

**ARCHIVES
NATIONALES**

Paris – Pierrefitte-sur-
Seine



**MINISTÈRE
DE LA CULTURE**

*Liberté
Égalité
Fraternité*

L'IA en appui des missions des archivistes

Analyser, classer, transcrire, valoriser

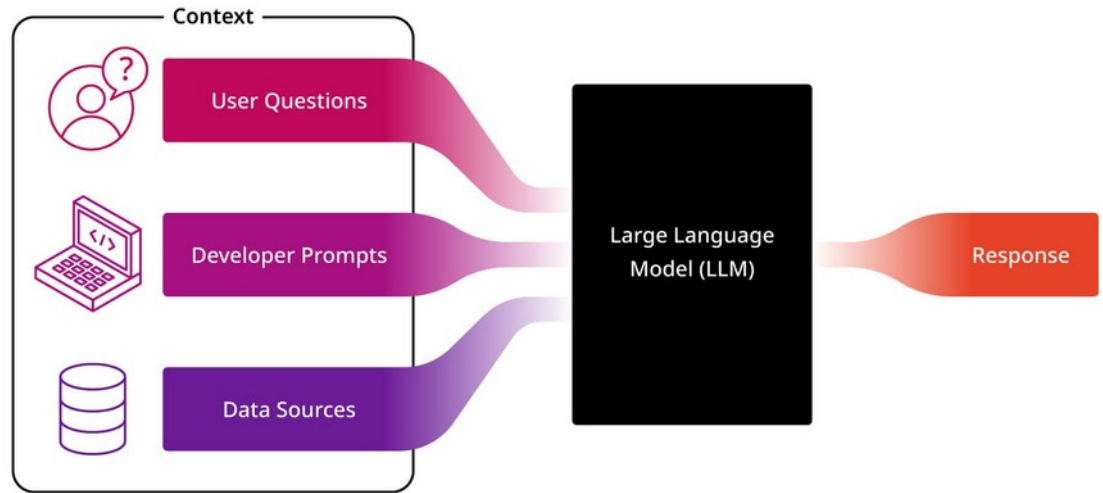
L'IA appliquée à la collecte et au traitement

Tirer parti de l'IA pour faire face à la masse

Technologies d'IA exploitées

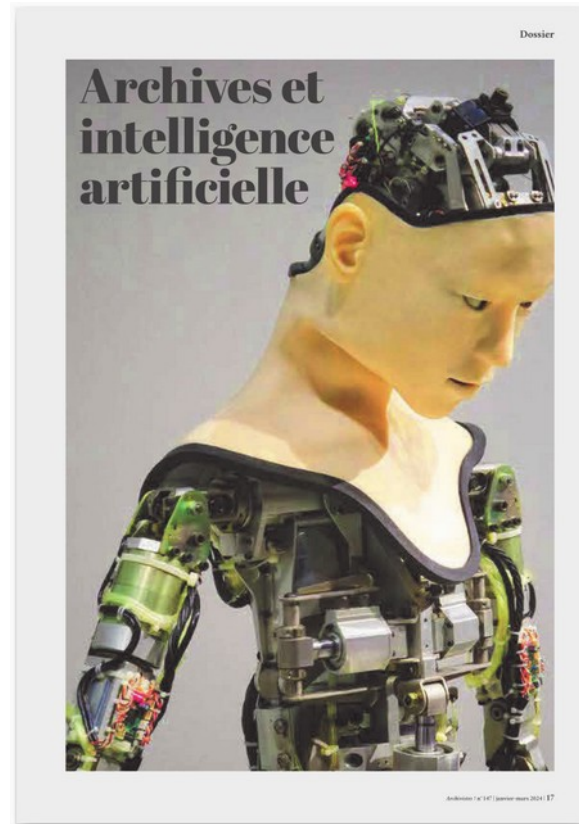
📦 *Data
vizualisation*

📦 Modèles de
langages (LLM)
et prompts



Les usages de l'IA dans les archives

En amont : une aide
à la collecte et à la
sélection ? Faire
face à la masse.



13/12/2024
Mediadix

Association des Archivistes français, *Archivistes !*, n° 147,
janv-mars 2024.

Les usages de l'IA dans les archives : collecte, sélection

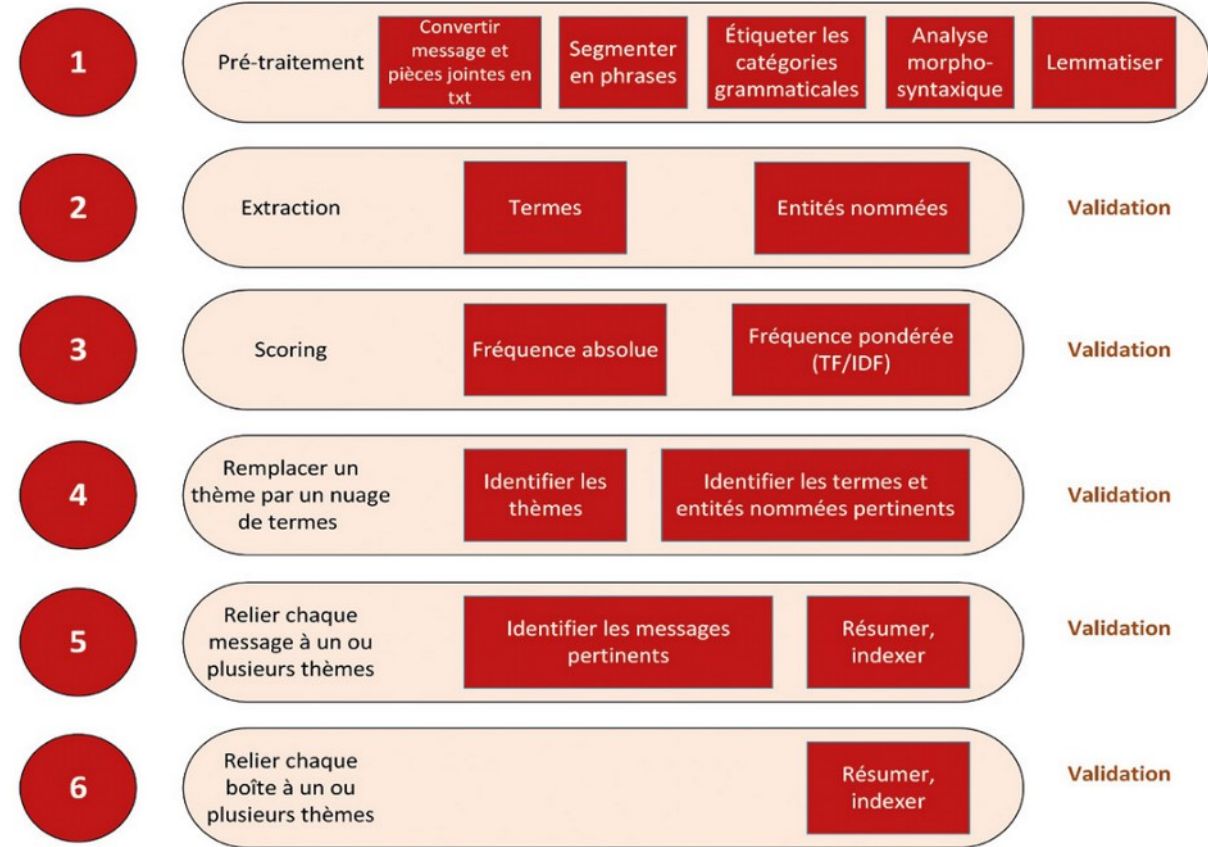
En amont : une aide à la collecte et à la sélection ? Faire face à la masse.


Pêlè-Mel : extraction et classification pour archivage des boîtes mail.

Suggestio : typage de milliers de fichiers bureautiques (par Action, Objet, Type).

Retour d'expérience sur le traitement des vracs bureautiques aux archives du Luxembourg par usage d'une IA générative.

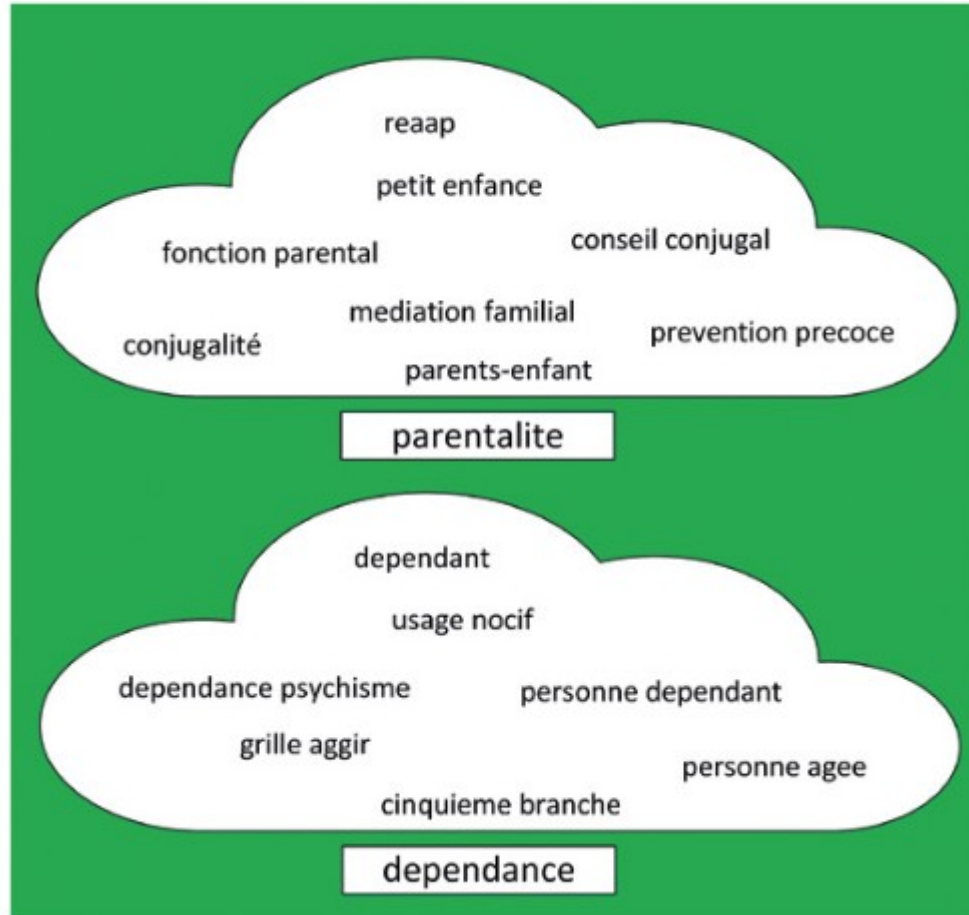
Exemple du projet Pêle-Mel




 Sélection de messageries à archiver en fonction de l'intérêt des mots-clés relevés

(B. Grailles, Touria Aït El Mekki, université d'Angers)

Exemple du projet Pêle-Mel



- Sélection de messageries à archiver en fonction de l'intérêt des mots-clés relevés

(B. Grailles, Touria Aït El Mekki, université d'Angers)

Voir aussi

13/12/2024
Mediadix

N° 147 AUTOMNE-HIVER 2024



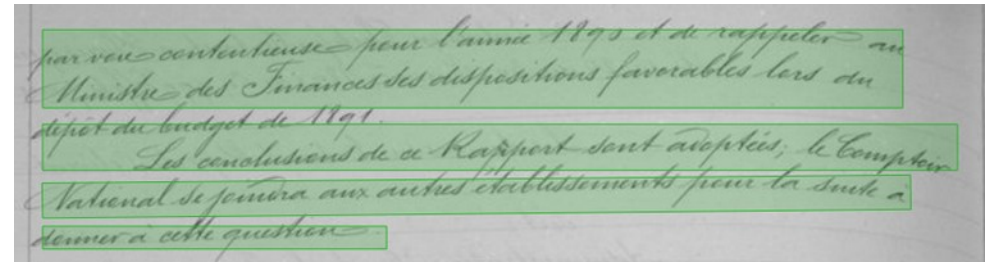
Recherche et intelligence artificielle

IA et valorisation du patrimoine archivistique

Au cœur de la source

Technologies d'IA exploitées

- ❖ Handwritten text recognition (HTR)
- ❖ Named Entities Recognition (NER)
- ❖ Modèles de langages (LLM)



Computation time on Intel Xeon 3rd Gen Scalable cpu: 0.038 s

Arrêt rendu sur la requête d' **Hugues Rollin** **PERSON** ministre de la religion
pretendue réformée a **TITLE** **Veyne** **LOCATION** en **Dauphiné** **LOCATION** ,
Bouteroue **PERSON** , **Murat** **PERSON** , **Blanc** **PERSON** , **Dize** **PERSON** et
Charles **PERSON** aussi **ministres de lad** **TITLE** . religion aud. pays de
Dauphiné **LOCATION** et **Ezechiel Benoît imprimeur** **PERSON** a Dye par lequel
le **Roy** **TITLE** évoque à soi et a son **Conseil les instances et procès mus et**
juelles **ORGANISATION** renvoie en la **Chambre de l'edit de**
Grenoble **ORGANISATION**

13/12/2024

Mediadix

Les usages de l'IA pour le traitement et la diffusion

Transcrire le texte...
Et après ?

**Diffusion et
moteur de recherche**

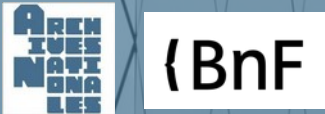
HIMANIS

- ❏ HISTorical MANuscripts
Indexing for user-
controlled Search
- ❏ Conception et conduite :
Dominique Stutzmann,
IRHT
- ❏ Registres de la chancellerie
royale française (XIII^e-
XV^e siècles)

*Joint
programming
initiatives*



Financement : 410 k€

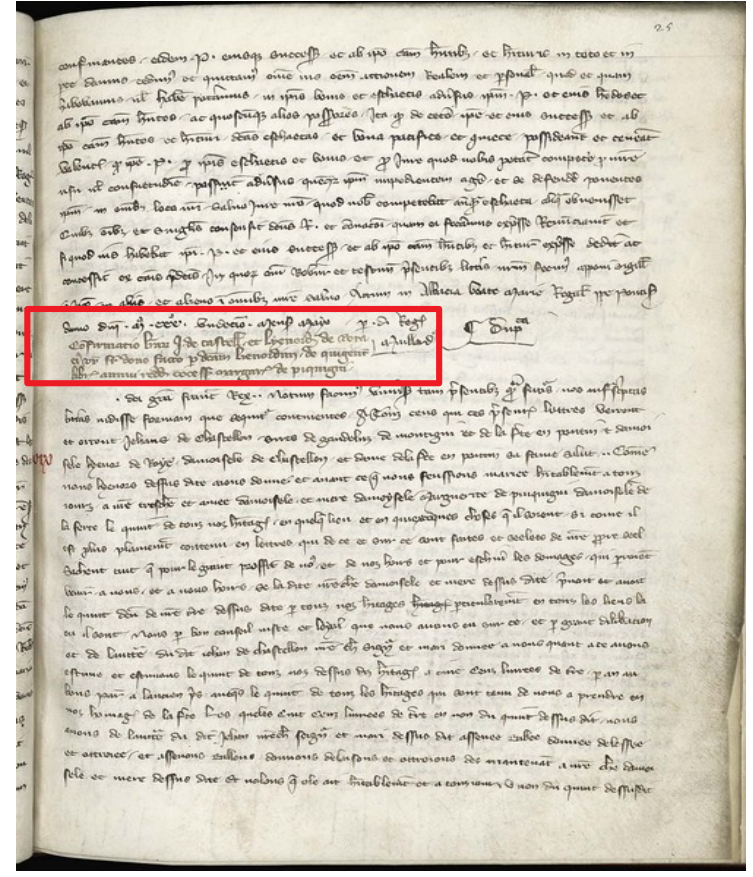
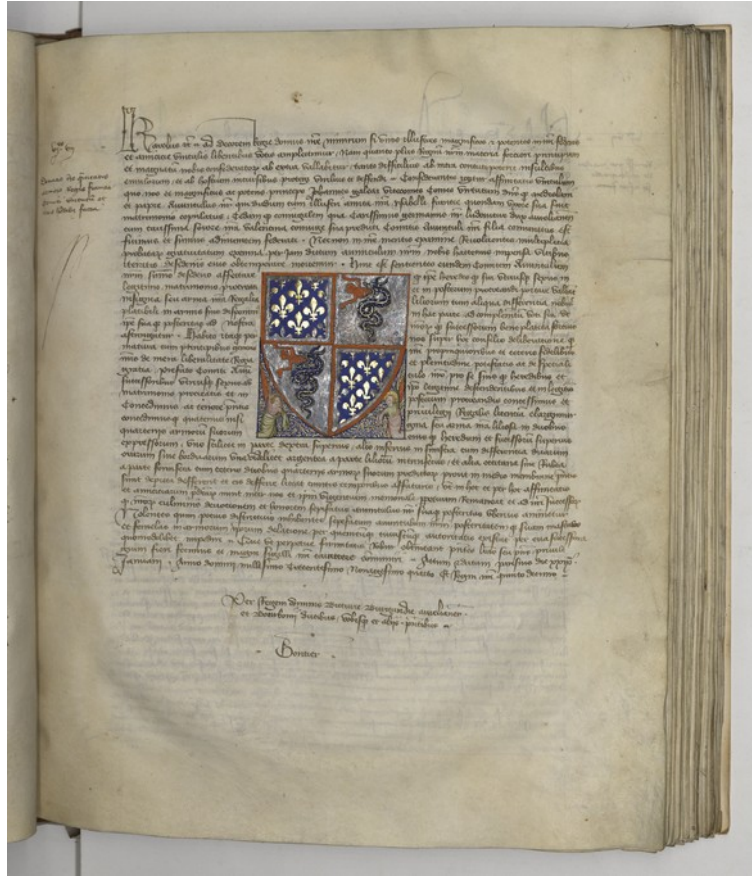


Les fonds concernés par HIMANIS

199 registres
du XIII^e au
XV^e siècle.

179 aux AN
(près de
80 000 pages).

20 à la BnF.



Registres de la chancellerie royale au Moyen Âge : un fonds primordial pour appréhender la construction du premier État moderne.

Interface de consultation

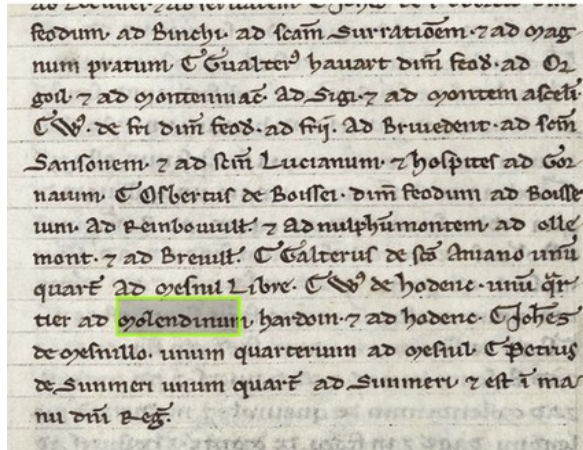
Diffusion : moteur de recherche en texte intégral

Exemple: recherche du terme "molendinum" (moulin)

me » chancery » JJ026 » page 80

dinum" with an average confidence of 69!

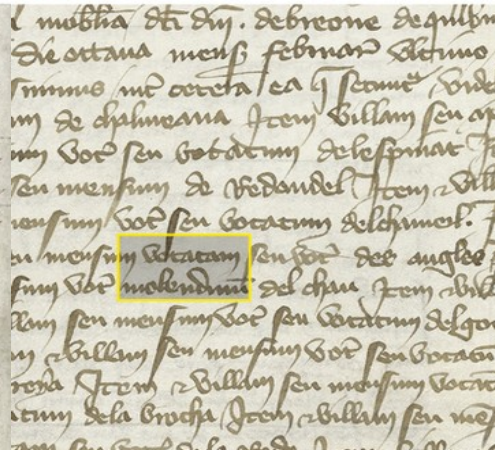
← Previous | Next →



me » chancery » JJ078 » page 357

dinum" with an average confidence of 46!

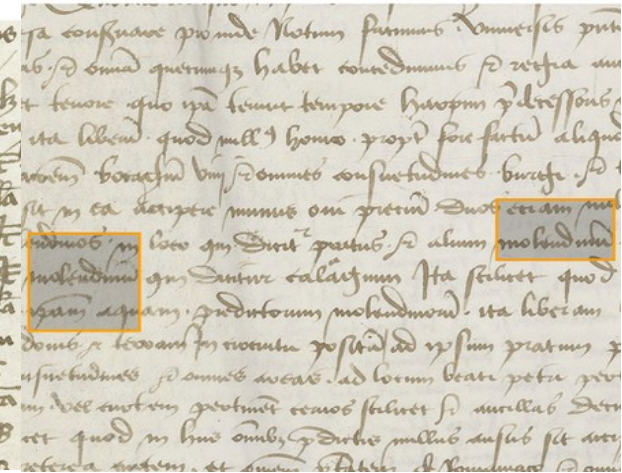
← Previous | Next →



me » chancery » JJ187 » page 159

dinum" with an average confidence of 31.2!

← Previous | Next →



13/12/2024

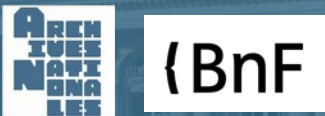
Mediadix

<http://himanis.huma-num.fr/himanis/>

eNDP. 2021-2024

Redécouverte des registres capitulaires de Notre-Dame de Paris

e-NDP



e-NDP : Notre Dame et son cloître

ANR 2021-2024

Présentation Partenaires Membres Objectifs Séminaires Ateliers HTR Formation Valorisation Publications

e-NDP

🕒 30/06/2022 📁 Non classé

Le projet ANR propose de renouveler la connaissance de Notre-Dame de Paris par l'édition collaborative numérique de 26 registres du chapitre de Notre-Dame (1326-1504) (conservés aux Archives nationales). Cette édition collaborative se fonde sur une procédure de reconnaissance automatique de l'écriture manuscrite (*handwriting text recognition* ou HTR).

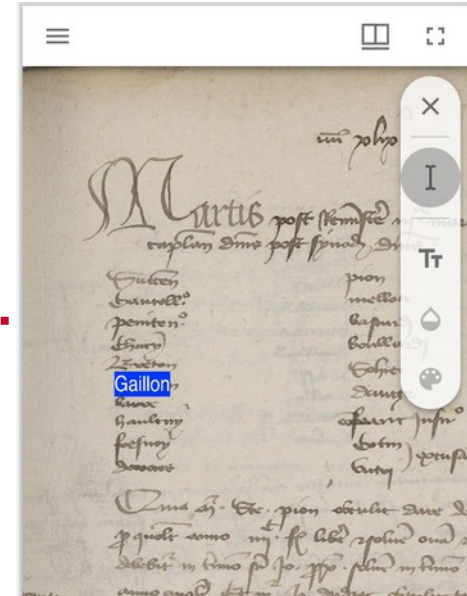
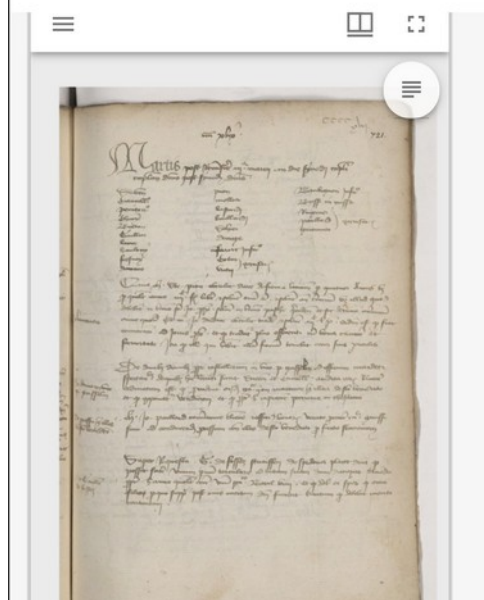
Le projet entend également reconstituer numériquement le parcellaire du cloître, en s'appuyant sur le Système d'information géographique ALPAGE.

13/12/2024

Mediadix



Fac-similé du registre de conclusions capitulaires LL116 de Notre-Dame de Paris - Mars 1450



13/12/2024
Mediadix

<https://endp.chartes.psl.eu/facsimile/LL116/716>

L. Terriel
(École des chartes)

Jean Aguenin le jeune

Chanoine

Identité

Prénom(s) - Nomen :

Johannes ; Jehan ; junior

Nom(s) - Cognomen :

Aguenin ; Aguenini ; Hagenin ; Hagenini

Mentions dans les registres

Première mention :

12 Septembre 1418

Dernière mention :

26 Août 1472

Relations familiales

neveu de :

Pierre Aguenin

fils de :

Jean Aguenin

frère de :

Guillaume Aguenin

frère de :

André Aguenin

Parcours

● Cliquer pour afficher l'évènement

12 Septembre 1418

26 Novembre 1423

23 Décembre 1440

29 Avril 1472

Sortie

Statut :

Chanoine de Paris (Canonicus Parisiensis)

Lieu : Maison claustrale (Domus claustralis)

Note : Motif de la sortie : Mortem Obiit magister J. Aguenin - Hodie hora quasi Va post meridiem magister Johannes Aguenin, canonicus Parisiensis, obiit in domo sua claustrali.

📄 Aller au fac-similé

Les usages de l'IA pour le traitement et la diffusion

Transcrire le texte...
Et après ?

Diffusion et
moteur de recherche

Créer de la métadonnée
à partir du document
d'archives



LECTAUREP (2018-2020)

Utiliser les ressources des notaires pour créer des
métadonnées descriptives

LECTure AUTomatique de REpertoires

Page	Word Error Rate (WER)	Error Rate
MC/RE/CXIII/20, image 100, page de droite	85.89	45.58
MC/RE/CXIII/22, image 65, page de droite	89.46*	52*
MC/RE/CXIII/19, image 77, page de gauche	84.69*	39.11*

* calculés avec wer++ [10]



<https://lectaurep.hypotheses.org/>



Registres des notaires parisiens mi XVIII^e-début du XX^e siècle.

Un inventaire en soi...

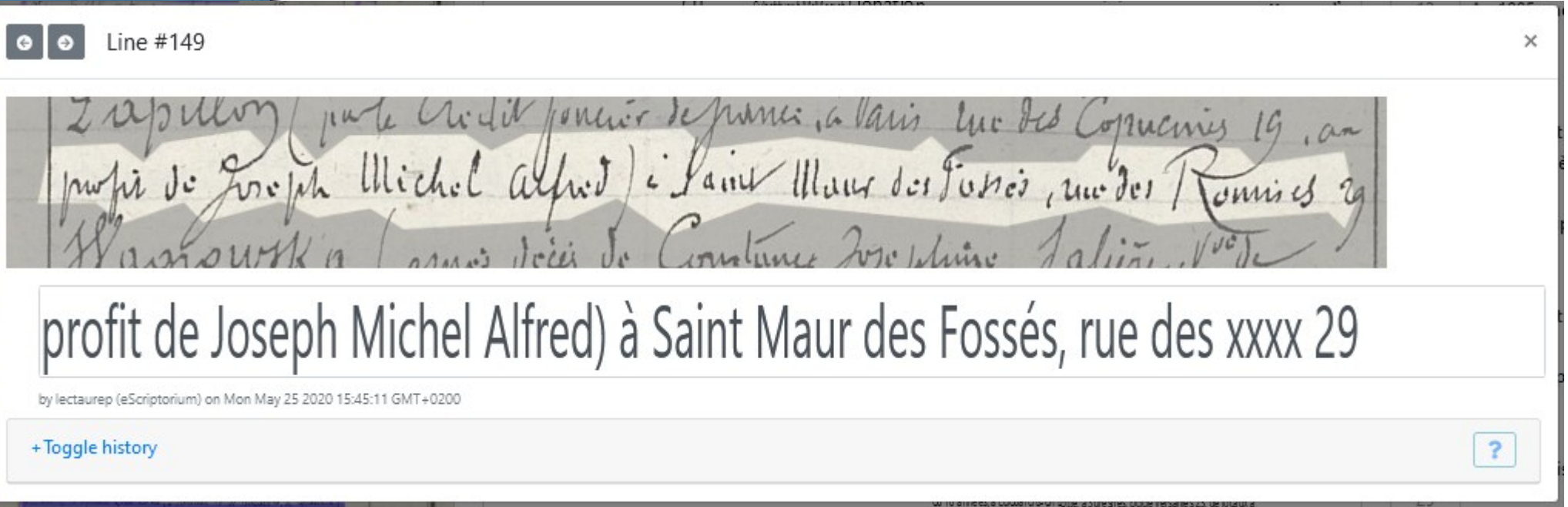


13/12/2024
Mediadix

DATES	NATURE ET ESPÈCE DES ACTES :		NOMS, PRÉNOMS ET DOMICILES DES PARTIES
	EN BRÈVETS	EN MINUTES	
			INDICATIONS, SITUATIONS ET PRIX DES BIENS
24	Acte	Inventaire	An 1903, mois de Juillet
24	Décharge		Contel (par Marie Marie) à Paris, rue Brochant 50, du 11 ^e Juin 1903
24	Procuration		Joufflet (par Pierre Joseph Marie) à Paris, rue Brochant 27, n. 11. Carrière de mandats, et de 2000 ^{fr} .
24		Donation	Redouze (par Marie Marie) à Paris, avenue Condorcet, 21, 1 ^e section 2 ^e .
24			Kouvetel (par Claude Pierre) à Paris, rue de la Colonne, 12, à Eugénie Hubert, son épouse (immobilière en toute propriété)
24			2 ^e (par M ^{lle}) à Paris (2 ^e)
25	Procuration		Letellier (par Marie Marie) à Paris, rue Blanche 101, 1 ^e section immuable
27		Mainlevée	Frisouls (par Marie Marie) à Paris, rue de la Colonne, 12, à Louis Olivier, à Clatques
27	(Paris 31 ^e juillet 1903)	Continuation d'inventaire	Marczewski (par Marie Marie) à Paris, rue de la Colonne, 12, à Marie Marie et Marie Marie
28	Consentement mariage		Jaslin (par Marie Marie) à Paris, rue de la Colonne, 12, à Marie Marie et Marie Marie
29		Mainlevée	Du Bois (par Marie Marie) à Paris, rue de la Colonne, 12, à Marie Marie et Marie Marie

Processus

Line #149



The screenshot shows a digital interface for transcribing handwritten text. At the top, there are navigation arrows and the text "Line #149". Below this is a horizontal strip of a handwritten document in cursive script. Underneath the strip is a white box containing the transcription of the visible text: "profit de Joseph Michel Alfred) à Saint Maur des Fossés, rue des xxxx 29". Below the transcription box, there is a small text attribution: "by lectaurep (eScriptorium) on Mon May 25 2020 15:45:11 GMT+0200". At the bottom left of the interface is a blue link "+ Toggle history", and at the bottom right is a blue question mark icon in a square box.

profit de Joseph Michel Alfred) à Saint Maur des Fossés, rue des xxxx 29

by lectaurep (eScriptorium) on Mon May 25 2020 15:45:11 GMT+0200

+ Toggle history

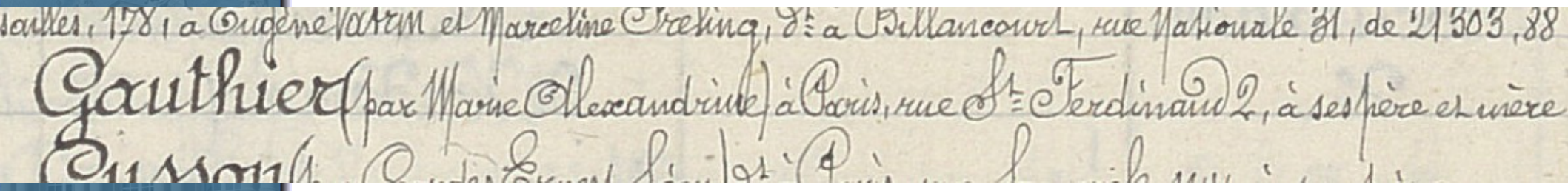
HTR avec Kraken et son interface eScriptorium (préférée à Transkribus pour son caractère ouvert).

13/12/2024

Mediadix

<https://lectaurep.hypotheses.org/>

Processus



Gauthier PER (par Marie Alexandrine PER) à Paris LOC , rue St Ferdinand 2 LOC , à ses père et mère

Débuts de l'utilisation de NER pour extraire des termes d'indexation.

**ARCHIVES
NATIONALES**

Pierrefitte-sur-Seine –
Paris



**MINISTÈRE
DE LA CULTURE**

*Liberté
Égalité
Fraternité*

NER4Archives

Faciliter la reconnaissance d'entités nommées dans les inventaires

Hosted inference API ⓘ

Token Classification

Examples ▾

Confirmation, par le roi Louis IX, du versement d'une rente annuelle de 20 livres parisis sur le Temple à Paris, au bénéfice de Philippe, fils de son défunt chambellan Philippe de Nemours, et de ses héritiers ; cette somme provient du partage d'une rente de 40 livres acquise par le chambellan à la suite d'un échange avec Guillaume, seigneur de Lignièrès, et dont l'autre moitié a été échangée, après accord du roi, avec l'abbaye de Saint-Port (Barbeau) par un autre fils du chambellan, Jean

Compute



NER4Archives

Computation time on Intel Xeon 3rd Gen Scalable cpu: cached

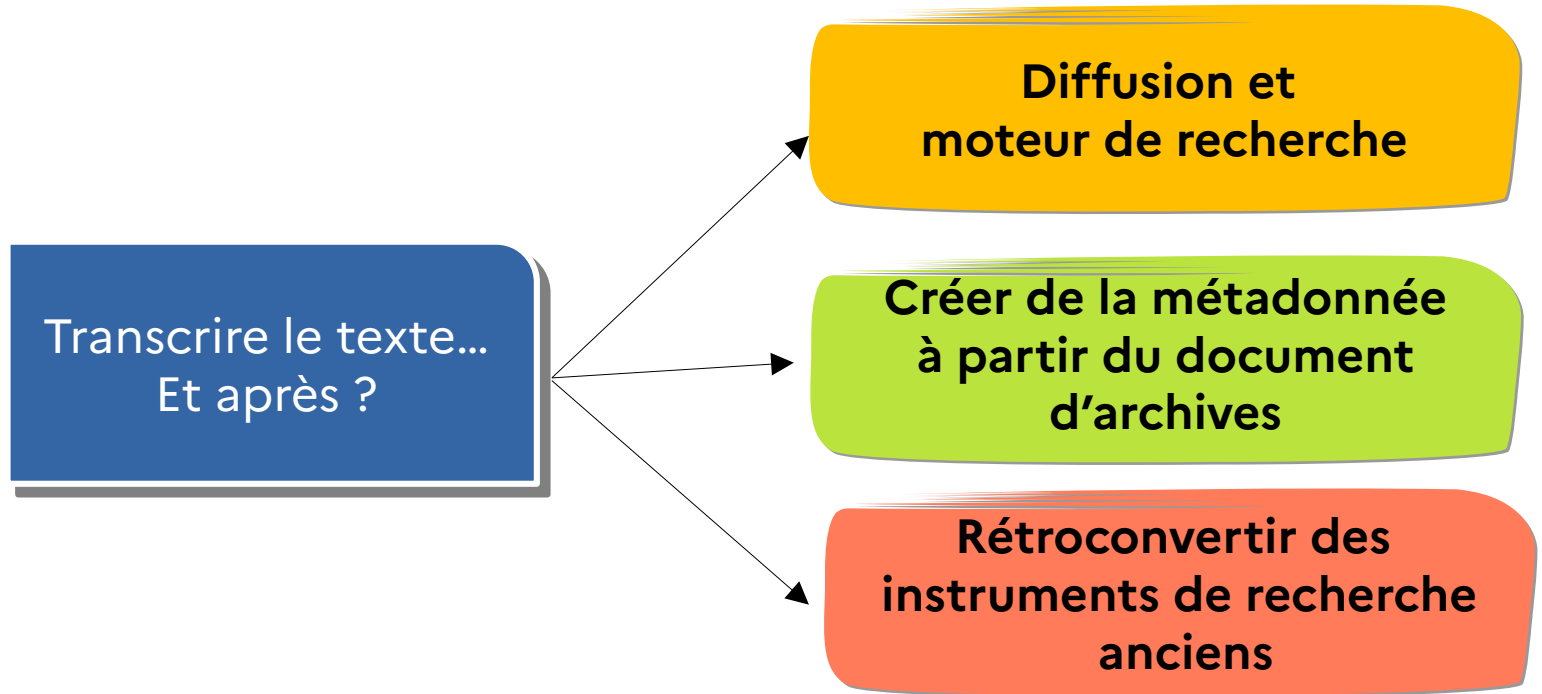
Confirmation, par le **roi** **TITLE** **Louis IX** **PERSON**, du versement d'une rente annuelle de 20 livres parisis sur le Temple à **Paris** **LOCATION**, au bénéfice de **Philippe** **PERSON**, fils de son défunt **chambellan** **TITLE** **Philippe de Nemours** **PERSON**, et de ses héritiers ; cette somme provient du partage d'une rente de 40 livres acquise par le **chambellan** **TITLE** à la suite d'un échange avec **Guillaume** **PERSON**, **seigneur de Lignièrès** **TITLE**, et dont l'autre moitié a été échangée, après accord du **roi** **TITLE**, avec l' **abbaye de Saint-Port** **ORGANISATION** (Barbeau) par un autre fils du **chambellan** **TITLE**, **Jean** **PERSON**

</> JSON Output

Maximize

Identifier grâce aux modèles de langage des entités nommées dans des inventaires non-indexés.

Les usages de l'IA pour le traitement et la diffusion



**ARCHIVES
NATIONALES**

Paris – Pierrefitte-sur-
Seine



**MINISTÈRE
DE LA CULTURE**

*Liberté
Égalité
Fraternité*

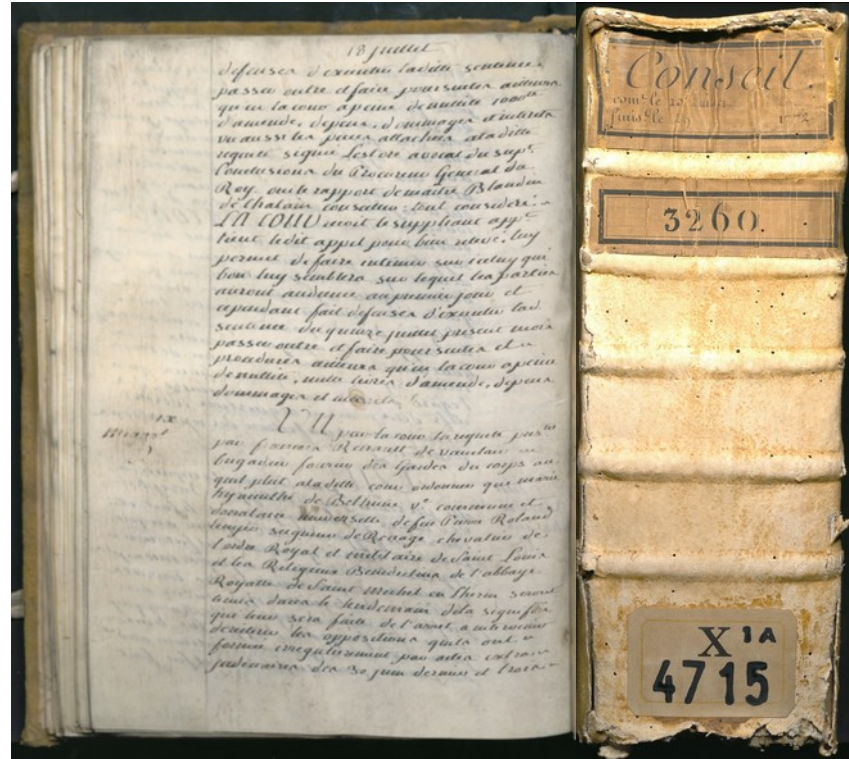
SIMARA

Saisie d'Inventaires Manuscrits Assistée par
Reconnaissance Automatique

Les inventaires, ou la clé du Trésor...

RENAGE (Marie Hyacinthe de
 Bellune, veuve de Pierre
 Rodou (mariage de), capitaine
 Chevalier de l'ordre royal et militaire
 de Saint-Louis.
 / François Renaud de Vaublair
 18 Juillet 1772 X^{1A} 4715. 95

Une fiche...



Pour trouver un acte...



Parmi ces registres !

Particularités des IR anciens

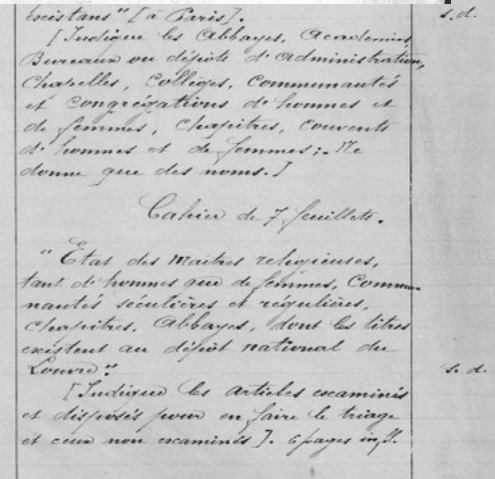
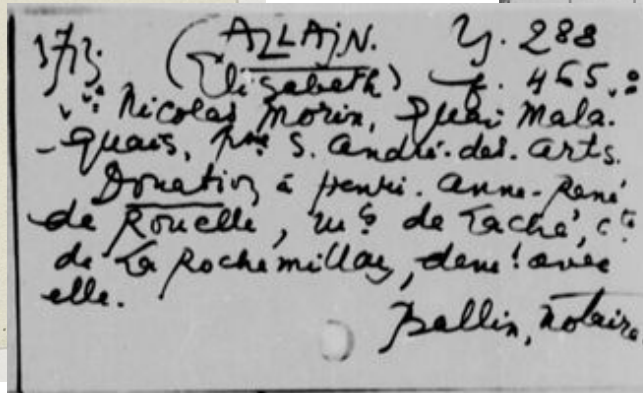
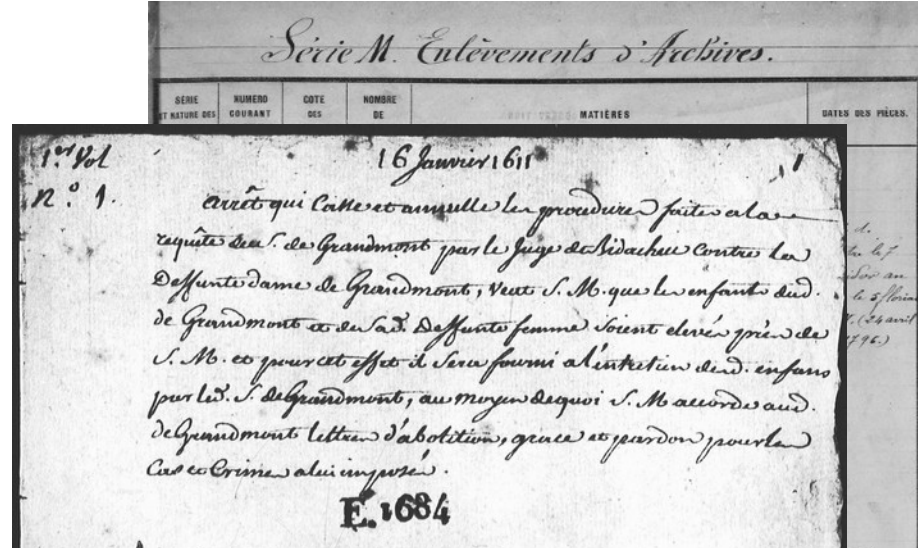
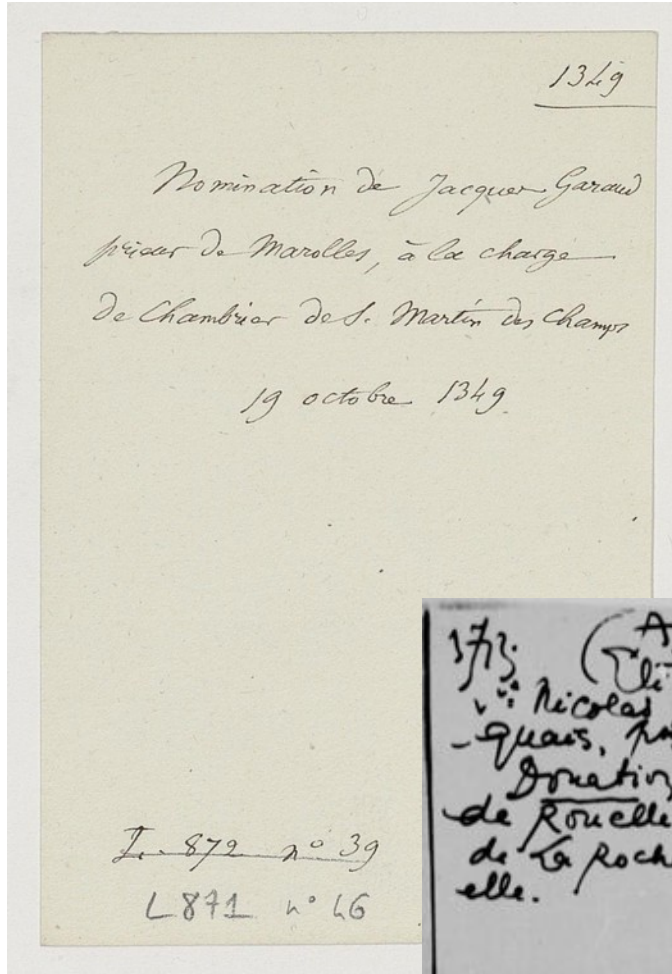
Volumétrie :

800 000 fiches,
10 000 pages.

Processus de

traitement manuel
long:

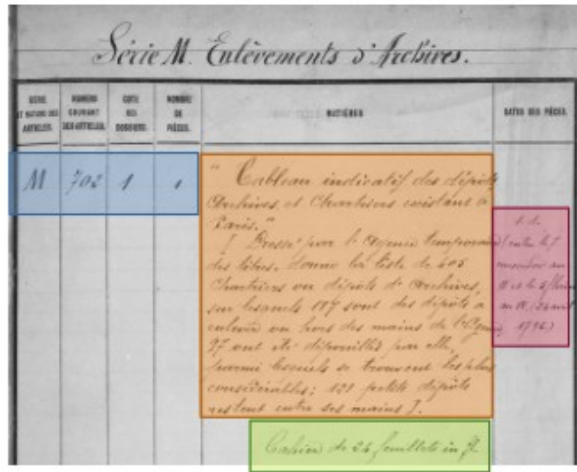
300 pages = 2 mois
1550 fiches = 1 mois



Objectifs de SIMARA

Transcrire le text avec l'HTR

et structurer en même temps le texte en XML EAD, selon les catégories d'information.



Cote	M 702 / 1 / 1
Intitulé	« Tableau indicatif des dépôts d'archives...
Date	S. d. (entre le 7 mess...
Support	Cahier de 26 feuillets...

```
<unitid>M 702/1/1</unitid>

<unittitle>Tableau
indicatif des dépôts
d'archives...</unittit
le>

<unitdate>S. d. (entre
le
7 mess...</unitdate>

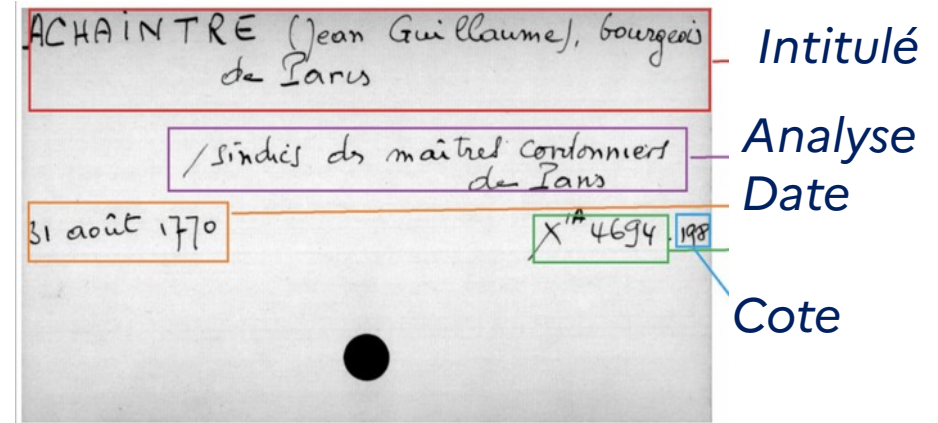
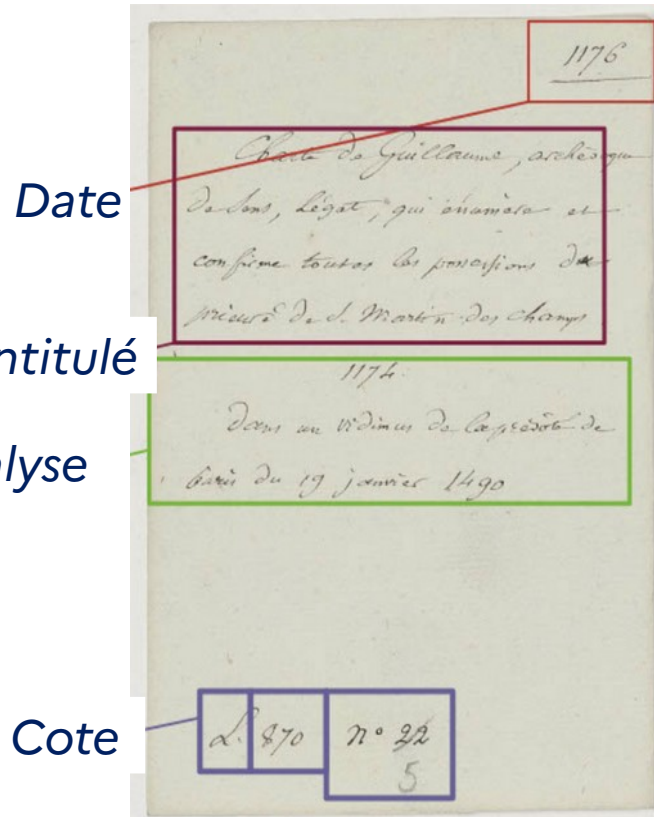
<physdesc>Cahier de 26
feuillets...</physdesc
>
```

1. Transcription et
identification des
catégories.

2. Répartition des
transcriptions dans un
formulaire.

3. Derrière le
formulaire, éléments
et attributs XML.

Analyse de l'inventaire



Saisie éventuelle d'une vérité de terrain

- Traduction des métadonnées en formulaire.
- Formulaire pour saisie de vérité de terrain.

Annotation d'une image entière

Nom du projet
FRAN_IR_5848_Série Y - Registres des insinuations

Nom de l'image
Page 605 (FRAN_IR_005848_6804.pdf)

Formulaire
Série Y - formulaire insinuations

[Retour](#)
[Ignorer cette tâche](#)

Intitulé *

Date

Date de début

Date exacte ou borne inférieure pour un intervalle de dates. Cette valeur est requise pour être conforme à la norme ISO 8601.

Date de fin

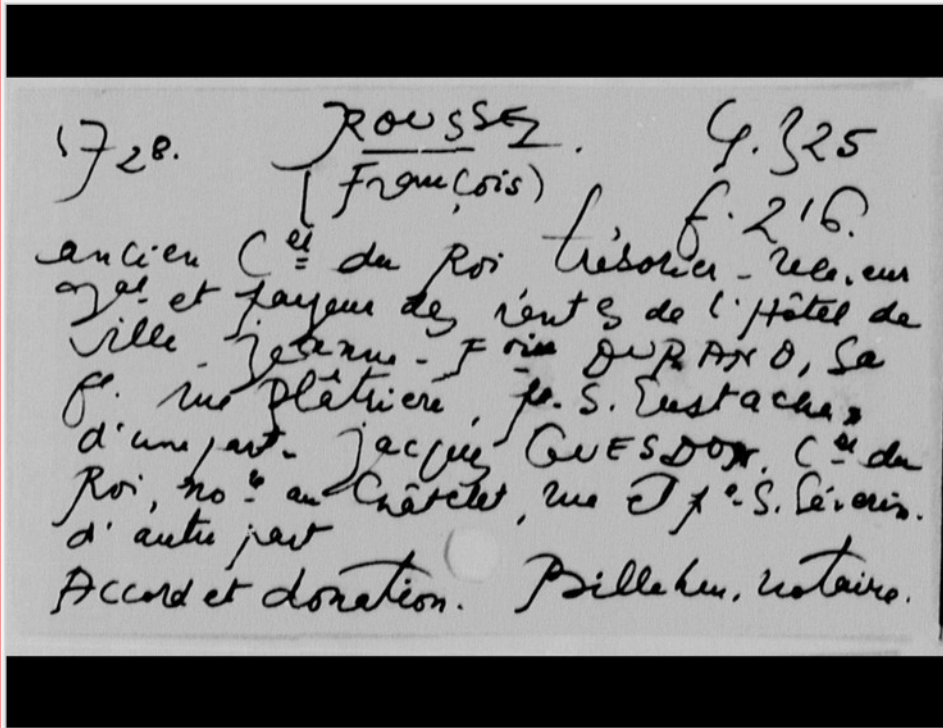
Borne supérieure optionnelle pour un intervalle de dates. Ce champ peut être laissé vide en cas de date exacte.

Cote : Série

Cote : Article

Cote : Pièce/Folio

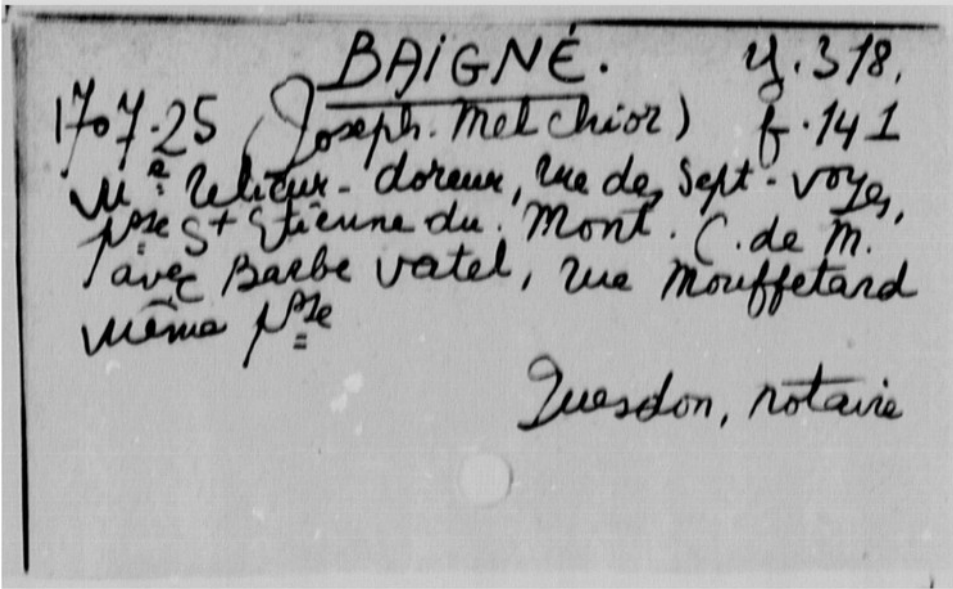
Analyse complémentaire



1728. ROUSSEZ. G. 325
 (François) f. 216.
 ancien C^{te} du Roi trésorier - receveur
 et payeur des rentes de l'Hôtel de
 ville - Jean - François DUPRAXO, Sa
 fr. rue Plâtrière, fr. S. Eustache
 d'une part. Jacques GUESDON, C^{te} du
 Roi, no^{te} au Châtelet, rue St^e S. Séverin.
 d'autre part
 Accord et donation. Pillehen, notaire.

Processus d'HTR et NER (Teklia)

- Le texte est généré et traité par une combinaison HTR+NER.
- Le texte est redistribué dans les champs du formulaire.



TRANSCRIPTIONS

Manual

Manual

INTITULE BAIGNÉ (Jospeh-Melchior) DATE 1707-1725 COTE_SERIE Y
COTE_ARTICLE 318 PRECISIONS_SUR_COTE 141 ANALYSE_COMPL Maître
relieur doreur, rue des Sept-Voyes, paroisse Saint-Étienne-du-
Mont.
Contrat de mariage avec Barbe Vatel, rue Mouffetard, même
paroisse.
Guesdon, notaire.

Correction des transcriptions HTR et export

Correction/
validation des
transcriptions
de l'IA.

Date

22 novembre 1681

Date de début

22 / 11 / 1681

Date exacte ou borne inférieure pour un intervalle de dates. Cette valeur est requise pour être conforme à la norme ISO 8601.

Date de fin

jj / mm / aaaa

Borne supérieure optionnelle pour un intervalle de dates. Ce champ peut être laissé vide en cas de date exacte.

Analyse complémentaire

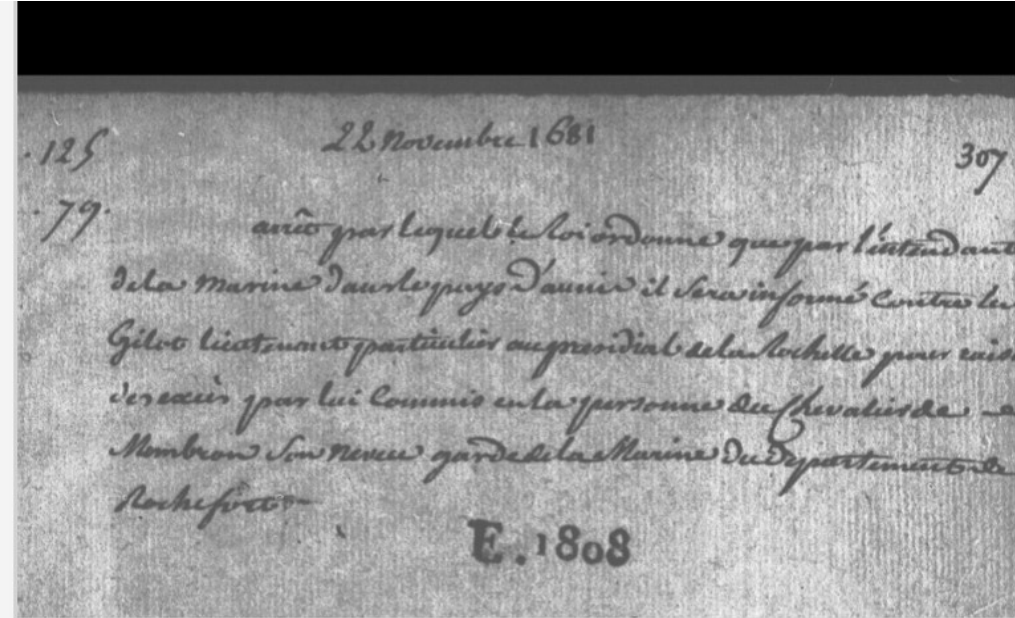
307

Intitulé *

arrêt par lequel à Roi rdonne que par l'intendant de la marine dans le pays d'aunis il sera informé contre le S. Gilot licutenant particulier au presidial de la Rochelle pour caison des excis par lui commis en la personne du chevalier de Membron son nevece garde de la Marine du departement le Rochefort on

Cote : Série *

E



Transcription
à relire.

Correction des transcriptions HTR et export

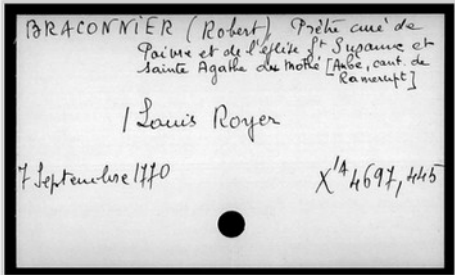
Données exportables en EAD et publiables dans le SI.

Tableau de bord Formulaires Assigner des tâches Jean-François Moufflet - Admin

Réponses sur Page 0007 (FRAN_IR_001428_08.pdf)

[Retour](#)

1. Tâche effectuée **TERMINÉE** [Détails de la tâche](#) [Télécharger au format XML EAD](#)



Annotation n°1 [Éditer](#)

Champ	Valeur
Cote : Série	X1A
Cote : Article	4697
Date	7 septembre 1770
Intitulé	BRACCONNIER (Robert) Prêtre curé de Poivres et de l'église Sainte-Suzanne et sainte Agathe du Mothé [Aube, canton de Ramerupt]
Analyse complémentaire	Contre Louis Royer
Cote : Pièce/Folio	445
Entité	Catégorie

administrateur

XML

VALIDE

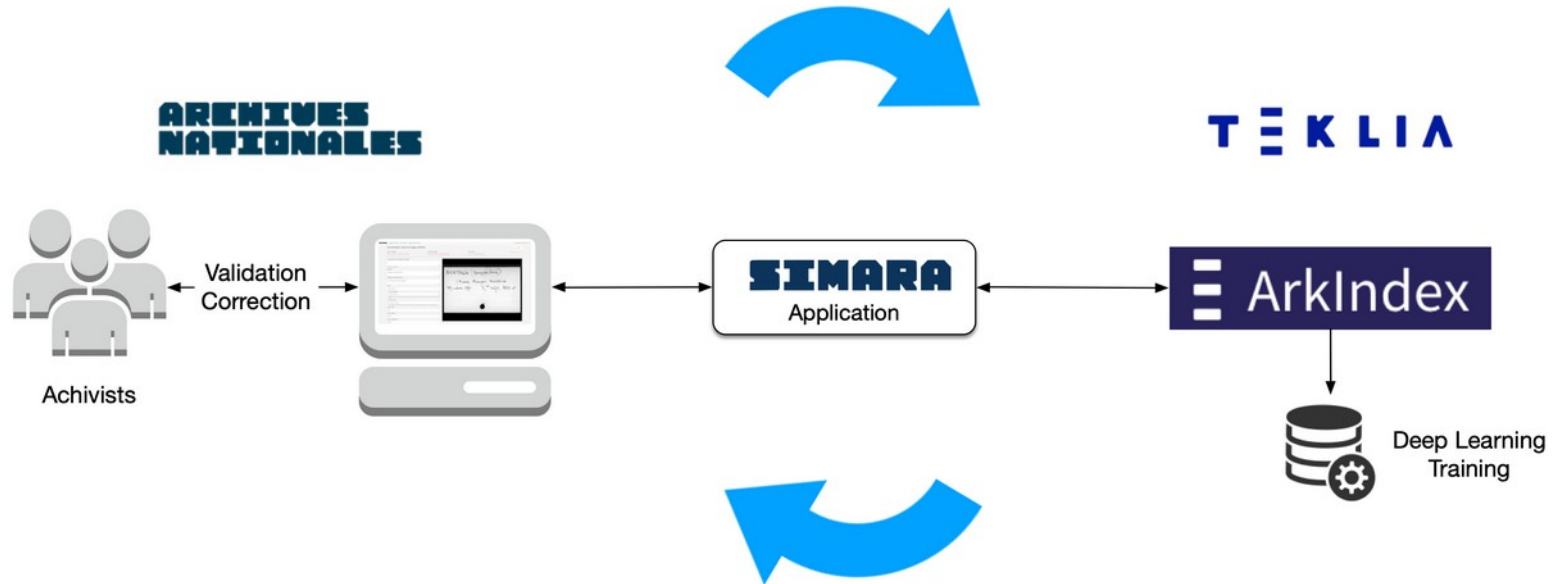
```

<accessrestrict>
  <p/>
</accessrestrict>
<dsc>
  <c>
    <did>
      <unitid type="cote-de-consultation">X1A//4697</unitid>
      <unitid type="pieces">445</unitid>
      <unittitle>BRACCONNIER (Robert) Prêtre curé de Poivres et de l'église Sainte-Suzanne et sainte Agathe du Mothé [Aube, canton de Ramerupt]</unittitle>
      <unitdate normal="1770-09-07" calendar="gregorian" era="ce">7 septembre 1770</unitdate>
    </did>
    <scopecontent>
      <p>Contre Louis Royer</p>
    </scopecontent>
  </c>
</dsc>
</archdesc>
</ead>
          
```

Prévisualisation du code XML EAD

Principales conclusions sur SIMARA

- Un cercle vertueux: les corrections apportées par le utilisateurs dans l'interface sont réutilisées pour améliorer les modèles.





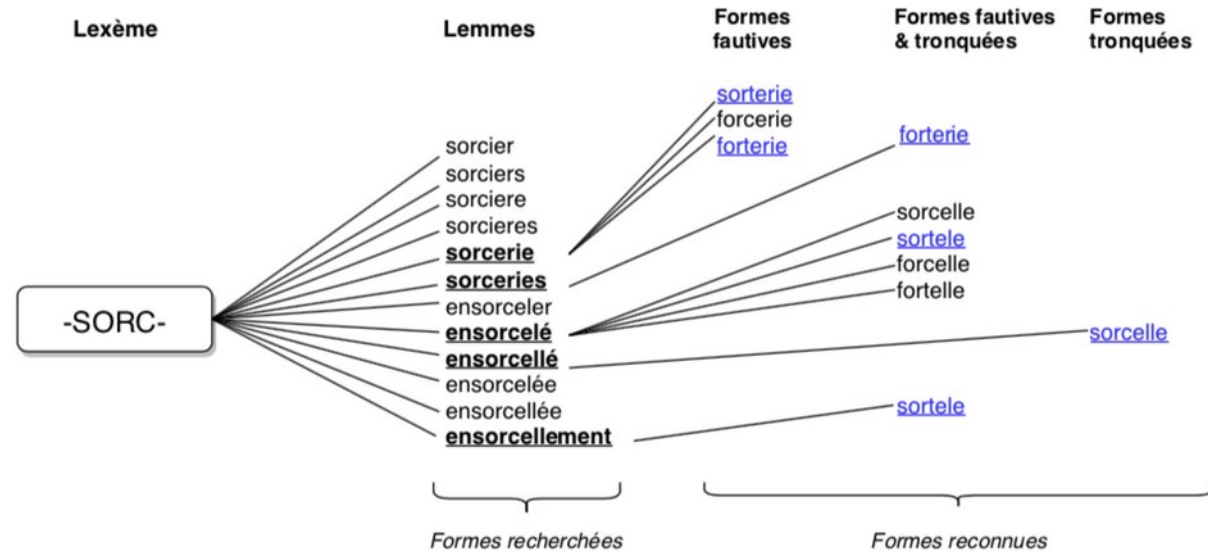
Quelques considérations

Sur la collecte et le traitement

- ❏ Question de la confidentialité des données
- ❏ L'IA **ne se substitue pas** à l'archiviste pour décider ce qu'il faut conserver

Pêlé-Mel : « Si l'IA ouvre des horizons, elle ne fait que proposer des outils de compréhension et de visualisation nouveaux : l'impulsion comme la décision restent du côté du service d'archives. »

Les archives replacées au cœur de la recherche



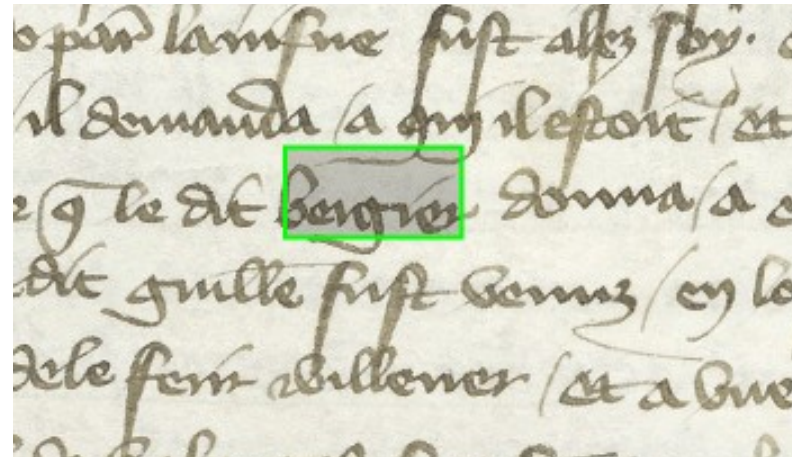
Himanis : analyse statistique lexicale autour de la sorcellerie.

La transcription ne remplace pas la médiation de l'archiviste et les compétences paléographiques

La recherche en plein texte :

- Ne remplace pas les index que pouvaient concevoir les archivistes
 - ▶ Identifications des noms de lieu et de personne, mais aussi des concepts...
 - ▶ Désambiguïisation
 - ▶ Connaissance de la langue ancienne pour chercher le bon terme ou concept...

BERGER : 2783, 6665, 7260, 7281. V.
CABANE.



Un gain de temps pour les tâches répétitives

Date

22 novembre 1681

Date de début

22 / 11 / 1681

Date exacte ou borne inférieure pour un intervalle de dates. Cette valeur est requise pour être conforme à la norme ISO 8601.

Date de fin

jj / mm / aaaa

Borne supérieure optionnelle pour un intervalle de dates. Ce champ peut être laissé vide en cas de date exacte.

Analyse complémentaire

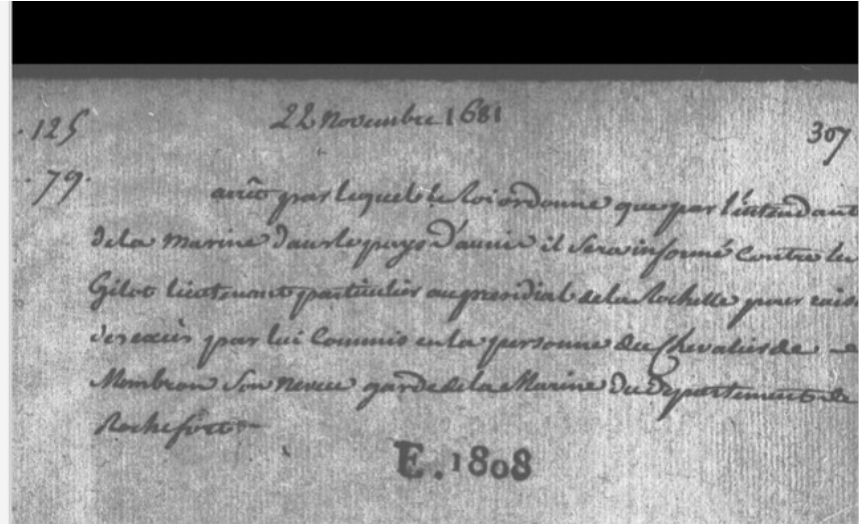
307

Intitulé *

arrêt par lequel à Roi rdonne que par l'intendant de la marine dans le pays d'aunis il sera informé contre le S. Gilot licutenant particulier au presidial de la Rochelle pour caison des excis par lui commis en la personne du chevalier de Membron son nevece garde de la Marine du departement le Rochefort on

Cote : Série *

E




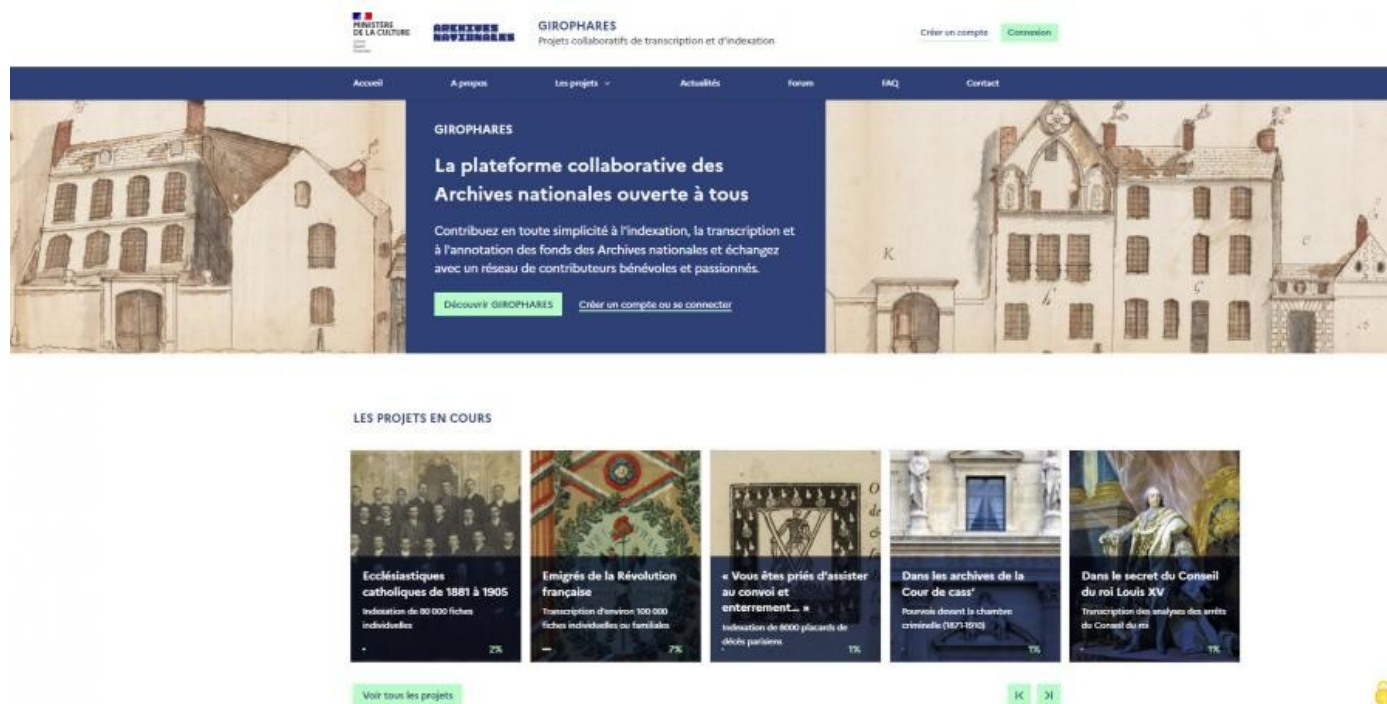
Mais qui ne supprime pas le temps de relecture et de vérification de la qualité des données !

13/12/2024

Mediadix

Le facteur humain

- 
SIMARA : rapidité de transcription, mais temps de relecture/correction très long.
 - Repose sur la motivation des responsables scientifiques
 - Faire appel aux bénévoles pour relire les transcriptions
 - ▶ cf. GIROPHARES, plate-forme de transcription collaborative des AN... mais sans IA !



The screenshot shows the homepage of the GIROPHARES platform. At the top, there is a navigation bar with links for 'Accueil', 'A propos', 'Les projets', 'Actualités', 'Forum', 'FAQ', and 'Contact'. The main header features the title 'GIROPHARES' and the subtitle 'La plateforme collaborative des Archives nationales ouverte à tous'. Below this, a central text block describes the platform's purpose: 'Contribuez en toute simplicité à l'indexation, la transcription et à l'annotation des fonds des Archives nationales et échangez avec un réseau de contributeurs bénévoles et passionnés.' There are two buttons: 'Découvrir GIROPHARES' and 'Créer un compte ou se connecter'. The page also displays a section titled 'LES PROJETS EN COURS' with five project cards, each showing a historical image and project details:

Project Name	Description	Progress
Ecclesiastiques catholiques de 1881 à 1905	Indexation de 80 000 fiches individuelles	25%
Emigrés de la Révolution française	Transcription d'environ 300 000 fiches individuelles ou familiales	75%
« Vous êtes priés d'assister au convoi et enterrement... »	Indexation de 8000 placards de décès parisiens	1%
Dans les archives de la Cour de cass'	Panonceau devant la chambre criminelle (187-1870)	1%
Dans le secret du Conseil du roi Louis XIV	Transcription des analyses des arrêtés du Conseil du roi	1%

At the bottom of the project cards, there are buttons for 'Voir tous les projets' and navigation arrows.

Dépendances technologiques

- ❏ Une difficulté à diffuser par nous-mêmes les données
 - ▶ Implémentation de IIF encore faible dans les archives
- ❏ L'expertise en IA n'est pas internalisée ; besoin de construire des partenariats (structures de recherche ou entreprises privées).

Merci pour votre attention !

jean-francois.moufflet@culture.gouv.fr

13/12/2024

Mediadix