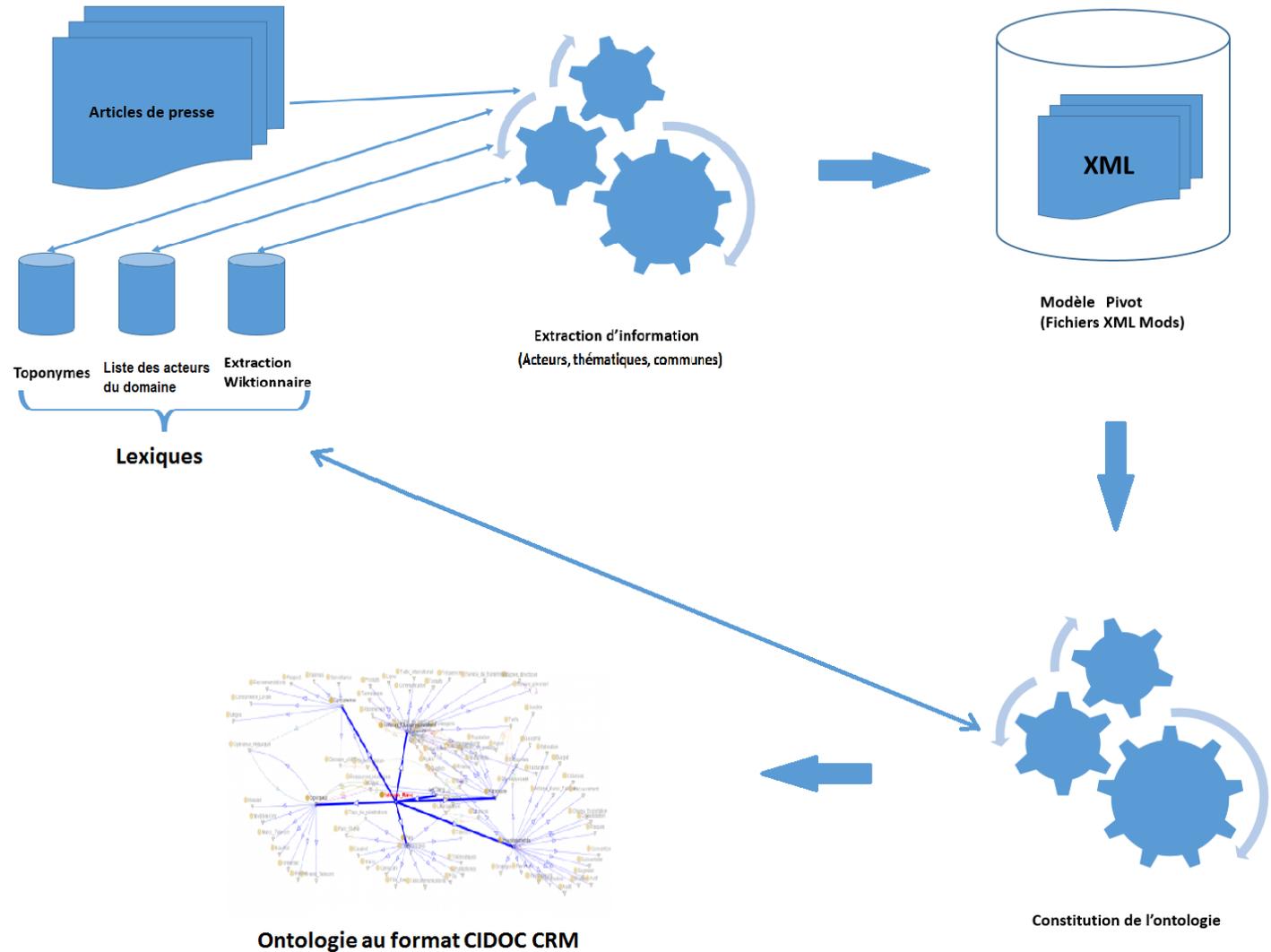
- **E. Kergosien, B. Jacquemin, M. Severo et S. Chaudiron**, Vers l'interopérabilité des données hétérogènes liée au patrimoine industriel textile, In 18ème colloque international sur le document électronique (CIDE'18), 15p, 2015(a),
- **M. Severo**, T. Venturini, *Intangible Cultural Heritage webs. Comparing national networks with digital methods*, in *New Media & Society*, to appear, 2015.
- **L. Favier, W. Mustafa El Hadi**, D. Vinck, Interopérabilité Culturelle, *Revue Communication*, Université de Laval, 2015
- **M.A. Berthelot, M. Severo, E. Kergosien**, *Cartographier les acteurs d'un territoire : une approche appliquée au patrimoine industriel textile du Nord Pas de Calais*, In **Colloque « En quête de territoire(s) / Looking for territories »**, to appear, 17-18 mars 2016, Grenoble
- **E. Kergosien, B. Jacquemin, M. Severo et S. Chaudiron**, The TECTONIQ project for valorization of digital textile industrial heritage in North of France, In the XVIth International TICCIH congress, 2015(b).
- **M.A. Berthelot**, Identifier et cartographier les acteurs d'un territoire : cas appliqué à la thématique du patrimoine de l'industrie textile de Lille Métropole, Mémoire de stage du Master 2 IDEMM, département Sciences de l'Information et du Document, Université Lille 3, 2016.
- **M. Roche, M. Teisseire**. Traitement automatique des données hétérogènes liées à l'aménagement des territoires, In the colloque ASRDLF'15 (Association de Science Régionale de Langue Française), 2015.
- M. A. Farvardin, **E. Kergosien, M. Roche and M. Teisseire**, A webtool for analyzing land-use planning documents. In *Proceedings of 14th International Semantic Web Conference (ISWC 2015)*, CEUR, Bethlehem, United-States, 2015.
- **E. Kergosien**, Le projet TECTONIQ pour la valorisation du patrimoine numérique de l'industrie textile : application au territoire Nord Pas de Calais en France, congrès de l'ACFAS (Sciences de l'information et la communication), Mai 2016, Montréal

# Questions



<http://tectoniq.meshs.fr>

## Extraction de thématiques



- Lexique de 330 termes construit automatiquement à partir WiktionaryX et de GLAWI (dictionnaire électronique libre au format XML issu du Wiktionnaire) :
  - termes dont au moins une définition comporte le mot 'textile' ;
  - termes dont au moins une définition comporte un attribut qui contient le mot 'textile' ;
  - termes dont au moins une catégorie contient le mot 'textile'.
- Travail en cours :
  - Alignement RAMEAU – GEONAMES : <http://data.bnf.fr/semanticweb#Ancre2>
  - Alignement lexiques Joconde :  
<http://www.culture.gouv.fr/documentation/joconde/fr/partenaires/AIDEMUSEES/telechargement.htm>

```

<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  <titleInfo> ☐☐ </titleInfo>
  <typeOfResource>text</typeOfResource>
  <genre> ☐☐ </genre>
  <originInfo> ☐☐ </originInfo>
  <subject>
    <topic>patrimoine textile</topic>
    <topic>industrie textile</topic>
    <hierarchicalGeographic>
      <country>France</country>
      <state>Nord</state>
      <city>Roubaix</city>
    </hierarchicalGeographic>
  </subject>
  <subject> ☐☐ </subject>
  <subject> ☐☐ </subject>
  <subject>
    <name type="corporate">
      <namePart type="given">musée de la Rubanerie</namePart>
      <affiliation>patrimoine textile</affiliation>
    </name>
    <hierarchicalGeographic>
      <country>France</country>
      <state>Nord</state>
      <city>Comines</city>
    </hierarchicalGeographic>
  </subject>
  <subject> ☐☐ </subject>
  </mods>

```

```
- <subject>
  - <hierarchicalGeographic>
    <country>France</country>
    <state>Nord</state>
    <city>Villeneuve-d'Ascq</city>
  </hierarchicalGeographic>
  - <cartographics>
    <coordinates> 50° 37' 10" N, 3° 07' 54" E</coordinates>
  </cartographics>
</subject>
```



```
<!-- http://www.semanticweb.org/kouku/ontologies/2016/5/untitled-ontology-7#Villeneuve-d&apos;Ascq -->

<owl:NamedIndividual rdf:about="&untitled-ontology-7;Villeneuve-d&apos;Ascq">
  <rdf:type rdf:resource="&untitled-ontology-7;E1_CRM_Entity"/>
  <rdf:type rdf:resource="&untitled-ontology-7;E28_Conceptual_object"/>
  <rdf:type rdf:resource="&untitled-ontology-7;E41_Appellation"/>
  <rdf:type rdf:resource="&untitled-ontology-7;E44_Place_Appellation"/>
  <rdf:type rdf:resource="&untitled-ontology-7;E53_Place"/>
  <rdf:type rdf:resource="&untitled-ontology-7;E70_Thing"/>
  <rdf:type rdf:resource="&untitled-ontology-7;E71_Man_made_thing"/>
  <rdf:type rdf:resource="&untitled-ontology-7;E77_Persistent_item"/>
  <rdf:type rdf:resource="&untitled-ontology-7;E90_Symbolic_Object"/>
  <untitled-ontology-7;P1_is_identified_by rdf:resource="&untitled-ontology-7;50°_37'_10"_N,_3°_07'_54"_E"/>
  <untitled-ontology-7;P87_is_identified_by rdf:resource="&untitled-ontology-7;50°_37'_10"_N,_3°_07'_54"_E"/>
  <untitled-ontology-7;P88_consists_of rdf:resource="&untitled-ontology-7;Nord"/>
  <untitled-ontology-7;P1_is_identified_by rdf:resource="&untitled-ontology-7;Villeneuve-d&apos;Ascq"/>
  <untitled-ontology-7;P87_is_identified_by rdf:resource="&untitled-ontology-7;Villeneuve-d&apos;Ascq"/>
</owl:NamedIndividual>
```

A	B	C	D	E	F	G	H	I	J	K
Id	Label	Status	ActeurProjet	PatrimoineVi	Statut	Catégorie	Rayonnement	Spécification	Locale	
c6188378-4452-4219-a2be-723c6f6febb9	Fems	IN	non	non	Association	Organisation	National	France		
8f12c5e2-fc65-4ffc-9971-300d3903495b	fr.Wikipedia /.../Industrie_textile	IN	non	oui	Association	Educatif	Internationale	Monde		
031bf5d5-6f8e-4ba2-8627-0f1f1c01373a	Ecomusee-Avesnois	IN	non	non	Institution	Culturel	Metropolitain	Fourmies		
0cd2801d-a846-4d95-bae0-3efcef4faf7f	Patrimoineindustriel-Apic	IN	non	non	Association	Culturel	Regional	Champagne-Ardenne		
033262e8-b9a2-4c4a-a3d1-30dc0e7715ff	Industrial-Archaeology	IN	non	non	Association	Culturel	National	Grande-Bretagne		
0b6a5c76-7d50-4c64-a8ba-697d8d837903	Nordpasdecalais	IN	non	oui	Institution	Actualite	Regional	Nord-Pas-de-Calais		
1861a9bd-ae0-4ee7-bcde-fede45e23658	Lamanufacture-Roubaix	IN	non	non	Institution	Culturel	Metropolitain	Roubaix		
220c76bb-05df-441c-a131-75f823a0a4d3	Museedelatoiledejouy	IN	non	non	Institution	Culturel	Metropolitain	Jouy-en-Josas		
23c771a4-eaca-46a4-85d2-249b55e809df	Etn-Net	IN	non	non	Association	Culturel	Internationale	Europe		
3f939336-5aad-4133-99a2-417a50648570	Textival	IN	non	oui	Association	Evenementiel	Regional	Rhone-Alpes		
2e82834d-3f64-41bb-a701-532140ea14cb	Ifth	IN	non	oui	Association	Organisation	Internationale	France		
5c8888bd-8354-4b75-9be4-54f998daa217	Textilemuseum	IN	non	non	Institution	Culturel	National	Canada		
4e829532-bbe1-4db5-b53d-0953256d6f4e	Unesco	IN	non	non	Institution	Culturel	Internationale	Monde		
5a692cb4-3d3e-4f77-b3a7-f7ed20af38ac	Fondation-Patrimoine	IN	non	non	Association	Culturel	National	France		
5f651cb0-5792-4669-b8c2-1e08e49c2df4	Musee-Impression	IN	non	non	Institution	Culturel	Metropolitain	Mulhouse		
4c254f1a-f860-41bf-93ae-8e8c64429bbd	Roubaixtourisme	IN	non	non	Institution	Culturel	Metropolitain	Roubaix		
5b8c8f54-232e-49e8-98f4-644c880a0477	Proscitec	IN	non	non	Association	Culturel	Regional	Nord-Pas-de-Calais		
55f3c92c-2108-4f4d-b88c-422b7a376ce8	Ticcih	IN	non	non	Association	Culturel	Internationale	Monde		
94ec7294-e9ec-46b1-a72f-67558c59e47c	La-Maison-Du-Textile	IN	non	non	Institution	Culturel	Regional	Aisne		
76d1060d-9a06-4a02-a773-d78bdc4c45a9	Lesartsdecoratifs	IN	non	non	Institution	Culturel	National	France		
7de01f6b-ad73-4127-a197-e819b479e3a7	Textile-Alsace	IN	non	oui	Association	Organisation	Regional	Alsace		
6d9598bb-0c71-4994-9b60-40ebdd8924d	Ici-Itineraire	IN	non	non	Institution	Culturel	Regional	Nord-Pas-de-Calais/Hainaut Belge		
9b87ef34-8560-4896-9061-93ef75d9e16f	Textile	IN	non	oui	Association	Organisation	Internationale	France		
a2158aed-a263-4abf-b036-da2ae0511cf4	Patrimoine-Mantois	IN	non	non	Association	Culturel	Metropolitain	Mantes		
bd656c8b-b5fe-468c-a1ba-76c3675b3730	Textile-Forum-Blog	IN	non	non	Association	Culturel	Internationale	Europe		
aeec00ec-0cd5-4780-9cd7-b294290a91d9	Grandsitedefrance	IN	non	non	Institution	Culturel	National	France		
c4f265c3-babc-4064-a5c6-400207ed5934	Roubaix-Lapiscine	IN	non	non	Institution	Culturel	Metropolitain	Roubaix		
b5189000-199b-4e6f-a448-236b6dfe09ab	Aiguille-En-Fete	IN	non	oui	Association	Evenementiel	National	France		
c05c833e-3bb7-444e-bed0-4dc3555378a0	Norddefrance	IN	non	oui	Institution	Actualite	Regional	Nord-Pas-de-Calais		
b86f3c40-d3ba-4c3b-9802-c3c7847ed613	Musee-Lozere	IN	non	non	Individu	Culturel	Regional	Lozere		
fc83b031-8e12-43db-b0fc-101f337b100b	Museudaindustriatextil	IN	non	non	Institution	Culturel	National	Portugal		
dd8c6f77-4b7c-4089-ac64-e3d3b8fff7cf	Museudedutextile	IN	non	non	Institution	Culturel	Metropolitain	Cholet		

Nombre de sites identifiés manuellement : 130

Nombre de sites crawlés automatiquement via NaviCrawler : 550 → 150 après nettoyage

Nombre de sites via Hyphe : 169

