

1. IRHIS Base Images

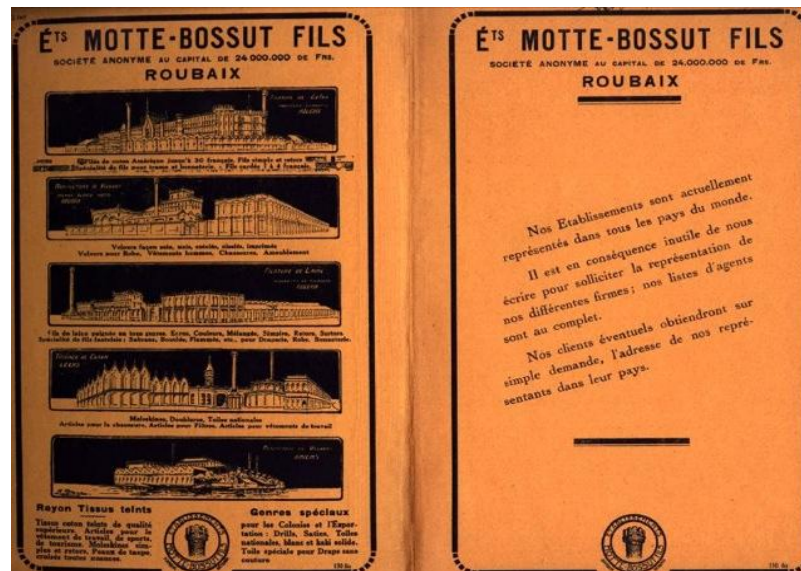
Institut de Recherches Historiques du Septentrion

"Institute of Historical Research of the North"

All the Images based in IRHIS with their properties.

1. TitreEnregistrement
2. Description
3. MotsClefsAnalytiques
4. MotsClefsGeographiques
5. EpoqueEvenement
6. AnneeEvenement
7. ReferenceBibliographique
8. ProvenanceDocument
9. EtablissementDepositaire
10. Format
11. Support

Example:



2. La Voix Du Nord

History of Textile industry and incidents happened. Those have the following attributes.

SOURCE

AUTEUR

REFERENCE

ENQUETES
DATE
TITRE
SOUS-TITRE
CONTEXTE
PRESENTATION

```
<!DOCTYPE CORPUS [  
  <!ELEMENT CORPUS (DOCUMENT)*>  
  <!ELEMENT DOCUMENT (DESCRIPTION|EVALUATION|TEXTE)*>  
  <!ATTLIST DOCUMENT  
    id CDATA #REQUIRED>  
  <!ELEMENT DESCRIPTION (SOURCE|AUTEUR|REFERENCE|ENQUETES|DATE|TITRE|SOUS-  
TITRE|CONTEXTE|PRESENTATION)*>  
  <!ELEMENT SOURCE (#PCDATA)>  
  <!ELEMENT AUTEUR (#PCDATA)>  
  <!ELEMENT REFERENCE (#PCDATA)>  
  <!ELEMENT ENQUETES (#PCDATA)>  
  <!ELEMENT DATE (#PCDATA)>  
  <!ELEMENT TITRE (#PCDATA)>  
  <!ELEMENT SOUS-TITRE (#PCDATA)>  
  <!ELEMENT CONTEXTE (#PCDATA)>  
  <!ELEMENT PRESENTATION (LANGUE|STYLE)*>  
  <!ELEMENT LANGUE (#PCDATA)>  
  <!ATTLIST LANGUE  
    utf-8 CDATA #REQUIRED>  
  <!ELEMENT STYLE (#PCDATA)>  
  <!ELEMENT EVALUATION (NOTE)*>  
  <!ATTLIST EVALUATION  
    nombre CDATA #REQUIRED>  
  <!ELEMENT NOTE (#PCDATA)>  
  <!ATTLIST NOTE  
    valeur CDATA #REQUIRED>  
  <!ELEMENT TEXTE (#PCDATA)>  
]>
```

3. MEL

PDF Files, managed to extract the text. Could not enrich it with any attributes yet. Trying to do that with NER or LDA.

4. Inventory Service

History of Buildings.

```
<!DOCTYPE noticeSInventaire [  
  <!ELEMENT noticeSInventaire  
(edifice|localisation|airEtude|adresse|destinationSuccessives|denomination|partiesNonEtudies|elem  
entsRemarquables|epoqueConstruction|auteurs|historique|description|elevation|etages|grosOeuvres  
|typeCouverture|materiauxCouverture|couvrement|etat|propriete|typeEtude|redacteur|copyright|d  
ateEnquete|dateVersement|dateMiseAJour|lienDossier)*>  
  <!ATTLIST noticeSInventaire  
    reference CDATA #REQUIRED>  
  <!ELEMENT edifice (#PCDATA)>  
  <!ELEMENT localisation (#PCDATA)>  
  <!ELEMENT airEtude (#PCDATA)>  
  <!ELEMENT adresse (#PCDATA)>  
  <!ELEMENT destinationSuccessives (#PCDATA)>  
  <!ELEMENT denomination (#PCDATA)>  
  <!ELEMENT partiesNonEtudies (#PCDATA)>  
  <!ELEMENT elementsRemarquables (#PCDATA)>  
  <!ELEMENT epoqueConstruction (#PCDATA)>  
  <!ELEMENT auteurs (#PCDATA)>  
  <!ELEMENT historique (#PCDATA)>  
  <!ELEMENT description (#PCDATA)>  
  <!ELEMENT elevation (#PCDATA)>  
  <!ELEMENT etages (#PCDATA)>  
  <!ELEMENT grosOeuvres (#PCDATA)>  
  <!ELEMENT typeCouverture (#PCDATA)>  
  <!ELEMENT materiauxCouverture (#PCDATA)>  
  <!ELEMENT couvrement (#PCDATA)>  
  <!ELEMENT etat (#PCDATA)>  
  <!ELEMENT propriete (#PCDATA)>  
  <!ELEMENT typeEtude (#PCDATA)>  
  <!ELEMENT redacteur (#PCDATA)>  
  <!ELEMENT copyright (#PCDATA)>  
  <!ELEMENT dateEnquete (#PCDATA)>  
  <!ELEMENT dateVersement (#PCDATA)>
```

<!ELEMENT dateMiseAJour (#PCDATA)>

<!ELEMENT lienDossier (#PCDATA)>