ARCHIVES NATIONALES
Paris – Pierrefitte-sur-Seine

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité



L'intelligence artificielle appliquée aux archives

03/11/2026

Webinaire AIAF

Jean-François Moufflet
Archives nationales (France)

ARCHIVES NATIONALES
Pierrefitte-sur-Seine – Paris

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

Quelques définitions.
Historique du développement de l'IA.

La première occurrence

- 1956 : première occurrence du terme

  Conférence à l'université de Dartmouth (New Hampshire)

  John McCarthy et Marvin Minsky (MIT)

- John McCarthy (1927-2011)

  « toute activité intellectuelle peut être décrite avec suffisamment de précision pour être **simulée** par une machine »

- Marvin Minsky (1926-2016) :

  « La construction de programmes informatiques capables d'accomplir des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains, car elles demandent des **processus mentaux de haut niveau** tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique. »

Quelques définitions

- « L'IA désigne la possibilité pour une machine de reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité. »

  Parlement européen

- « Ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine. »

  Encyclopédie Larousse, reprise par Wikipédia

Quelques définitions

« Capacité d'une machine à exécuter des tâches associées à l'intelligence humaine, comme la planification, la compréhension du langage, la reconnaissance d'objets ou de sons, l'apprentissage et la résolution de problèmes. Le test de Turing en est une illustration emblématique. »

Centre de Recherche Informatique de Montréal (CRIM)

Quelques définitions

- « Toutes ces technologies ont pour but de nous *assister* dans des tâches ponctuelles, souvent *répétitives* et fortement *codifiées*. Elles nous fournissent une aide qui vient amplifier notre humanité, et augmenter nos capacités intellectuelles, mais elles ne peuvent en aucun cas nous remplacer. »

Luc Julia (inventeur commande vocale d'Apple Siri), 2019

HISTOIRE DE L'IA

Temps forts de 1940 à aujourd'hui

1950 - 1970

Naissance et développement de l'IA



1943

Premier réseau de neurones artificiel par Warren McCulloch et Walter Pitts

1950

Travaux de John Von Neumann et Alan Turing, création du Test de Turing

1956

Terme "intelligence artificielle" introduit à la conférence de Dartmouth.

1980 - 1990

"Hiver de l'IA" durant les années 1980 et début des années 1990



2010

Essor de l'IA nouvelle génération



1996 - 1997

IBM Deep Blue bat Garry Kasparov aux échecs

2011

IA Watson d'IBM remporte Jeopardy!

2012

Google X crée une IA de reconnaissance d'image



2016

Alpha Go de Google bat le champion du monde de go

2020

Démocratisation de l'IA

Lancement de GPT-3 par OpenAI



2021

Introduction de Codex et DALL-E par OpenAI





2023

Lancement de GPT-4, intégration de DALL-E dans ChatGPT Plus

Ouverture au public de Google Bard et de Bing Chat



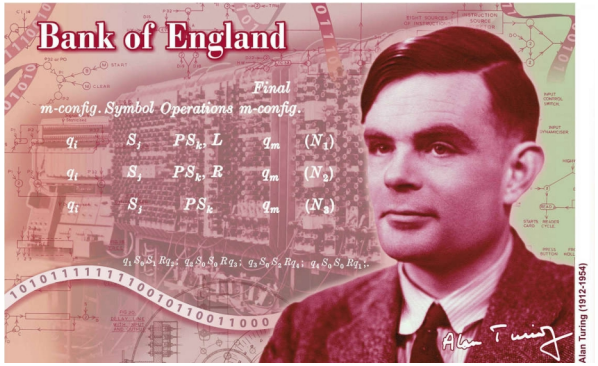
Source : WiziShop

Développement de l'IA

1950 : test de Turing

Une machine peut-elle avoir un dialogue avec un humain ?

Ce test, toujours d'actualité, pousse les machines à résoudre des problèmes toujours plus complexes soumis par l'homme.



Alan Turing (1912-1954)

Développement de l'IA

- Années 90
  Deep Blue
    Un ordinateur capable de battre un humain



Développement de l'IA

- Années 2010
  - massification des données (« Big Data »)
  - capacités **d'apprentissage** (le programme n'est plus dans l'itératif pur)
  - compréhension du langage humain et interactions



2011 : Waston (IBM)-Jeopardy !



Années 2010 : « assistants personnels » – langage naturel

Développement de l'IA

- Années 2010
  - 2012 : classification des images.
  - 2012 : Google X. Identification de chats sur des vidéos YouTube.

Using large-scale brain simulations for machine learning and A.I.

Jun 26, 2012 · 3 min read

Jeff Dean
Google Senior Fellow and SVP, Google AI

Andrew Ng
Visiting Faculty

Share

Listen to article 4 minutes

You probably use machine learning technology dozens of times a day without knowing it—it's a way of training computers on real-world data, and it enables high-quality speech recognition, practical computer vision, email spam blocking and even self-driving cars. But it's far from perfect—you've probably chuckled at poorly transcribed text, a bad translation or a misidentified image. We believe machine learning could be far more accurate, and that smarter computers could make everyday tasks much easier. So our research team has been working on some new approaches to large-scale machine learning.





One of the neurons in the artificial neural network, trained from still frames from unlabeled YouTube videos, learned to detect cats.

Développement de l'IA

- Années 2010
  - 2016 : Alpha Go (Google). Résolution de cas toujours plus complexes. Entraînement de l'IA contre elle-même.
  - 2015-2020 : essor des techniques d'entraînement des IA (*machine learning*, *deep learning*)



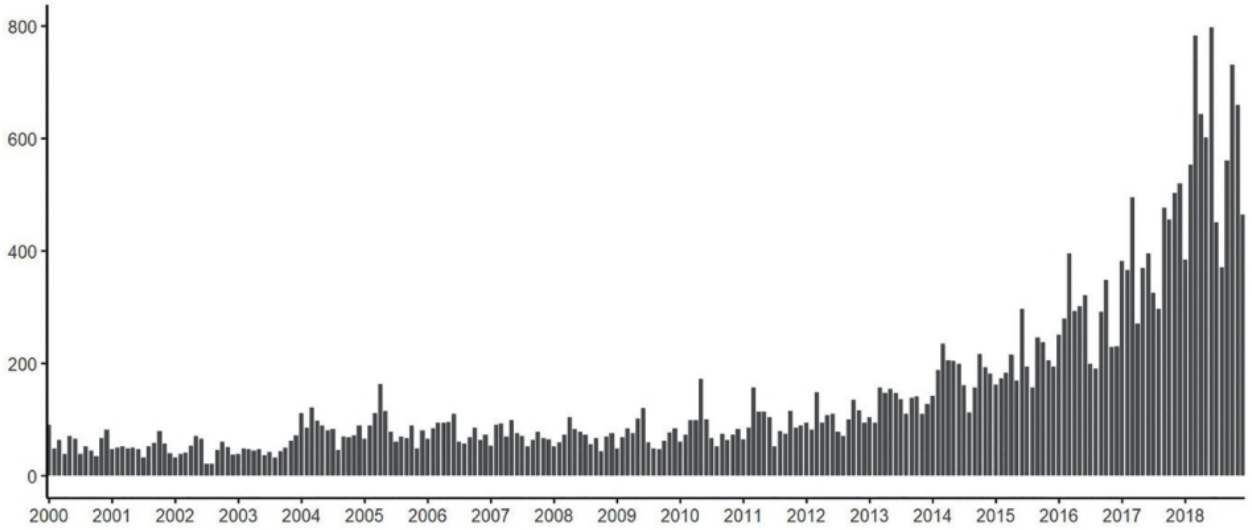
2016 : Lee Sedol vs Alpha Go

Développement de l'IA

- Médiatisation
- *« Emballement »*

en 2015

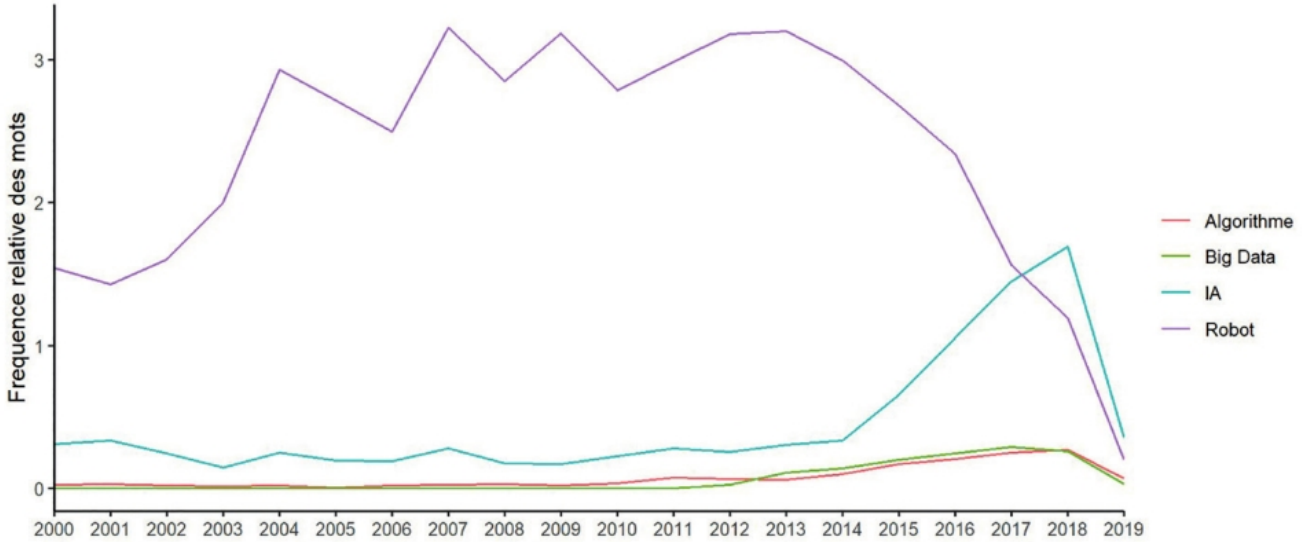
Figure 1. Évolution du nombre de publications mensuelles sur l'intelligence artificielle (2000-2019)



Source : auteures

Développement de l'IA

Figure 2. Fréquence relative des mots dans le corpus de presse (2000-2019) : des robots à l'IA



Source : auteures

Développement de l'IA

- 2020-2026 : essor de l'**IA générative**
  - Un pas de plus dans la « démocratisation » de l'IA : accès sur Internet de ChatGPT.
  - Juin 2020 : GPT-3 (OpenAI), modèle de traitement du langage naturel.
  - 2022 : API pour utilisation sur Internet
  - Mars 2023 : ChatGPT-4 : interface permettant de générer des contenus toujours plus variés.
  - Août 2025 : ChatGPT-5.
  - Sept. 2026 : ChatGPT-6



- Fév. 2023 : Llama (Meta)
- Mars 2023 : Claude (Anthropic)
- Avril 2023 : Mistral AI
- Sept. 2023 : Qwen (Alibaba)
- Déc. 2023 : Gemini (Google)

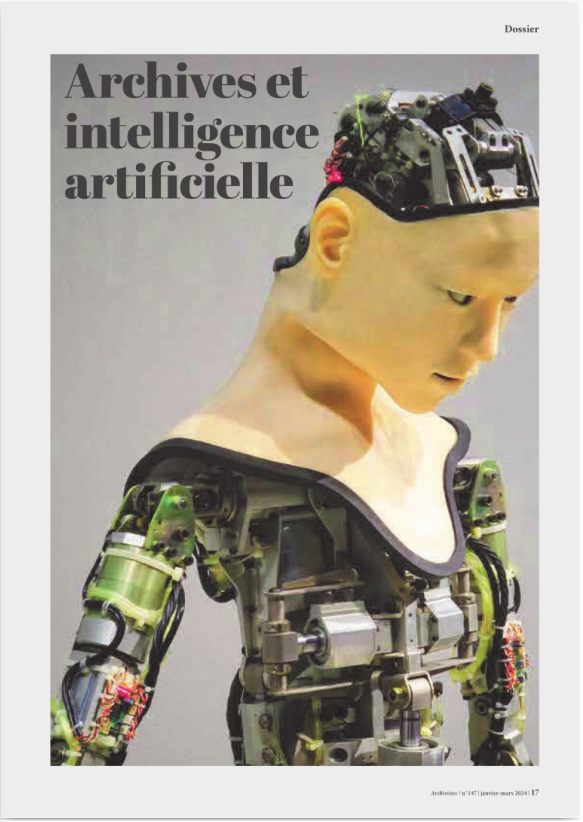
ARCHIVES NATIONALES
Pierrefitte-sur-Seine – Paris

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

Les usages de l'IA dans les archives : pourquoi ?

Les usages de l'IA dans les archives

- Depuis 2015, utilisation de l'IA
  - d'abord avec une visée de **recherche**
  - puis de **valorisation**
  - à présent dans tous les aspects du métier (**collecte** et **description archivistique**)



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

Les usages de l'IA dans les archives : exemples

- Transcrire des archives manuscrites.
- Océriser des bordereaux de versement papier.
- Création de scripts et de macros pour automatiser des traitements sur des archives bureautiques.
- Description d'archives textuelles ou iconographiques (détection et identification de formes).
- Aide à l'indexation d'inventaires.
- Chatbots pour l'orientation du public…

ARCHIVES NATIONALES
Pierrefitte-sur-Seine – Paris

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

IA extractive vs IA générative

Des familles d'IA

IA extractive

Machine Learning

Années 90-2000

Deap Learning

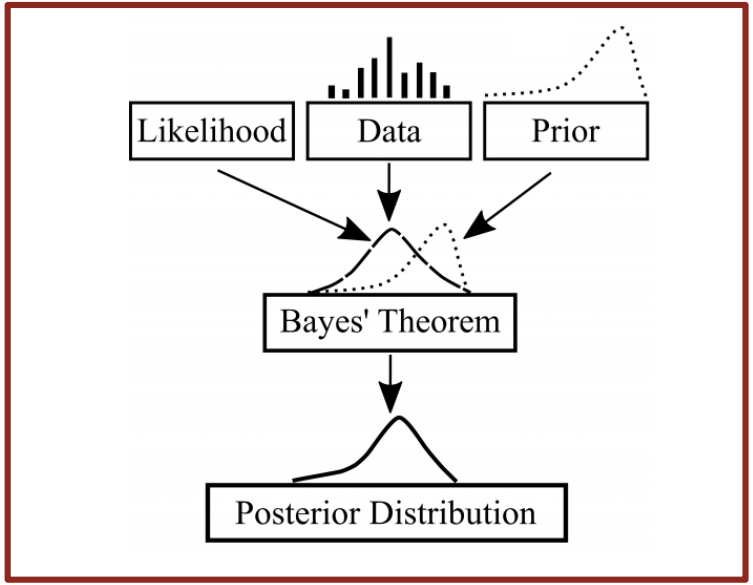
Années 2010

IA générative

Dep. 2020

Principes du Machine Learning/Deep Learning

- Entraîner une IA à effectuer certaines tâches à partir d'un échantillon de données.
- Cet échantillon est fondamental : de lui dépend la qualité du modèle.



= algorithme « apprenant », évolutif ; définit lui-même les résultats en fonction des données qu'on lui fournit.



Images étiquetées avec la valeur du chiffre qu'elles représentent.

Algorithme apprenant

- Exemple : comment développer un modèle pour reconnaître l'écriture ?
- En fournissant des transcriptions, ou **vérité de terrain** (*groundtruth*)

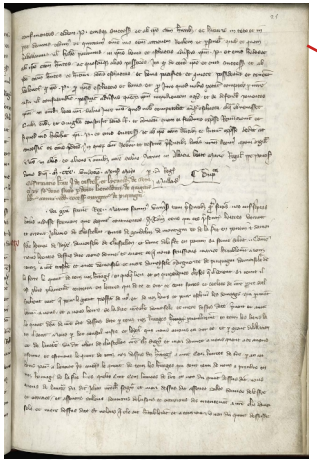


Image.



Télécharger Documenta

‹ CXCIX — Actes royaux du Poitou, t. 2 (1334-1348) CC — CCI ›

[p. 18]

CC

28 janvier 1305

Commission donnée à Étienne de Bourret, sous-doyen de Poitiers, pour la levée de la dime accordée au roi par l'archevêque de Bordeaux, les évêques suffragants et le clergé de la province.

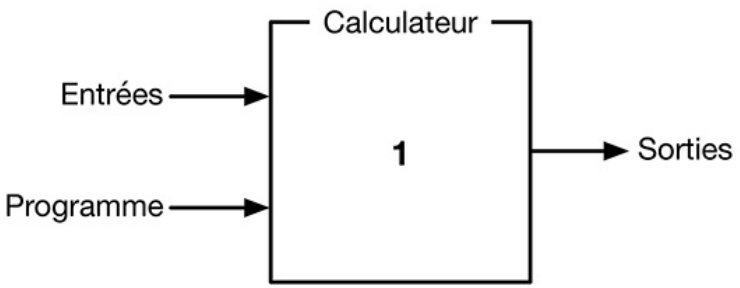
- B AN JJ. 35, n° 213, fol. 110 et JJ. 36, n° 203, fol. 87 v°
- a P. Guérin, Archives historiques du Poitou, 13, p. 18

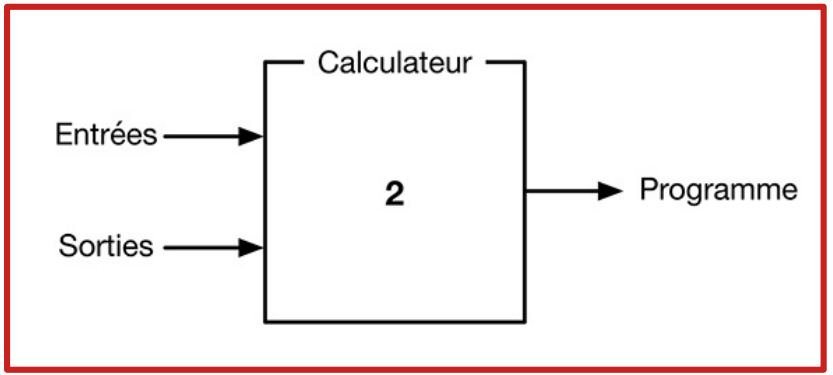
D'après a.

Philippus, Dei gratia, Francorum rex. Universis senescallis, ballivis, prepositis ceterisque justiciariis regni nostri, ad quos presentes littere pervenerint, salutem. Dilectum nostrum magistrum Stephanum de Bourreto[1], subdecanum Pictavensem, de cujus fidelitate et industria plenè confidimus, ad partes provincie Burdegalensis, pro exigenda et levanda pro nobis et nomine nostro decima nobis liberaliter concessa à prelatis et personis ecclesiasticis dicte provincie, et ad substituendum et subrogandum in singulis dyocesibus et locis ejusdem provincie, ubi et de quibus expedire viderit, loco sui ad hoc personas ydoneas, et faciendum omnia et singula que circa hoc fuerint facienda, tenore presencium, specialiter destinantes, vobis et vestrum singulis damus presentibus in mandatis ut eidem magistro Stephano et deputatis seu deputandis ab ipso in hac parte pareatis et intendatis efficaciter parerique et intendi faciatis, alter vestrum ad hoc alteram non expectans. Actum Parisius, xxviii. die januarii anno Domini m. ccc. quarto.

Transcription du texte de l'image.

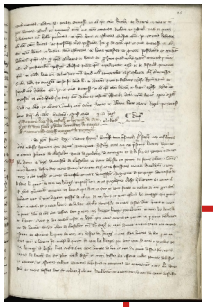
Algorithme apprenant

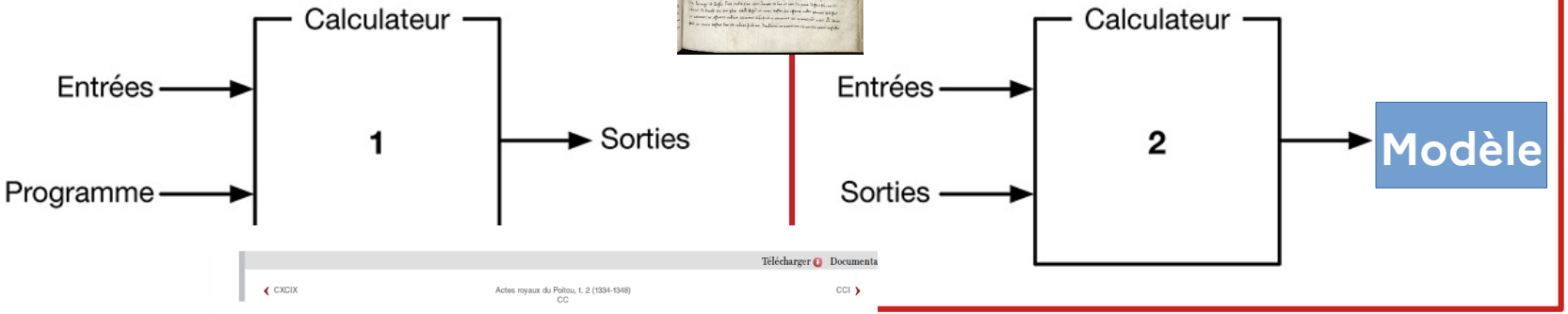


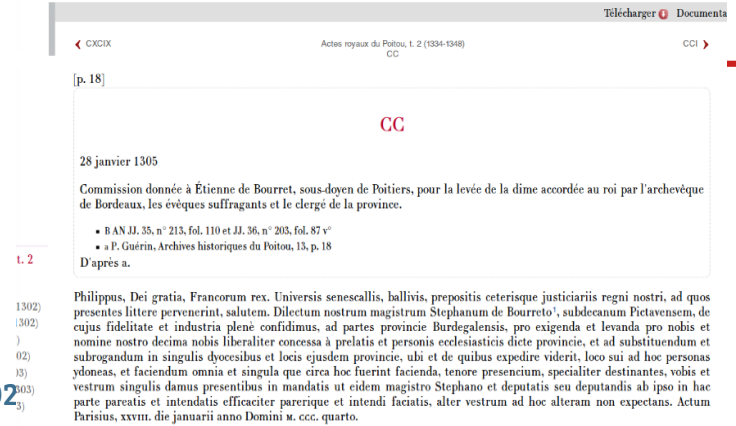


Systèmes « apprenant » :
on se passe de règles.
Elles sont déduites par
comparaison des données.

Algorithme apprenant







Télécharger Documents

CXCIX — Actes royaux du Poitou, t. 2 (1334-1348) — CCI

CC

[p. 18]

CC

28 janvier 1305

Commission donnée à Étienne de Bourret, sous-doyen de Poitiers, pour la levée de la dime accordée au roi par l'archevêque de Bordeaux, les évêques suffragants et le clergé de la province.

- B AN JJ. 35, n° 213, fol. 110 et JJ. 36, n° 203, fol. 87 v°
- a P. Guérin, Archives historiques du Poitou, 13, p. 18

D'après a.

t. 2

Philippus, Dei gratia, Francorum rex. Universis senescallis, ballivis, prepositis ceterisque justiciariis regni nostri, ad quos presentes littere pervenerint, salutem. Dilectum nostrum magistrum Stephanum de Bourreto[1], subdecanum Pictavensem, de cujus fidelitate et industria plenè confidimus, ad partes provincie Burdegalensis, pro exigenda et levanda pro nobis et nomine nostro decima nobis liberaliter concessa à prelatis et personis ecclesiasticis dicte provincie, et ad substituendum et subrogandum in singulis dyocesibus et locis ejusdem provincie, ubi et de quibus expedire viderit, loco sui ad hoc personas ydoneas, et faciendum omnia et singula que circa hoc fuerint facienda, tenore presencium, specialiter destinantes, vobis et vestrum singulis damus presentibus in mandatis ut eidem magistro Stephano et deputatis seu deputandis ab ipso in hac parte pareatis et intendatis efficaciter parerique et intendi faciatis, alter vestrum ad hoc alteram non expectans. Actum Parisius, xxviii. die januarii anno Domini m. ccc. quarto.

Systèmes « apprenant » :
on se passe de règles.
Elles sont déduites par
comparaison des données.

IA extractive

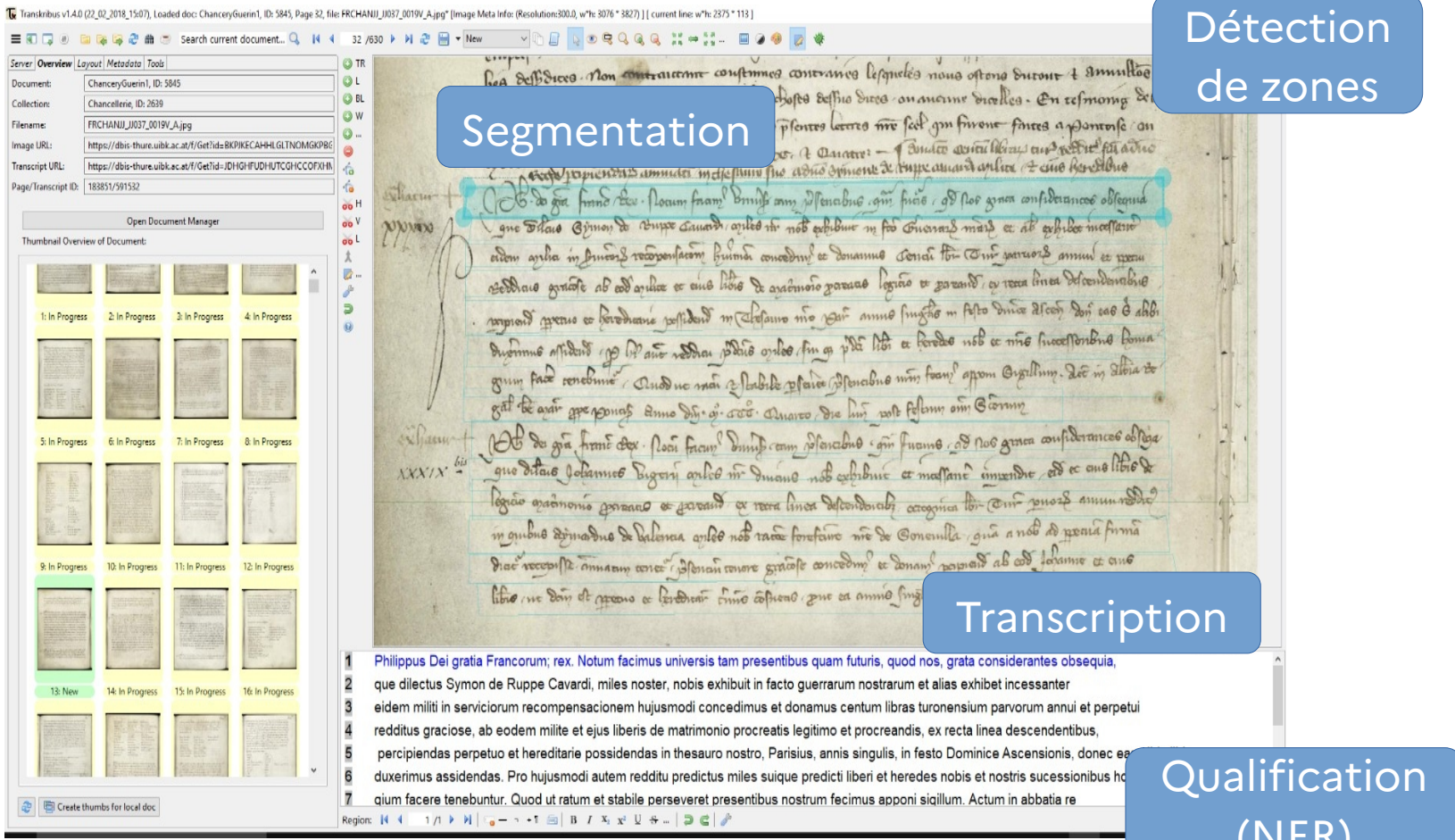
- Développement de modèles spécialisés pour certaines tâches d'extraction de l'information.

- Ex. :
  - Détection de mise en page/zonage (*Yolo*)
  - Transcription de texte manuscrit (*Kraken*)
  - Reconnaissance de formes dans une image (*SAM2*)
  - Identification de termes clés… (*spaCy*)

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

ARCHIVES NATIONALES

Exemple :
transcription de texte manuscrit



MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

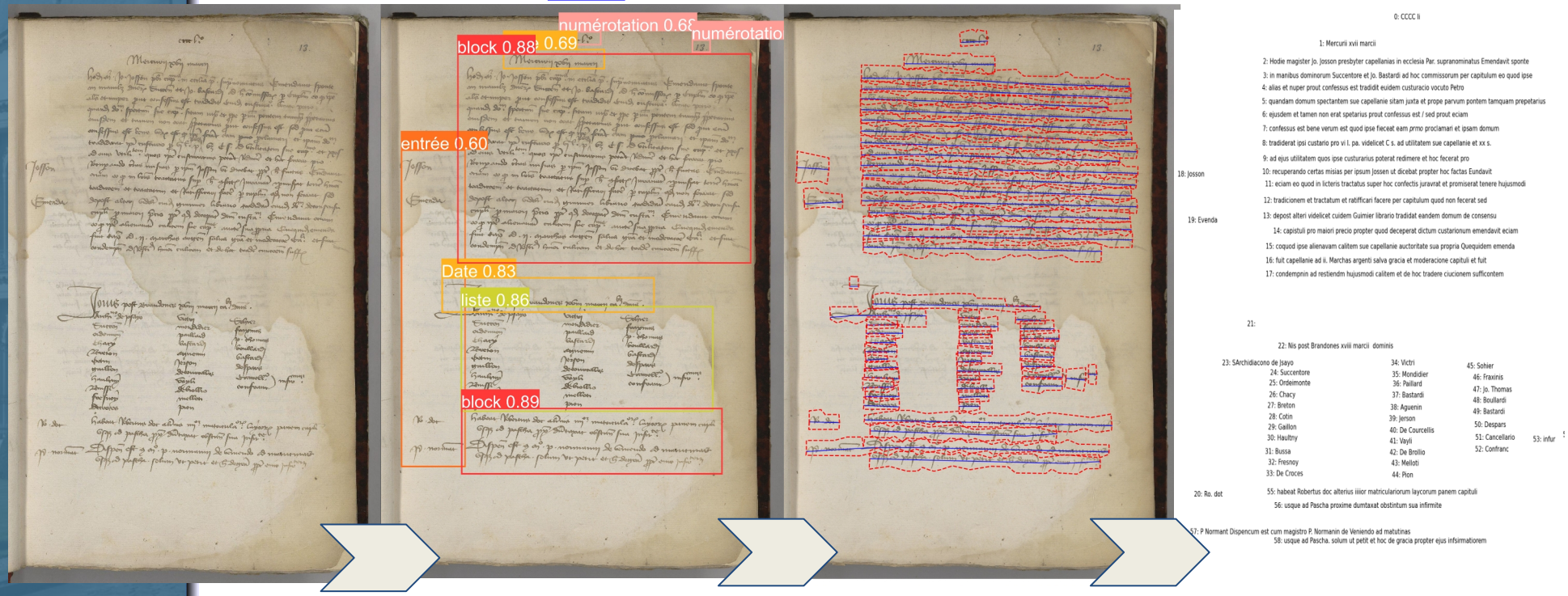
ARCHIVES NATIONALES

Image source

1) **Analyse de la mise en page** (optionnel) - réalisé avec modèle Yolov5

2) **Segmentation** (lignes de bases et polygones) - ici avec eScriptorium/Kraken

3) **Transcription** - ici avec eScriptorium/Kraken



Approche « analytique ». Combinaison de différents modèles spécialisés dans une chaîne de traitement.

IA extractive vs IA générative

\+

- Utilisation de modèles spécifiques pour extraire certains types d'information => **performants**.
- Possibilité d'améliorer (« fine-tuner ») en fournissant d'autres données d'entraînement.
- Peu énergivore.

\–

- Découpage des tâches => complexification, risques d'erreurs
- Plusieurs modèles à entraîner

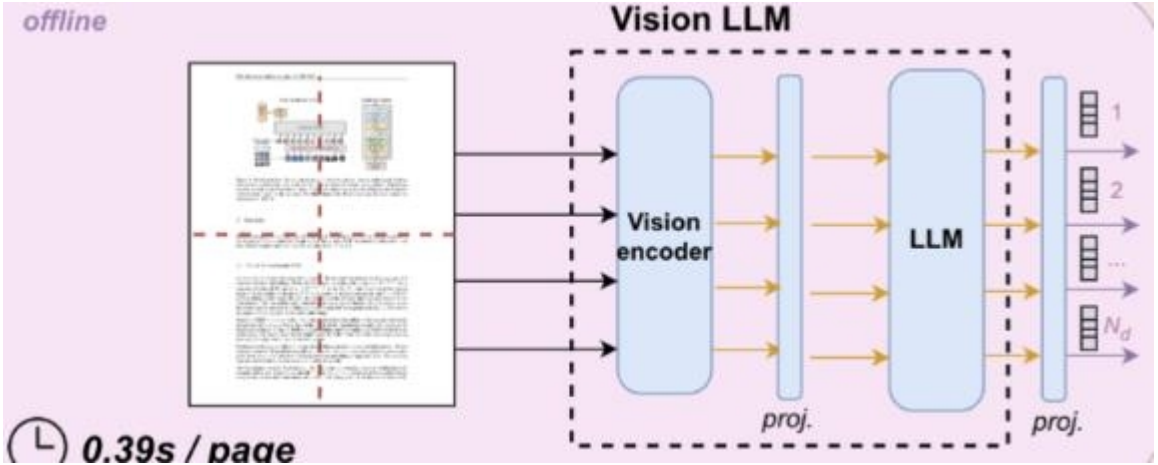
Principes de l'IA générative

- Pour les outils grands publics, interaction par des instructions écrites => ***prompts***

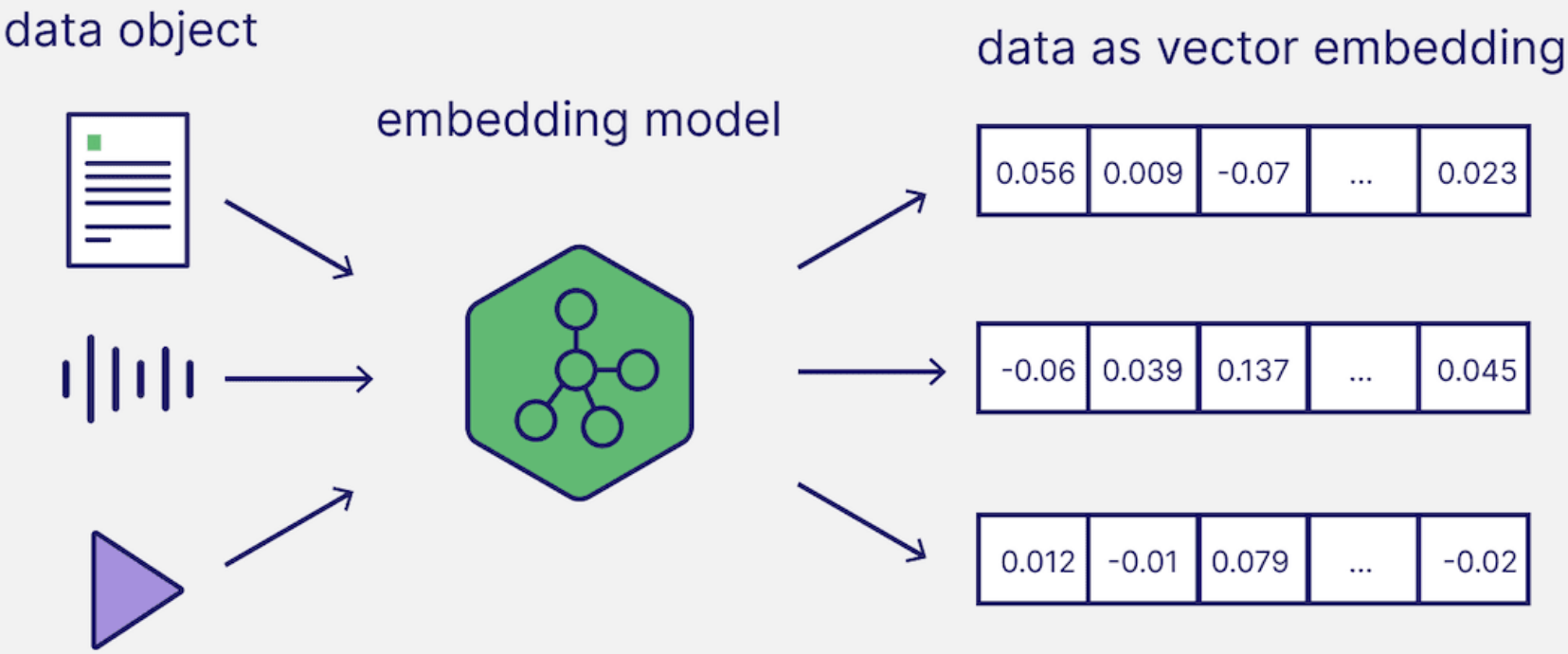
- Disponibilité de modèles pré-entraînés à partir d'un grand corpus de données : le Web.

Principes de l'IA générative

- Approche globale :
  - création d'une représentation mathématique (***embedding***, vectorisation) de la donnée.
  - génération d'une réponse ou d'un résultat à partir de cette représentation mathématique.



Principes de l'IA générative



Principes de l'IA générative



Analyse de la matrice de pixels

```
[
 [[34, 120, 200],
  [35, 118, 199], ...],
 ...
]
```

Analyse du contenu visuel (bords, coins, textures, couleurs, motifs, yeux, moustaches, forme de la tête, posture, etc.)

```
[
 [0.22, -1.13, 0.55, ...]
]
```

Principes de l'IA générative

- Ex. :
  - Chat : `[0.21, -0.55, 1.02, …]`
  - Chien : `vecteur proche`
  - Voiture : `vecteur éloigné`

Exemple : instruction prompt (Gemini)

Bonjour. Es-tu capable d'analyser une image et de détecter des formes.
Je voudrais te soumettre l'image d'un rouleau médiéval composé de
plusieurs peaux de parchemins. Les peaux de parchemin sont cousues entre
elles par du fil de lin. Pourrais-tu me donner les coordonnées sur
chaque image des différentes peaux de parchemin ?

Bonjour ! Oui, je suis tout à fait capable d'analyser des images et d'identifier des éléments spécifiques comme des formes, des textures ou des délimitations structurelles.

Dans le cas de votre **rouleau médiéval**, je peux examiner les zones de chevauchement et les coutures en fil de lin pour tenter de distinguer les différentes peaux (ou "membranes").

Exemple : segmentation par prompt d'un rouleau médiéval

Ce que je peux faire pour vous :

- **Identification visuelle :** Repérer les transitions entre les peaux de parchemin.
- **Localisation :** Décrire précisément où se trouvent les limites ou fournir des coordonnées relatives (par exemple, en pourcentage de la hauteur/largeur de l'image) pour chaque section identifiée.
- **Analyse de l'état :** Noter d'éventuelles irrégularités dans les coutures ou des différences de grain entre les peaux.

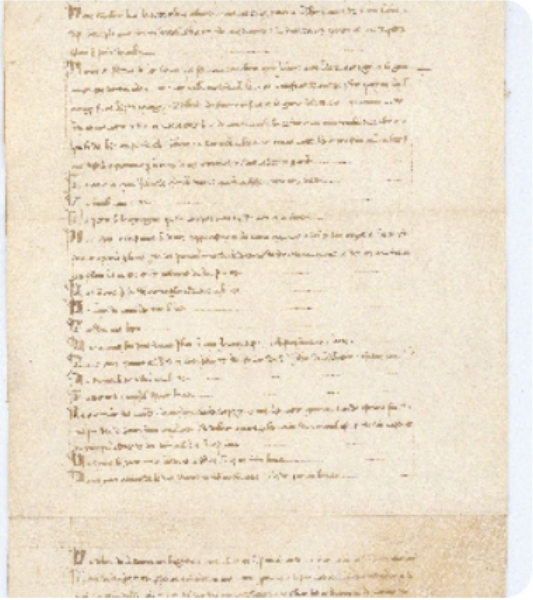
Comment procéder ?

1. **Téléchargez l'image** (ou les images) du rouleau.
2. Si possible, assurez-vous que la résolution permet de bien voir les **points de couture**, car c'est souvent l'indice le plus fiable pour séparer deux peaux.

Dès que vous m'aurez transmis le fichier, je commencerai l'analyse pour vous fournir les coordonnées de chaque segment.

Souhaitez-vous me soumettre l'image dès maintenant ?

Exemple : segmentation par prompt d'un rouleau médiéval



Oui, je vais te transmettre une image jpeg de la fin d'un testament sur rouleau. Je voudrais les coordonnées des pixels des différentes membranes que tu seras capable de repérer pour ensuite pouvoir proposer un zonage de l'image en fonction des différentes membranes.

Exemple : segmentation par prompt d'un rouleau médiéval

Analyse

C'est un magnifique document. Sur cette image, on distingue nettement les zones de recouvrement où les membranes ont été assemblées.

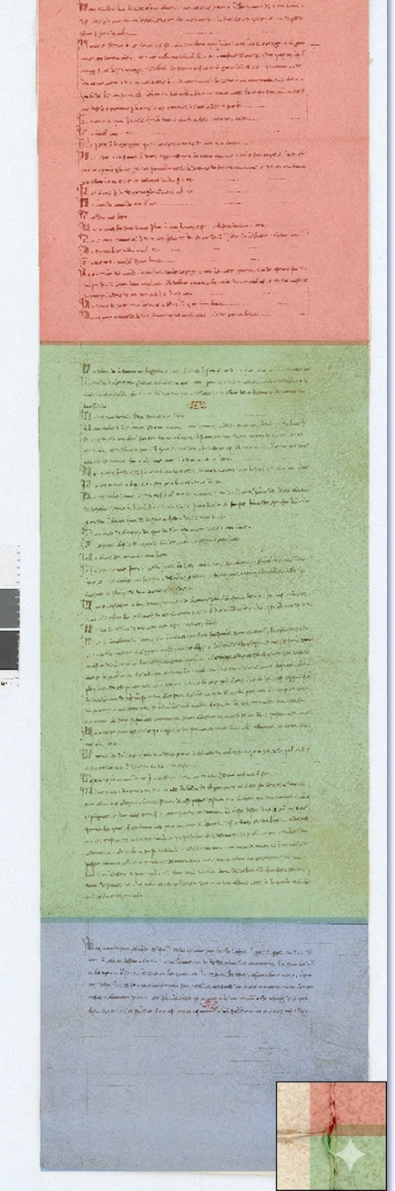
D'après l'examen visuel des coutures et des changements de texture/couleur du parchemin, j'ai identifié **trois membranes (peaux) distinctes**.

Voici les coordonnées approximatives sous forme de boîtes de délimitation (bounding boxes). Les coordonnées sont exprimées en pixels [y_min, x_min, y_max, x_max] sur la base de votre image (environ 343 x 1000 px) :

Exemple : segmentation par prompt d'un rouleau médiéval

Coordonnées des membranes

Membrane	Zone (Pixels [y_min, x_min, y_max, x_max])	Description visuelle
Membrane 1	`[0, 68, 285, 935]`	La partie supérieure. La couture de fin est visible juste avant le bloc de texte central.
Membrane 2	`[285, 85, 765, 948]`	La section centrale, la plus longue, contenant le corps principal du texte.
Membrane 3	`[765, 98, 988, 938]`	La section inférieure (la fin du testament), avec une marge plus importante en bas.



Exemple : segmentation par prompt d'un rouleau médiéval

Oui je veux bien que tu me génères un masque visuel avec les zones colorées sur mon image. Cela me permettrait de vérifier la segmentation que tu as faite.

Voici une visualisation de la segmentation de l'image en trois membranes, superposée à votre document original.

- **Zone Rouge (Haut) :** Première membrane.
- **Zone Verte (Milieu) :** Deuxième membrane (la section la plus longue).
- **Zone Bleue (Bas) :** Troisième membrane.

Le masque utilise une légère transparence pour vous permettre de vérifier que les délimitations colorées correspondent bien aux coutures et aux changements de texture du parchemin visibles sur l'original.

IA extractive vs IA générative

+

- Utilisation de modèles spécifiques pour extraire certains types d'information => **performants**.
- Possibilité d'améliorer (« fine-tuner ») en fournissant d'autres données d'entraînement.
- Peu énergivore.

–

- Découpage des tâches => complexification, risques d'erreurs
- Plusieurs modèles à entraîner

+

- Rapidité : une seule étape pour réaliser des tâches complexes.
- Facilement accessibles.

–

- Très énergivore.
- Modèles difficiles à spécialiser.
- Risques d'hallucination ou d'inventions plus grand.

ARCHIVES NATIONALES
Pierrefitte-sur-Seine – Paris

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

Retour sur quelques projets d'IA appliquée aux archives (en France principalement)

Les usages de l'IA aux Archives nationales

IA « extractive »

(a. 1990-2010)

- 2015-2017 : HIMANIS (projet IRHT)
  - Transcription par IA (*HTR*) des registres de la chancellerie royale française (XIII^{e}-XV^{e} siècles)
- 2018-2021 : LectAuRep (Minutier Central)
  - Transcription de registres de notaires parisiens (fin XVIII^{e}-début XX^{e} siècles).
- 2020-2024 : e-NDP (Paris Cité, Paris 1, ENC)
  - Transcription des registres capitulaires de Notre-Dame de Paris (XIV^{e} siècle)
- 2020-2025 : NER4Archives (Lab des AN)
  - Extraction d'entités nommées (qualification de données textuelles)
- 2021-… : SIMARA
  - Rétroconversion d'inventaires manuscrits anciens (avec HTR et NER)

L'usage de l'IA dans le réseau territorial

IA « extractive »

(a. 1990-2010)

- 2019-2024 : Lettres en Lumières (AD Côte-d'Or)
  - Transcription par IA (*HTR*) des registres des registres des états de Bourgogne (XVIII^{e} siècle)
- 2021-2025 : SOCFace (réseau AD-SIAF)
  - Transcription par IA (*HTR*) des recensements de population de 1836 à 1936

ARCHIVES NATIONALES
Paris – Pierrefitte-sur-Seine

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

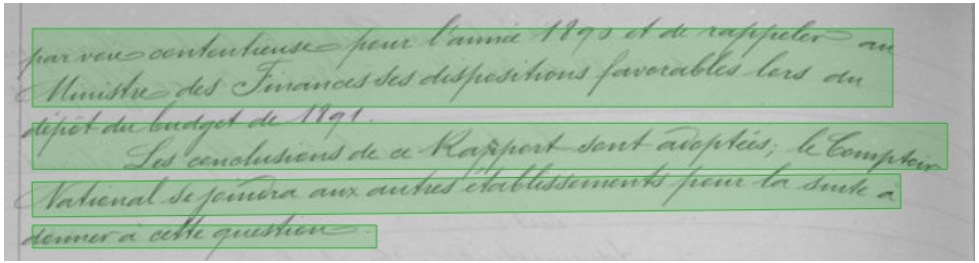


Himanis (2015-2017)

La recherche en plein texte dans les archives royales

Les technologies d'IA exploitées

- Handwritten text recognition (HTR) : reconnaissance des écritures manuscrites.



- Named Entities Recognition (NER) : reconnaissance des entités nommées.

prétendue reformée à TITLE Veyne LOCATION en Dauphine LOCATION, Bouteroue PERSON, Murat PERSON, Blanc PERSON, Dize PERSON et Charles PERSON aussi ministres de lad TITLE. religien 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 renvoye en la Chambre de l'edit de Grenoble ORGANISATION

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

CULTURAL HERITAGE
A CHALLENGE FOR EUROPE

Financement : 410 k€

HIMANIS

- **2015-2017**
- Conception et portage : Dominique Stutzmann (IRHT). Partenaires : A2iA, Teklia, universités de Valence et Groningue.
- Institutions associées : Archives nationales, Bibliothèque nationale de France.
- Financement : Europe. JPI CH (*Joint Programming Initiative on Cultural Heritage*) — 410 000 €.
- Corpus : registres de la chancellerie royale (XIII^{e}-XV^{e} siècles) ; 80 000 images.
- Outils : Transkribus (HTR), Transkriptorium (interface de consultation)
- Finalités : recherche en texte intégral, statistiques pour la paléographie.
- Données et modèles : éditions de P. Guérin et OCR Gallica d'éditions (vérité de terrain), github (index HTR, inventaires TEI, manifestes IIIF).

HIMANIS

- **Carnet de recherche**
  - https://himanis.hypotheses.org/
- **Interface de recherche**
  - http://himanis.huma-num.fr/

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

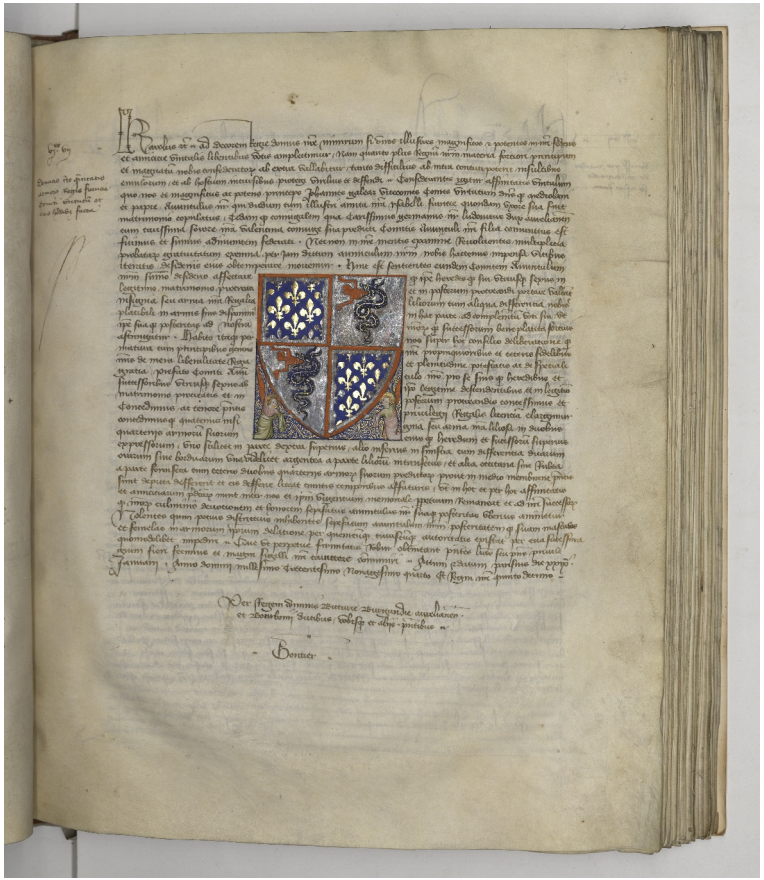
ARCHIVES NATIONALES

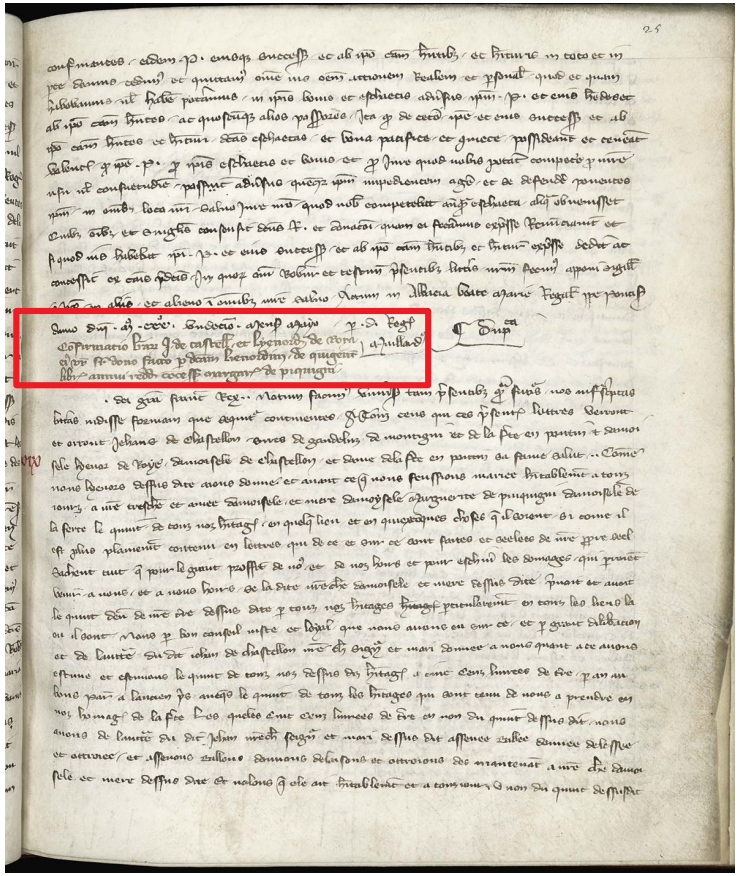
Les fonds concernés par HIMANIS

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.

HIMANIS

- **Finalités :** moteur de recherche en texte intégral ; exploitation statistique des index (termes) ; études pour la paléographie (personnel de chancellerie)

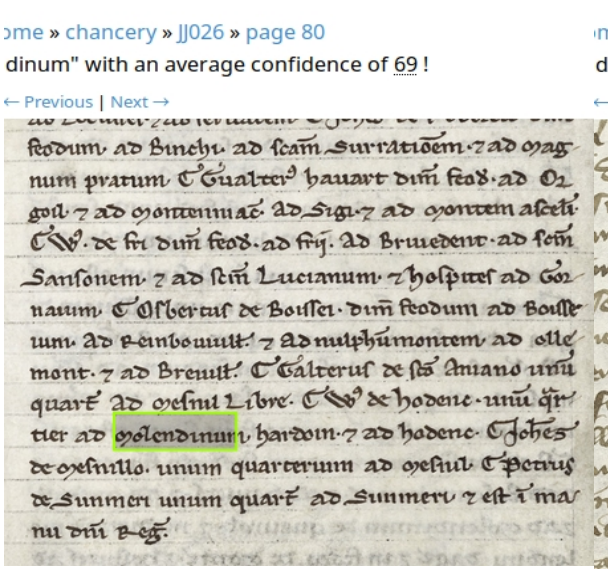
- **Outils :** **<u>un</u>** des outils utilisés : Transkribus (HTR) (pour le contrôle de la segmentation), Transkriptorium (interface de consultation reposant sur le technologie de consultation développée par Valencia)

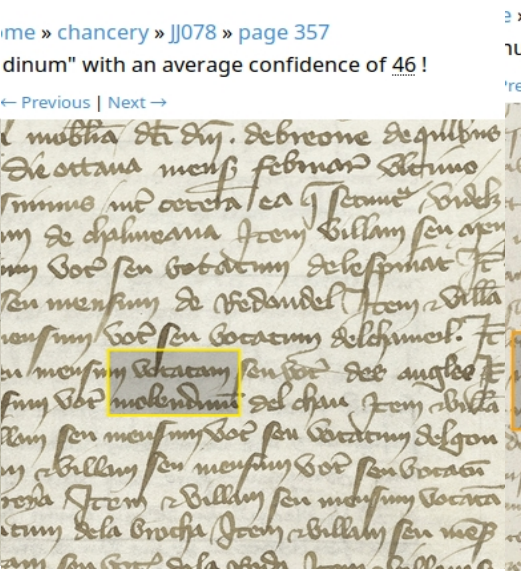
Transkribus tranSkriptorium

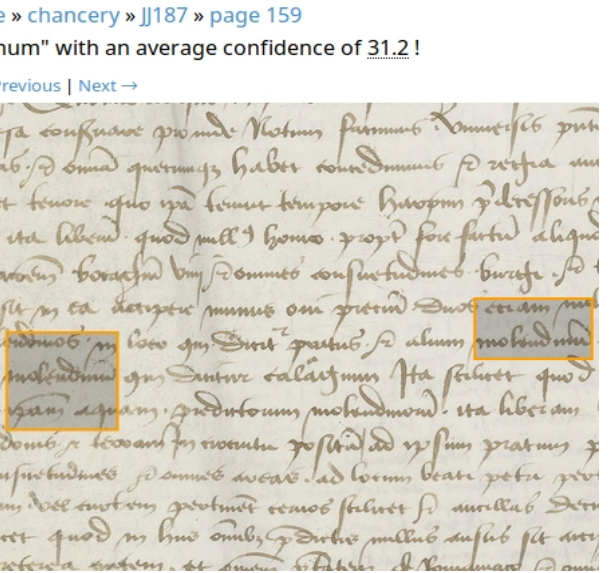
Interface de consultation

- **Diffusion :** moteur de recherche en texte intégral

Exemple : recherche du terme “molendinum” (moulin)

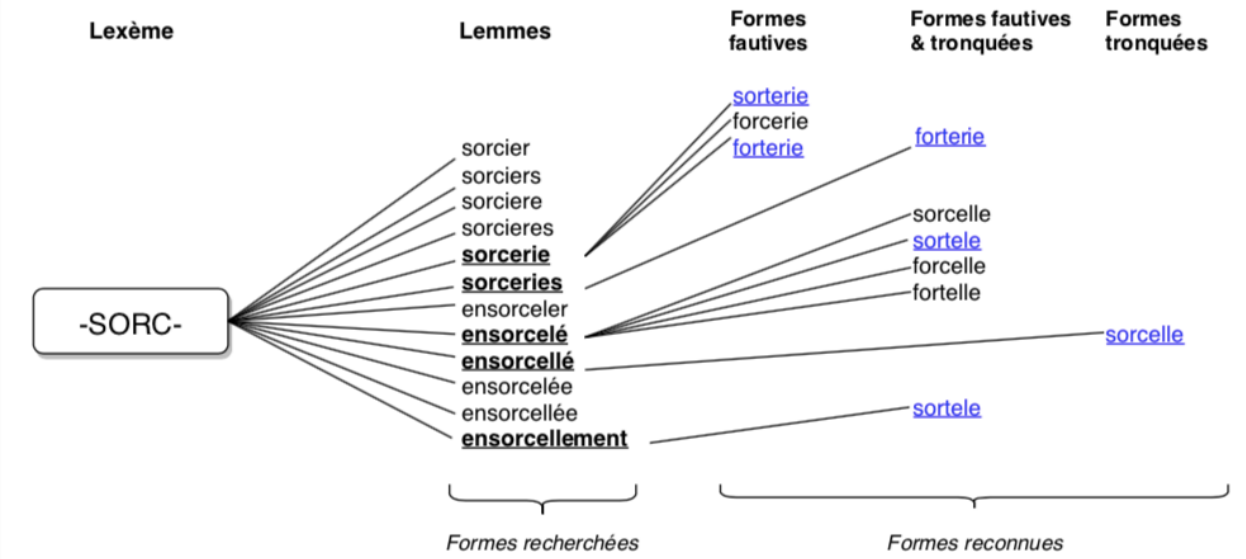






http://himanis.huma-num.fr/himanis/

Les archives replacées au cœur de la recherche

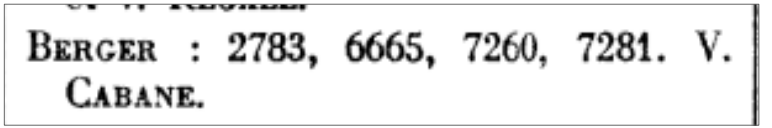


https://himanis.hypotheses.org/470

Himanis : analyse statistique lexicale autour de la sorcellerie.

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

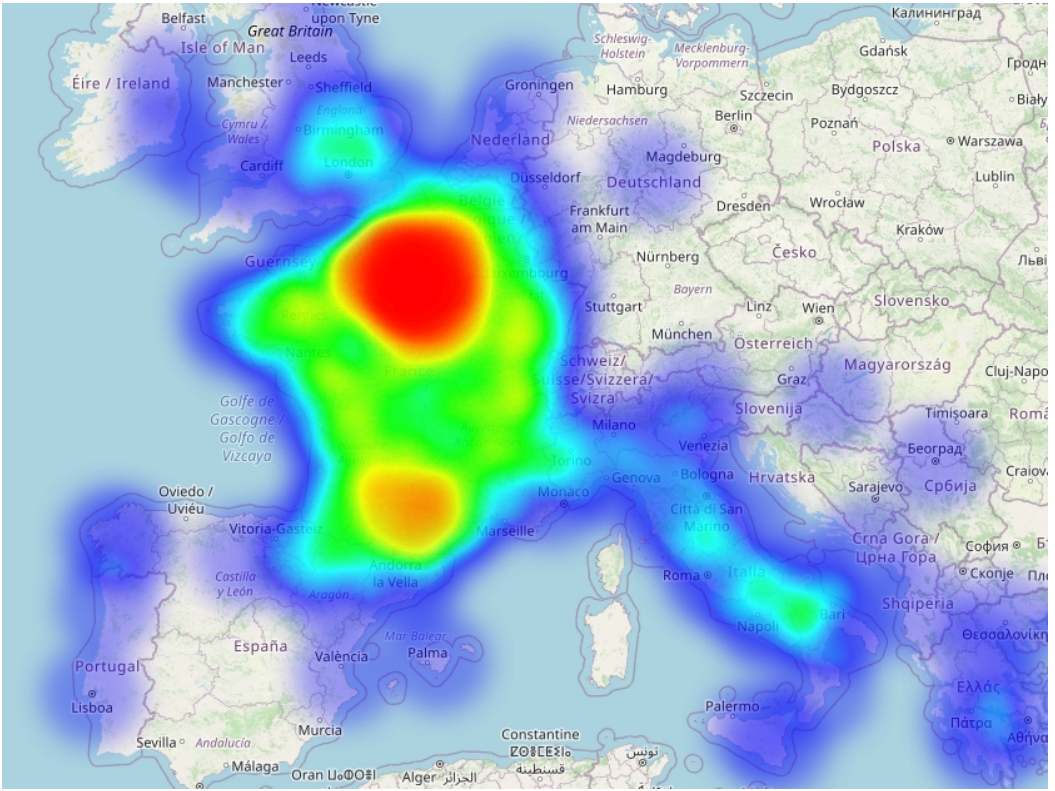
- La recherche en plein texte :
  - Ne remplace pas les index que pouvaient concevoir les archivistes
    - Identifications des noms de lieu et de personne, mais aussi des concepts…
    - Désambiguïsation
    - Connaissance de la langue ancienne pour chercher le bon terme ou concept…





https://himanis.hypotheses.org/1106

Le projet est officiellement clos, mais l'IRHT continue à explorer les données en s'intéressant aux index et à leur amélioration en croisant avec les données textuelles.





ARCHIVES NATIONALES
Paris – Pierrefitte-sur-Seine

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

LECTAUREP (2018-2021)

Utiliser les ressources des notaires pour créer des métadonnées descriptives

LECTure AUtomatique de REPertoires

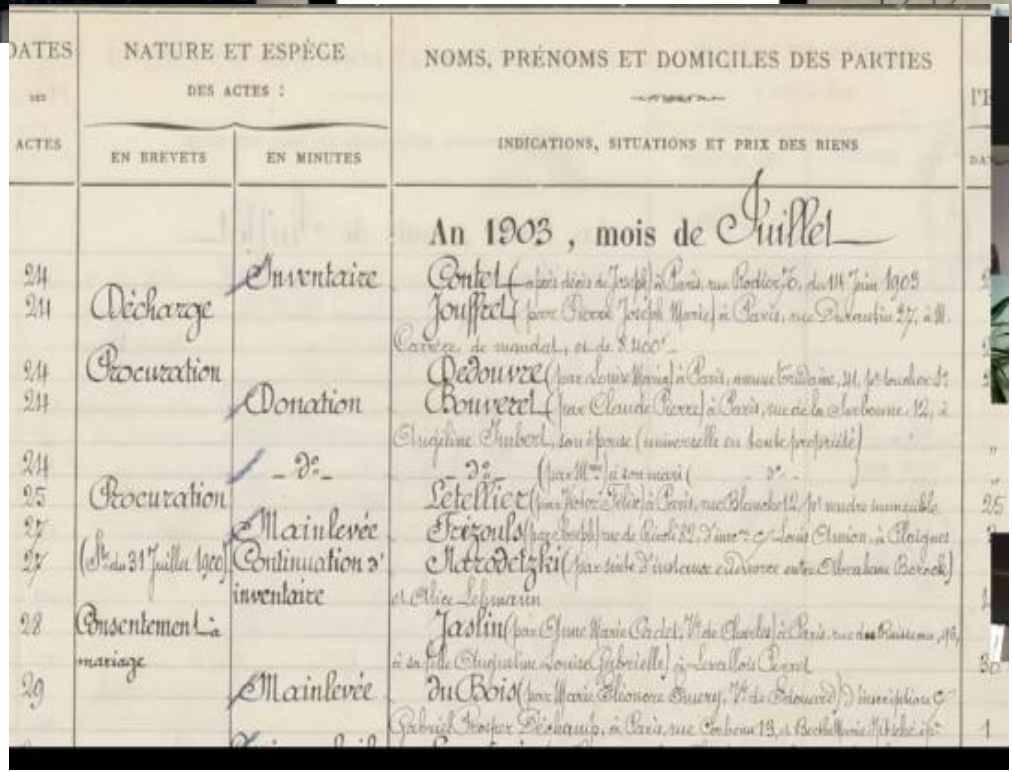
https://lectaurep.hypotheses.org/





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

Un inventaire en soi…



03/11/2026

Webinaire AIAF

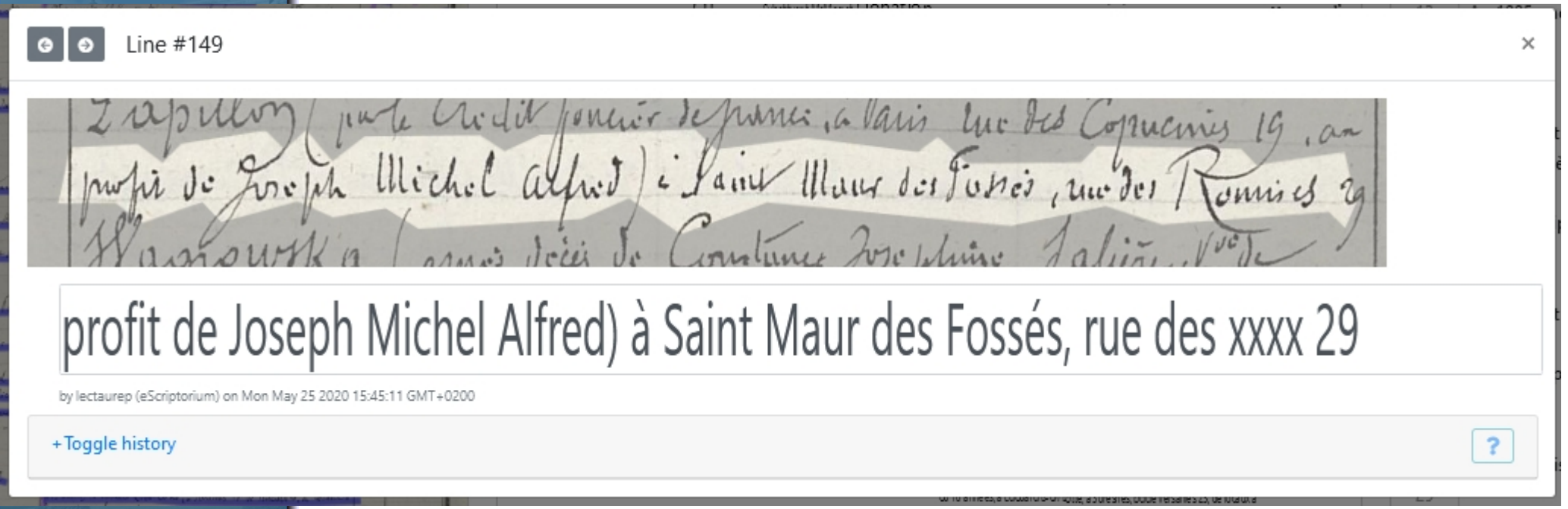
LECTure AUtomatique de REPertoires

https://lectaurep.hypotheses.org/



- Conception et portage : Aurélia Rostaing (DMC/Archives nationales)
- Partenaires : INRIA (convention MC/INRIA).
- Financement : Ministère de la Culture
- Corpus : registres du minutier central des notaires parisiens (XVIIIe-XXe s.).
- Finalités : transformer des répertoires de notaires en métadonnées ; reconnaissance d'entités nommées.
- Outils : eScriptorium

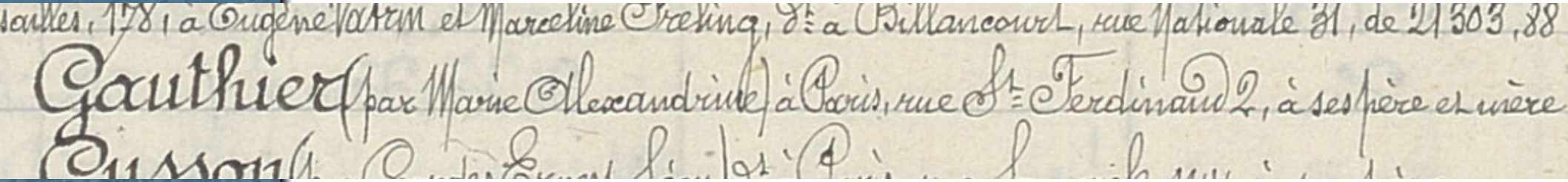
Processus



HTR avec le logiciel Kraken et son interface eScriptorium (à l'époque préféré à Transkribus pour son caractère non commercial).

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 la NER pour extraire des termes d'indexation.

https://lectaurep.hypotheses.org/

Données du projet

Publication d'un modèle de transcription et de la vérité de terrain !

alixchague

12/05/2022

Création des modèles, Tutoriel

Laisser un commentaire

Afin de rendre visible le travail colossale d'annotation réalisé par les équipes des Archives nationales, les transcriptions susceptibles de servir de données d'entraînement sont désormais publiées au sein du projet HTR-United, dans 3 répertoires Github distincts. Leur publication permet désormais à d'autres utilisateurs et utilisatrices de l'HTR de les réemployer pour leur propres expérimentations.

Il s'agit des répertoires suivants :

– LECTAUREP Repertoires : qui contient les pages de répertoires, transcrites et validées (de nombreuses nouvelles pages sont en cours de relecture).

– LECTAUREP Bronod : qui contient les minutes rédigées par Maître Bronod, dont le format diffère du corpus de répertoires car elles datent de 1742.

– LECTAUREP Mariages et Divorces : qui contient les registres des Mariages et Divorces, pour le début du XXe siècle principalement.

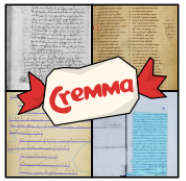
Carnet de recherche :
https://lectaurep.hypotheses.org/

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

ARCHIVES NATIONALES

eScriptorium

- Fortement adossé à des communautés des humanités numériques et de l'ENC
  - CREMMA-Lab



Consortium pour la Reconnaissance d'Écriture Manuscrite des Matériaux Anciens

Ouverture du service HTR Cremma

31 déc. 2025 - 1 janv. 2028 Paris (France)

Connexion

NAVIGATION

Accueil

Gabarit et instructions sur les dépots

Déposer une demande

Partenaires

SUPPORT

@ Contact

BIENVENUE !

Le Consortium Reconnaissance d'Écriture Manuscrite des Matériaux Anciens (CREMMA) offre aux projets retenus un accès gratuit à l'application de transcription automatique eScriptorium, déployée sur des serveurs permettant l'entraînement de modèles performants et leur application à de grandes quantités de documents.

Cet accès permet une utilisation normale de l'application (création de données, gestion de projets de transcription, application des modèles de segmentation ou de transcription). Le droit d'entraîner des modèles est activé seulement s'il est demandé.

Le Consortium est porté par Inria et l'École nationale des chartes (Centre Jean-Mabillon), et est financé par le DIM Matériaux anciens et patrimoniaux. Il est soutenu par le LAMOP, l'IRHT, l'EFEO et l'EPHE.

Sont invités à solliciter l'accès au service tous les projets en lien avec l'écrit patrimonial, quel que soit le matériau (manuscrits médiévaux ou modernes, supports muraux, etc.), sans préjuger des langues et des écritures cibles, et quel que soit la taille du projet ou de l'équipe. Leur recevabilité est validée par le Consortium au regard de la pertinence des propositions et de la disponibilité du service.

Les projets retenus s'engagent en contrepartie à partager, sous licence ouverte et selon une méthode documentée, la vérité de terrain produite, ainsi que les éventuels modèles de transcription automatique. Ces jeux de données partagés permettent d'entraîner de nouveaux modèles performants et, partagés en retour, ils bénéficient à d'autres utilisateurs de la technologie.

DÉPOSER UNE DEMANDE

Nous testons une nouvelle méthode pour recueillir les informations sur les projets: n'hésitez pas à remplire le formulaire suivant:

- **Pour les projets,** https://forms.gle/emyFeCdnRuUZVgGL7
- **Pour les doctorant-e-s et étudiant-e-s**, https://forms.gle/BQTACy67S1M8F8YB6

Pour déposer une demande, rendez-vous d'abord sur la "Gabarit et instructions sur les dépôts" pour prendre connaissances des recommandations concernant le dépôt de demande et obtenir le gabarit proposé par le consortium. Ensuite, vous pouvez vous rendre sur la page "Déposer une demande" pour soumettre vos documents.

LOGOS

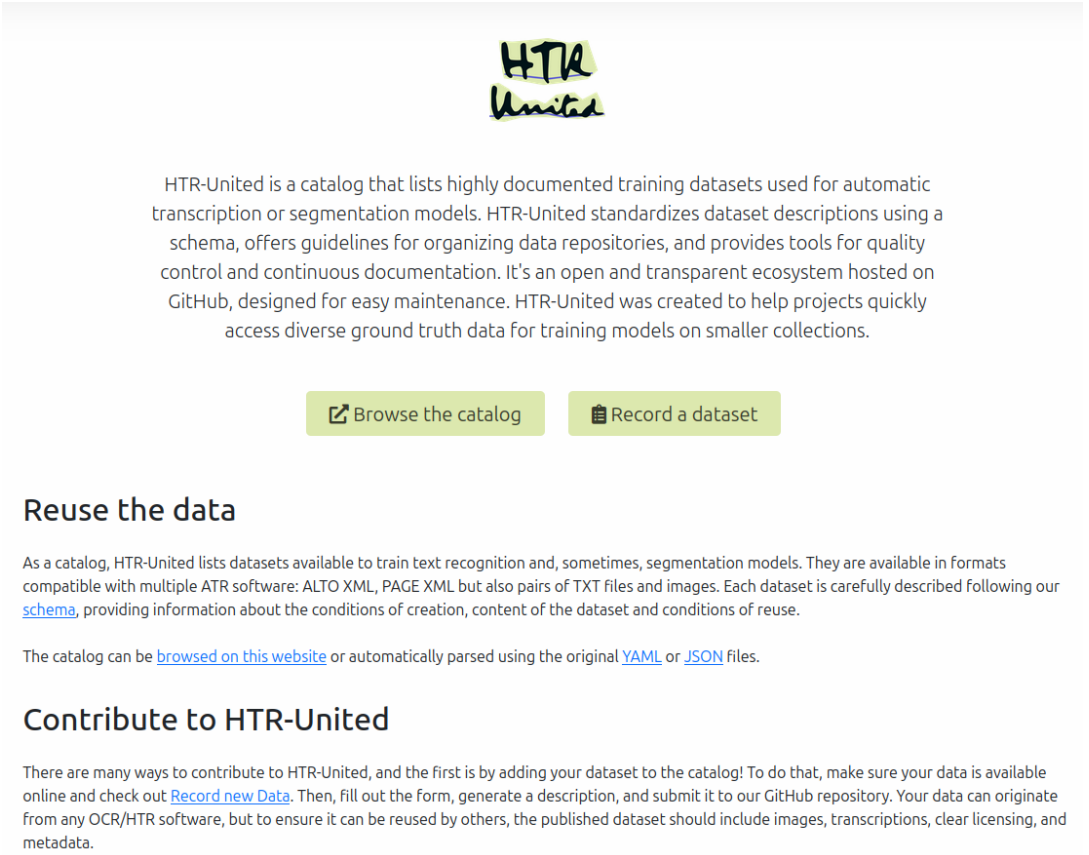
dim matériaux anciens et patrimoniaux — Région Île-de-France — Inria — École nationale des chartes | PSL

03/11/2026

Webinaire AIAF

eScriptorium

- Fortement adossé à des communautés des humanités numériques et de l'ENC
  - HTR-United



03/11/2026

Webinaire AIAF

Catalogue

La langue utilisée pour la description des jeux de données correspond à celle employée lors du dépôt d'un signalement dans HTR-United.

Filtres

Langue: Toutes les langues

Système d'écriture: Tous les systèmes d'écriture

Type d'écriture: Manuscrit uniquement

Projet: Tous les projets

Dates: Not before: -250 | Not after: 2024

Nombre de résultats: 71

Description statistique des résultats

Unité	Total	Projets avec cette unité
Caractères	49 121 231	47
Lignes	1 259 548	60

Options

- [ ] Afficher les normes de transcription
- [ ] Afficher les informations de citation

CREMMA-AN Testament De Poilus

Testaments de Poilus
1914 - 1918

Lien Dépôt de données | Lien Fichier Citation (CFF)

Language fra | Script Latn | Script Type only-manuscript | Hands 1-per-file

Volume 87 726 characters | Volume 226 files | Volume 3 330 lines

ANR e-NDP Ground Truth

ANR e-NDP
1326 - 1504

Lien Dépôt de données

Language fra | Language lat | Script Latn | Script Type only-manuscript

Hands more-than-10



ARCHIVES NATIONALES
Paris – Pierrefitte-sur-Seine

NER4Archives. 2020-2025

L'utilisation de la reconnaissance des entités nommées pour qualifier les données archivistiques

NER4Archives : reconnaissance d'entités nommées pour les services d'archives

Depuis fin 2020, les Archives nationales et l'Inria, avec le soutien du ministère de la Culture, travaillent à la reconnaissance des entités nommées dans les instruments de recherche au format XML/EAD. C'est le projet NER4Archives (Named Entity Recognition For Archives) ou « reconnaissance d'entités nommées pour les services d'archives ».

https://www.archives-nationales.culture.gouv.fr/innovation-et-recherche/ner4archives-reconnaissance-dentites-nommees-pour-les-services-darchives

Historique du projet

- Première étape (relative à la reconnaissance des entités nommées) réalisée entre fin 2020 et fin 2022 dans le cadre d'une première convention cadre passée entre l'Inria, les Archives nationales et le ministère de la Culture.
  - présentation (mars 2022)
- Mise en pause du projet en 2023
- Deuxième étape (relative à la réconciliation et à l'identification des entités nommées) dans le cadre d'une deuxième convention passée entre les mêmes partenaires (avec Cecilia Graiff comme responsable du projet à l'Inria) : décembre 2023-novembre 2024.
  - présentation (mai 2024)
  - Élaborer une méthodologie générale ; la mettre à l'épreuve, entraîner et évaluer les solutions d'IA retenues ; réaliser, documenter et publier en open source un prototype fonctionnel permettant d'avancer concrètement et de préparer l'avenir

NER4Archives-project

GitHub organization for the research project NER4Archives. Developments on named entity recognition methods for XML EAD findings aids. Inria-Archives nationales

10 followers · France

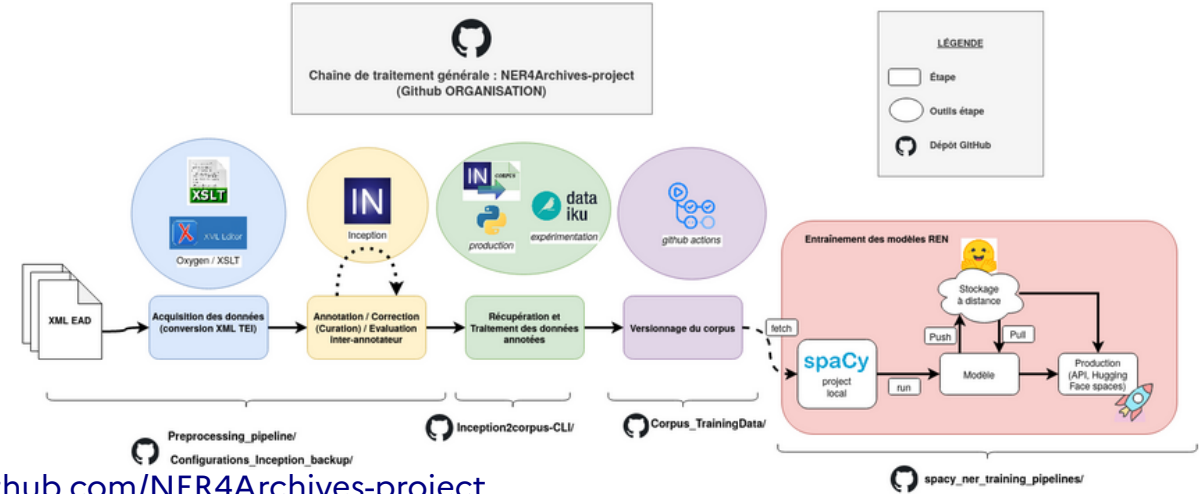
Overview | Repositories 3 | Projects | Packages | People

README.md

NER4Archives Project

🔎 NER4Archives project overview / Présentation générale du projet NER4Archives:

- 🇬🇧 Developments made in collaboration with the French National Archives on named entity recognition & linking methods for XML-EAD finding aids.
- 🇫🇷 Développements réalisés en commun avec les Archives Nationales de France sur les méthodes de reconnaissance et de liage des entités nommées dans les instruments de recherche XML-EAD.



People

This organization has no public members. You must be a member to see who's a part of this organization.

Top languages

Python

https://github.com/NER4Archives-project

NER4Archives



Illustration avec outil displaCy de spaCy

Traitement automatique du langage (TAL)

NER4Archives

PARIS I

J 148

1-5. Rentes provenant de la succession du chambellan royal, Philippe de Nemours.

1. Paris, juin 1256.

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 Lignières, 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 [cf. 2a-2d].

Acte en latin, cancellé, scellé de cire verte sur lacs de soie rouge.

Éd. : *Layettes du Trésor des chartes*, t. III, n° 4266, p. 307.

NER4Archives

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 Lignières, 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



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 Lignières TITLE, et dont l'autre moitié a été échangée, après accord du roi TITLE, avec l'abbaye de Saint-Port (Barbeau ORGANISATION) par un autre fils du chambellan TITLE, Jean PERSON

</> JSON Output

Maximize



ARCHIVES NATIONALES
Paris – Pierrefitte-sur-Seine

eNDP. 2020-2024

Redécouverte des registres capitulaires de Notre-Dame de Paris

e-NDP



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.

e-NDP

- **2020-2024**
- Conception et portage : Julie Claustre, Darwin Smith (LaMOP, Paris-VII)
- Partenaires : Archives nationales, Bibliothèque nationale, École nationale des chartes, bibliothèque Mazarine
- Financement : ANR ; 467 505 €.
- Corpus : registres capitulaires de Notre-Dame de Paris (XIVe siècle).
- Outils : eScriptorium
- Finalités : transcrire les registres capitulaires ; mettre au point des outils d'exploration et de fouille des données textuelles.
- Données et modèles : ex-nihilo.

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

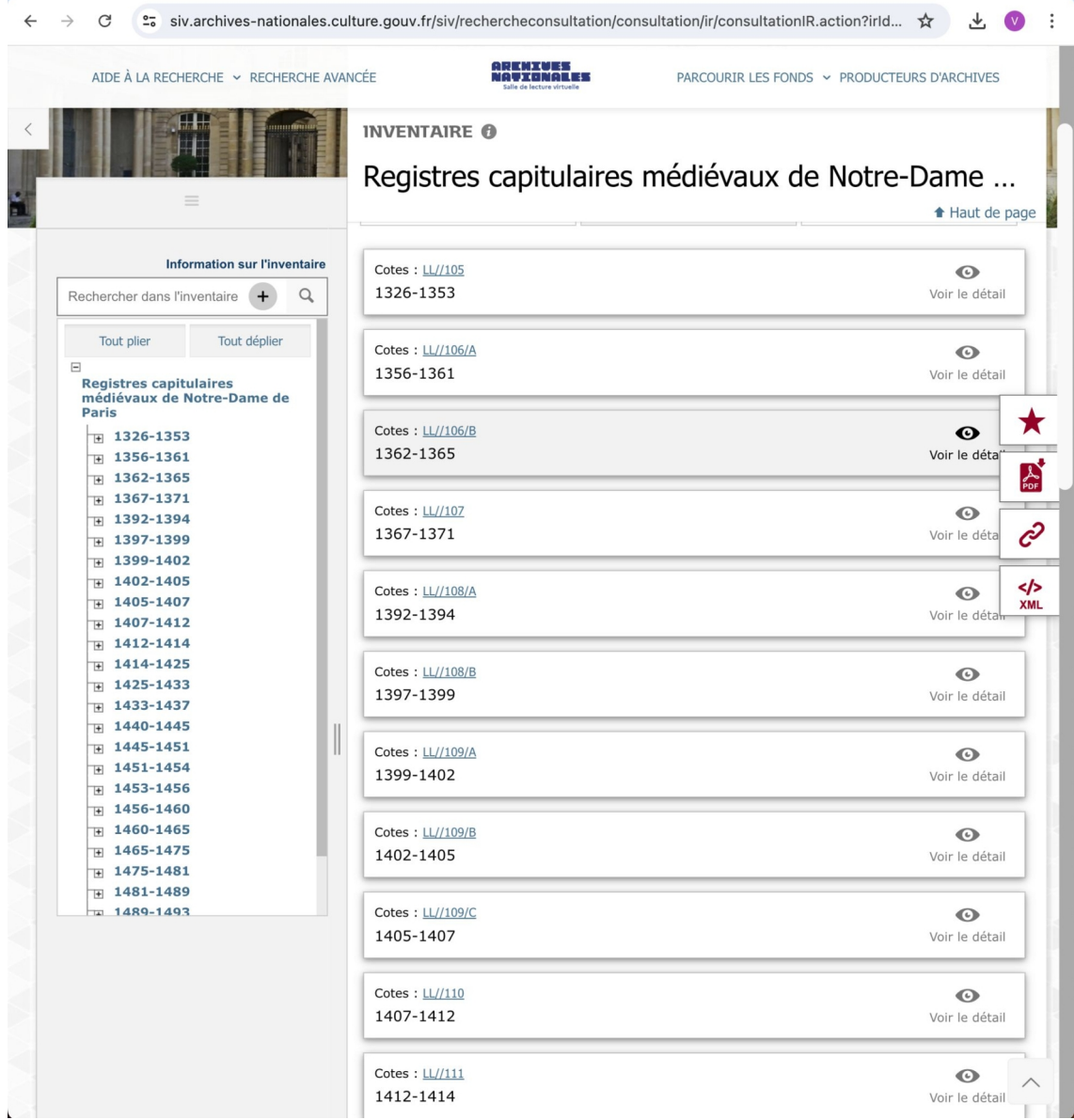
ARCHIVES NATIONALES

Les registres capitulaires (1325-1504)



MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

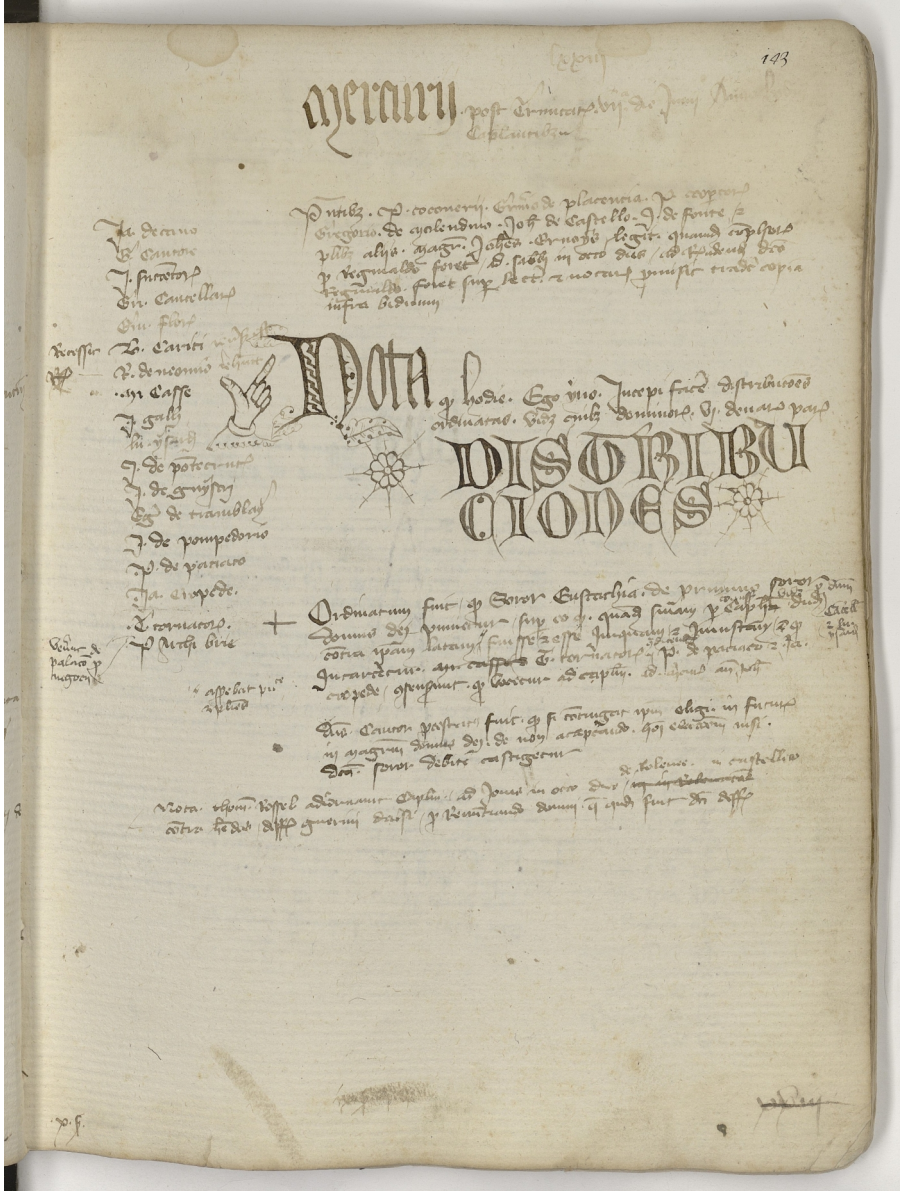
ARCHIVES NATIONALES



03/11/2026

Webinaire AIAF

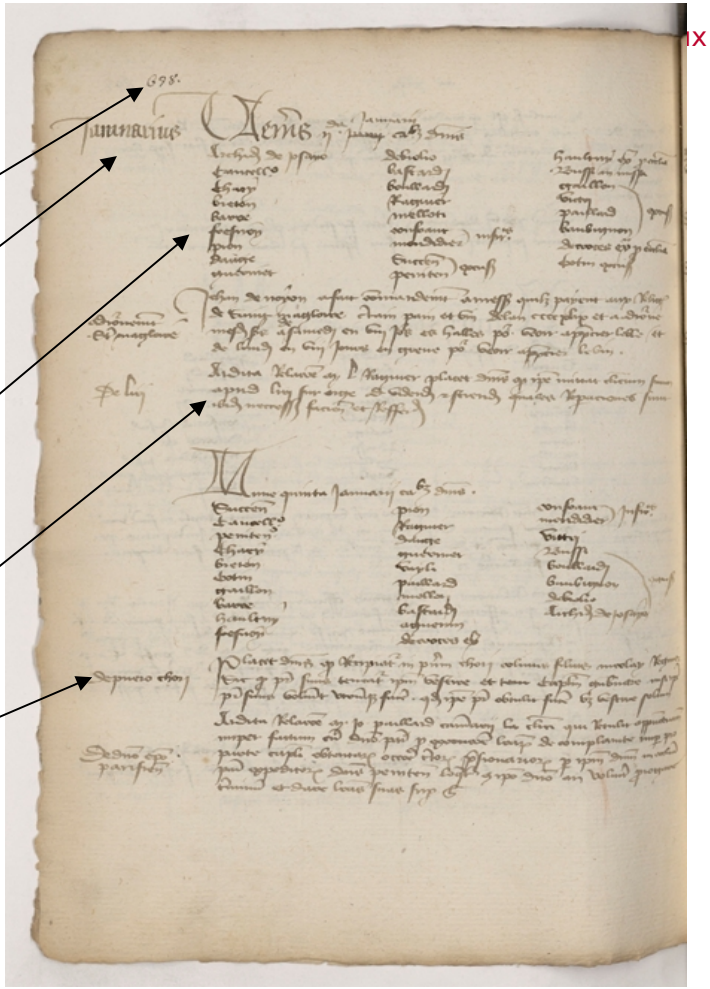
e-NDP



LL 107 p. 143

7 juin 1368

e-NDP



Pagination

Datation

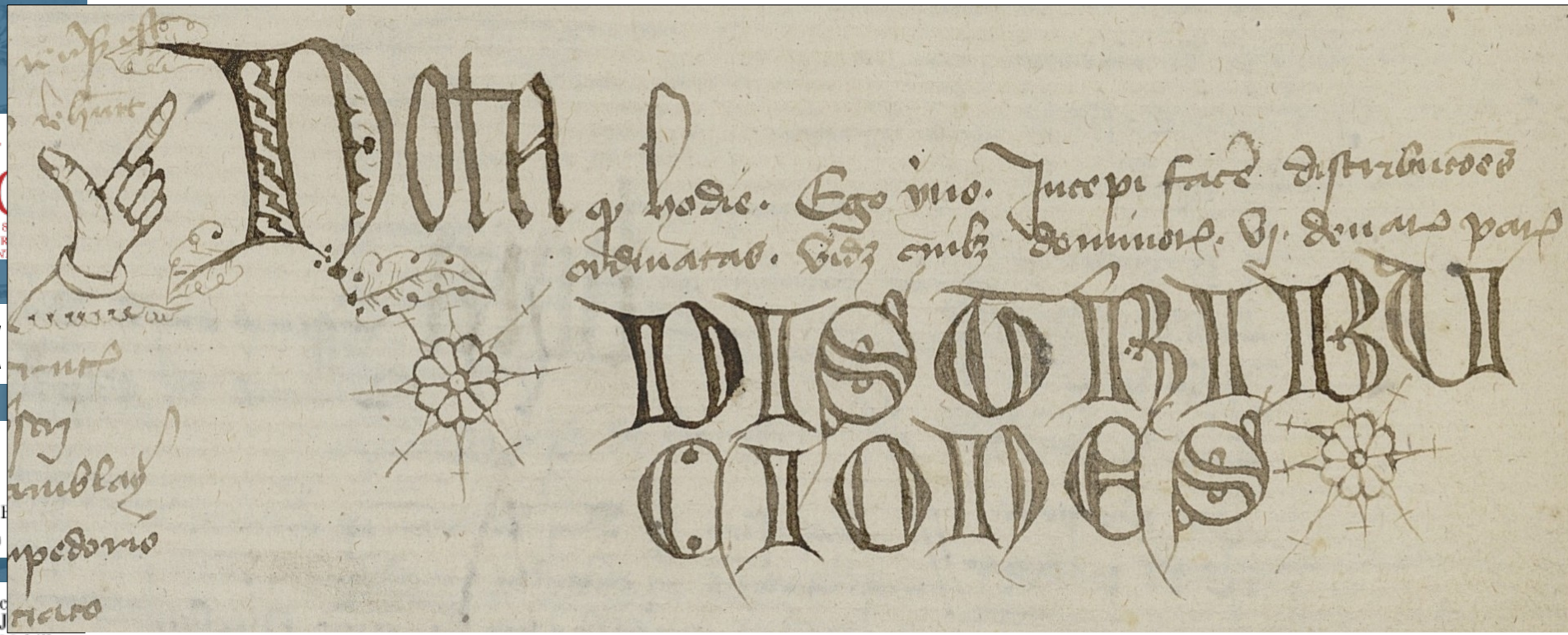
Liste des chanoines présents

conclusions

Index

➔**NOTA** quod **h**odie. **E**go yvo. **I**ncepi facere distribuciones ordinatas. videlicet cuilibet dominorum. VI. denarios parisienses

⊙ **DISTRIBUCIONES** ⊙



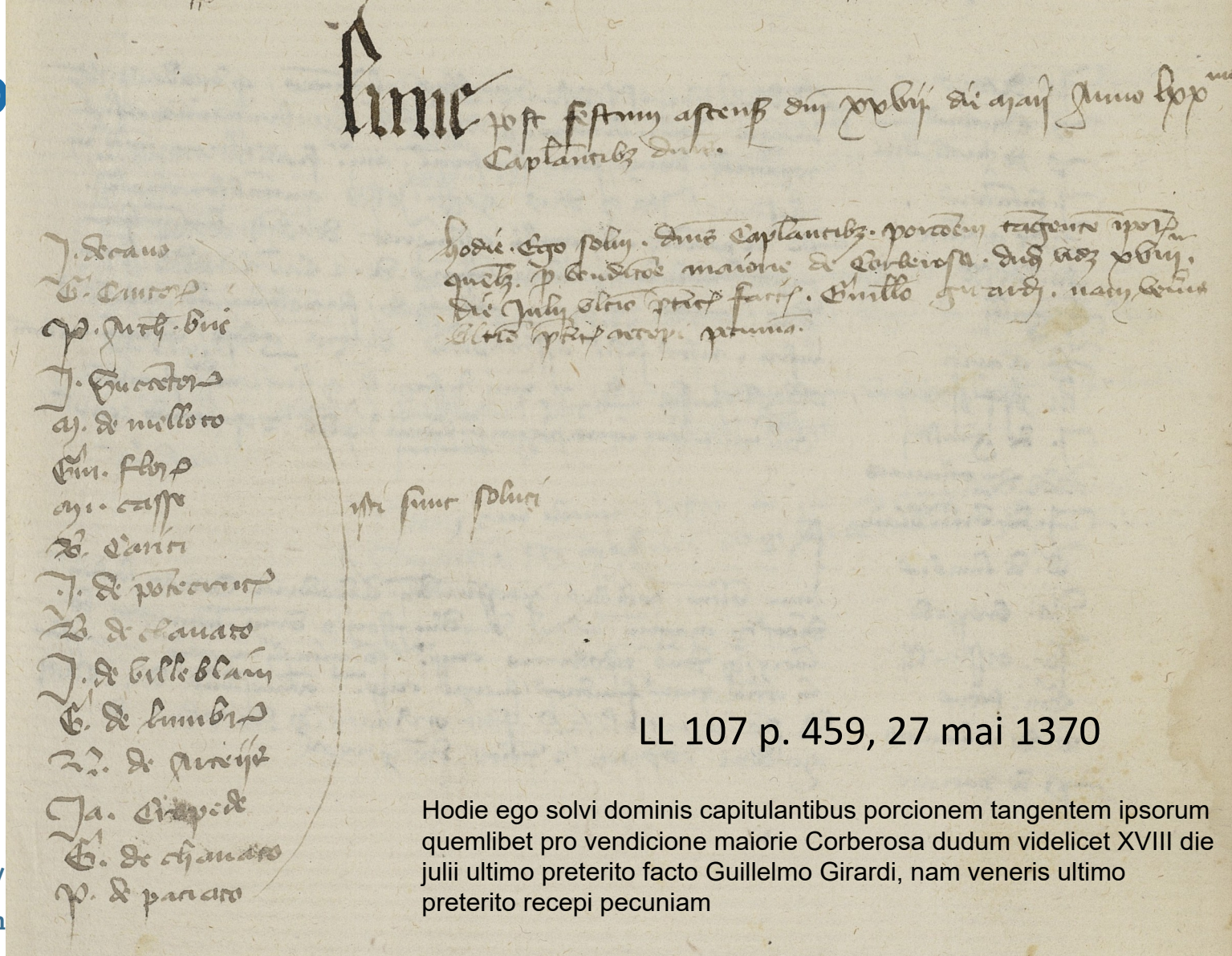
➔**NOTEZ BIEN** qu'**A**ujoud'hui, **M**oi yves, j'ai **C**ommencé à faire les distributions ordonnées, à savoir 6 deniers parisis à chacun de messeigneurs

! DISTRIBUTIONS !

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

ARCHIVES NATIONALES

e-ND



LL 107 p. 459, 27 mai 1370

Hodie ego solvi dominis capitulantibus porcionem tangentem ipsorum quemlibet pro vendicione maiorie Corberosa dudum videlicet XVIII die julii ultimo preterito facto Guillelmo Girardi, nam veneris ultimo preterito recepi pecuniam

IAMOP
UMR 8589
CNRS
UNIVERSITE PARIS 1 PANTHEON-SORBONNE

ARCHIVES NATIONALES

BnF

Bibliothèque Mazarine

centre Jean Mabillon
École nationale des chartes

03/11/

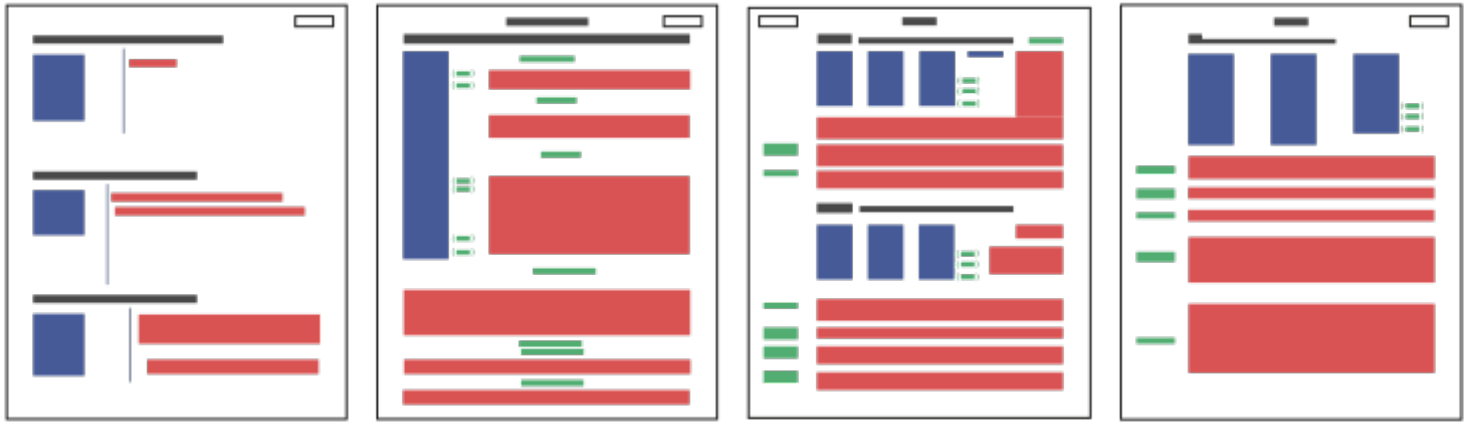
Webin

e-NDP

- Développements HTR : ENC / Sergio Torres
- Utilisation d'eScriptorium

e-NDP

mise en page des registres



LL 105
1326-1330 1346-1353

LL 108
1392-1399

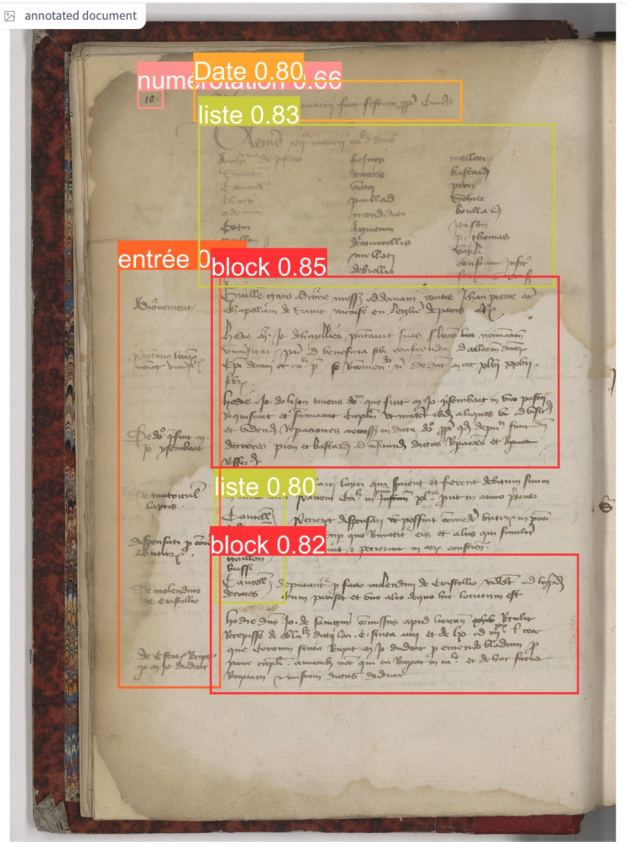
LL 112
1414-1424

LL 117
1451-1454

numéros de page
date
liste des capitulants
conclusions
entrées/compléments

Structure très stable sur la durée.

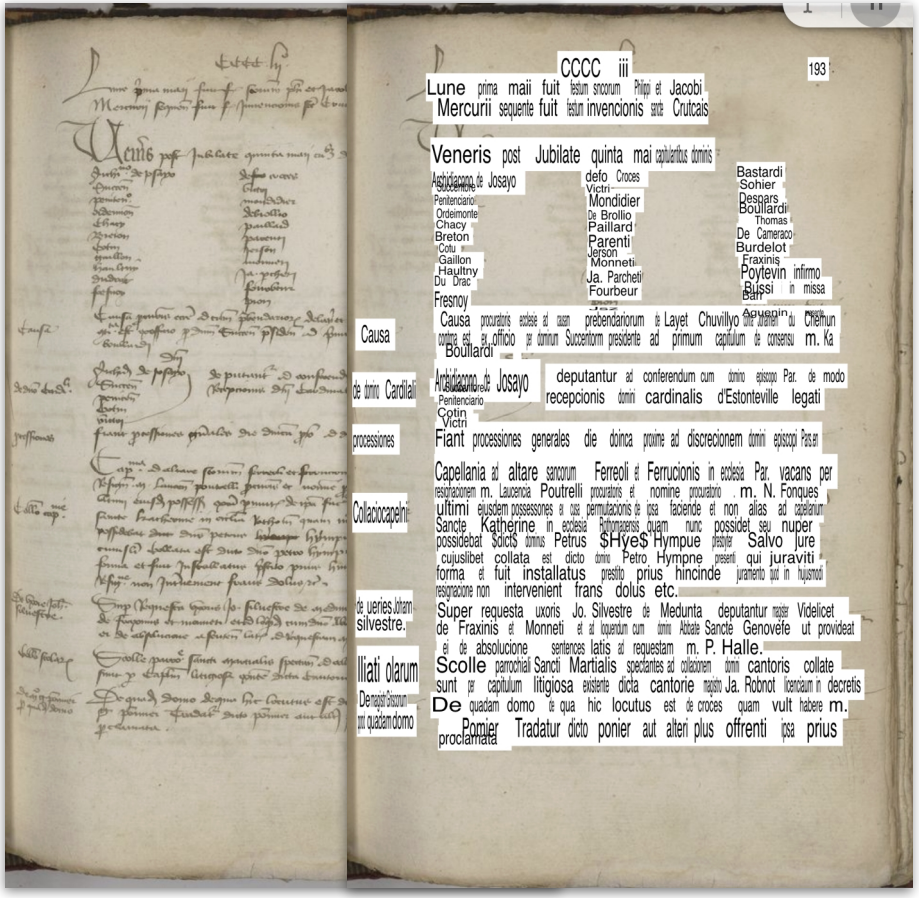
e-NDP



Les modèles de reconnaissance de la mise en page (*layout analysis*) dans e-NDP s'appuie sur cette typologie historique.

Ici un exemple de prédiction avec un modèle de segmentation Yolov5. Test ici

Phase 1 : acquisition du texte (modèle extractif)



MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

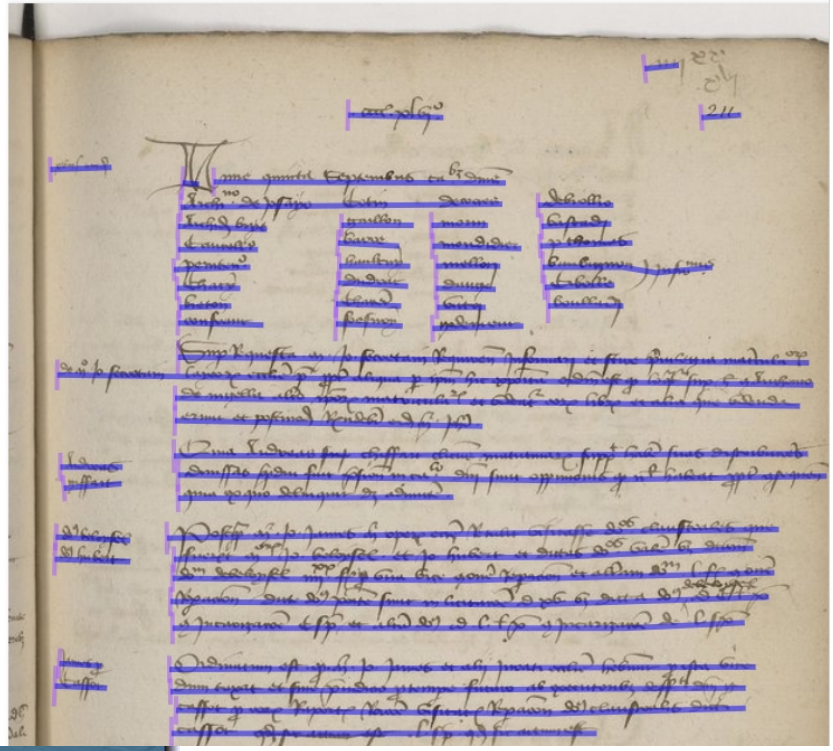
ARCHIVES NATIONALES

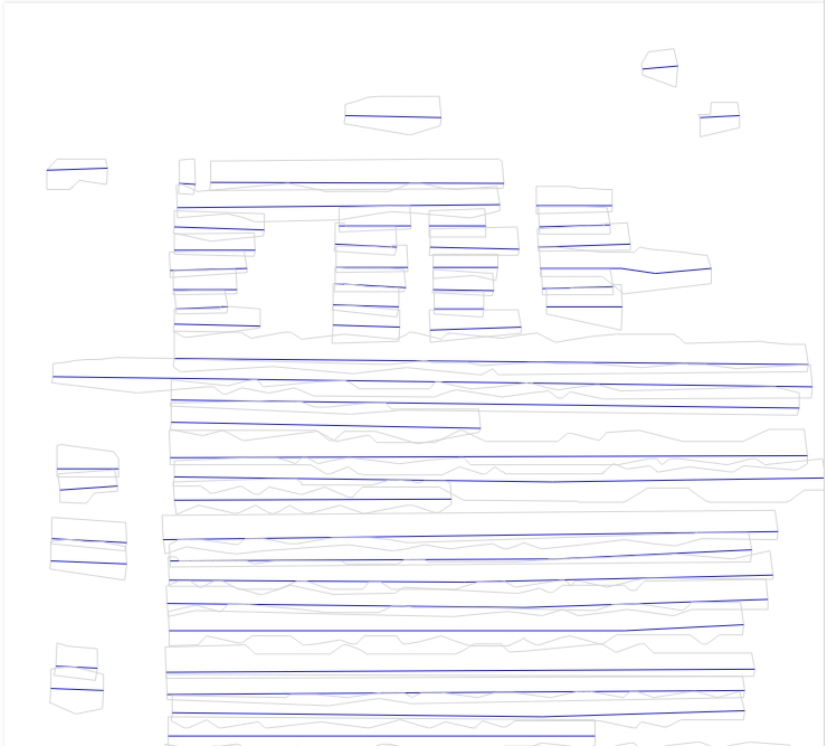
e-NDP

eScriptorium Home Contact All | Search in LL_116 My Projects My Models Hello Sebastien

Description Ontology Images Edit Models Reports LL_116
Element 218 - FRAN_0393_04567_L.jpg - (2538x3846) - 1.33 MB

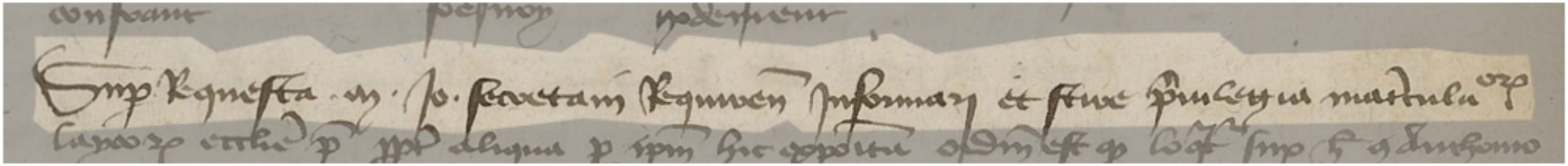
kraken:medieval_early_modern



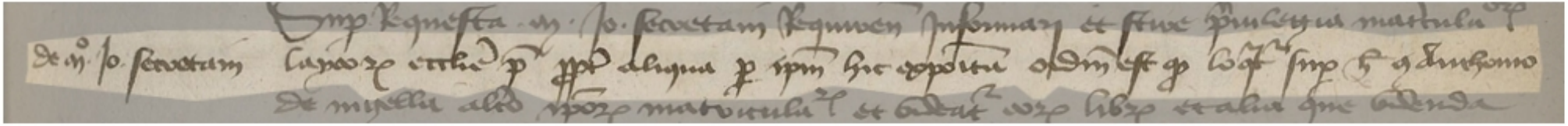


e-NDP

Line #33



Line #34



MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

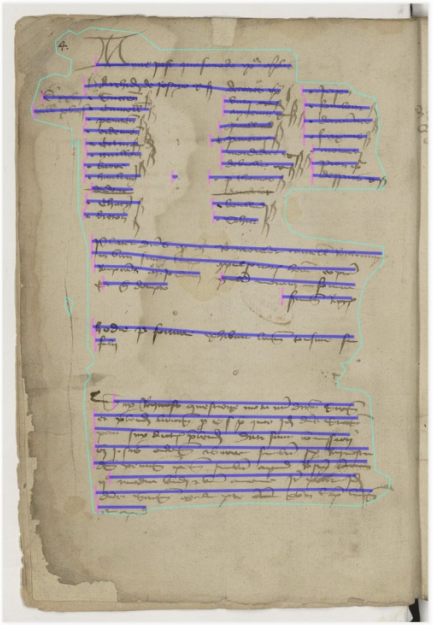
ARCHIVES NATIONALES

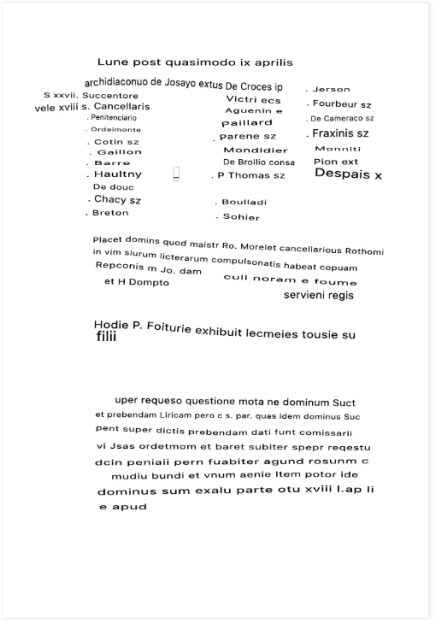
eScriptorium Home Contact | AI | Search in LL_118 | My Projects | My Models | Hello Lucas

Description | Ontology | Images | Edit | Models | Reports | LL_118

Element 8 - FRAN_0393_10996_L.jpg - (1869x2718) - 663.11 KB

kraken:medieval_early_modern





1	Lune post quasimodo ix aprilis
2	S xxvii. Succentore
3	archidiaconuo de Josayo extus De Croces ip
4	. Jerson
5	vele xviii s. Cancellaris
6	Victri ecs
7	. Penitenciario
8	Aguenin e
9	. Fourbeur sz
10	. Ordeimonte
11	. Parene sz
12	Paillard
13	. De Cameraco sz
14	. Cotin sz
15	. Fraxinis sz
16	. Gaillon
17	. Barre
18	Mondidier
19	Monniti
20	. Haultny
21	Pion ext
22	De douc
23	
24	. P Thomas sz
25	De Brollio consa
26	Despais x

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

ARCHIVES NATIONALES

eNDP | REGISTRES | PERSONNES | RESSOURCES

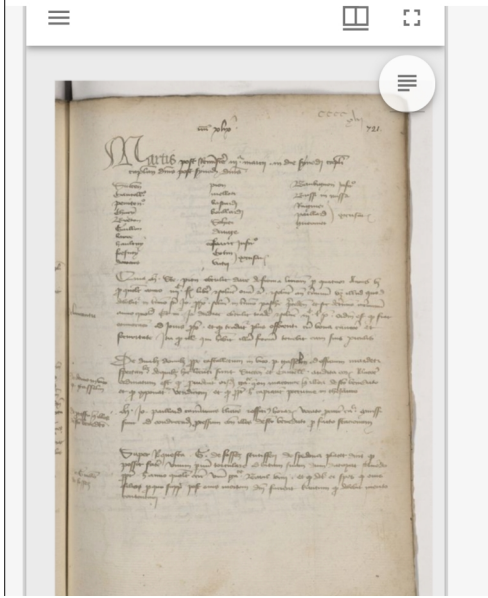
eNDP

Registres

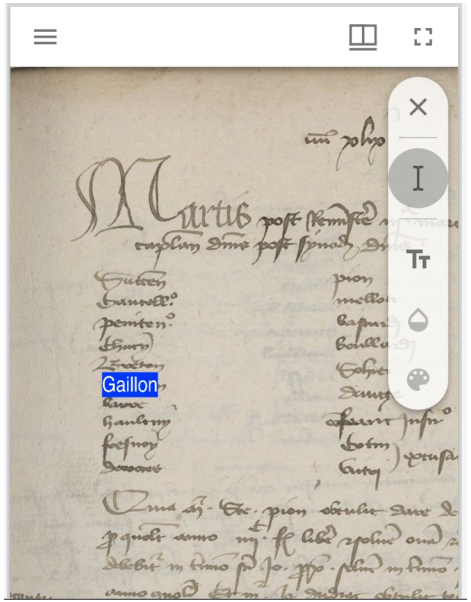
Fac-similé du registre de conclusions capitulaires LL116 de Notre-Dame de Paris - Mars 1450

1504

26 *Registres disponibles*







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

L. Terriel
(École des chartes)

ARCHIVES NATIONALES
Paris – Pierrefitte-sur-Seine



Un exemple suisse : les procès-verbaux du Conseil fédéral

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

ARCHIVES NATIONALES

Recherche simple | Recherche par date | Information

Procès-verbaux du Conseil fédéral (1848-1963)

La page *Procès-verbaux du Conseil fédéral (1848-1963)* vous permet d'effectuer une recherche par mot-clé dans les procès-verbaux manuscrits et dactylographiés du Conseil fédéral.

Saisir le terme de recherche

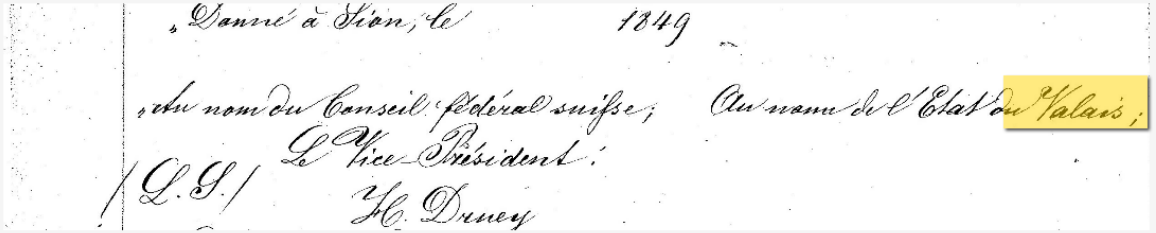
Valais

Période

1848 — 1872

Correspondances trouvées dans 63. documents trouvés dans 1626ms.

Beschlussprotokoll(-e) 24.10.1849



1849 retu nom die Conseil fédéral suisse, An mann des Etat der Valais; C Vice-Präsident. .G.)

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Schweizerisches Bundesarchiv BAR
Archives fédérales suisses AFS
Archivio federale svizzero AFS

E1004.1#1000/9#1325*

Rechercher

Chercher un mot

3 sur 9

payement de la quote hart de ce dernier Stat aus frai

2 2

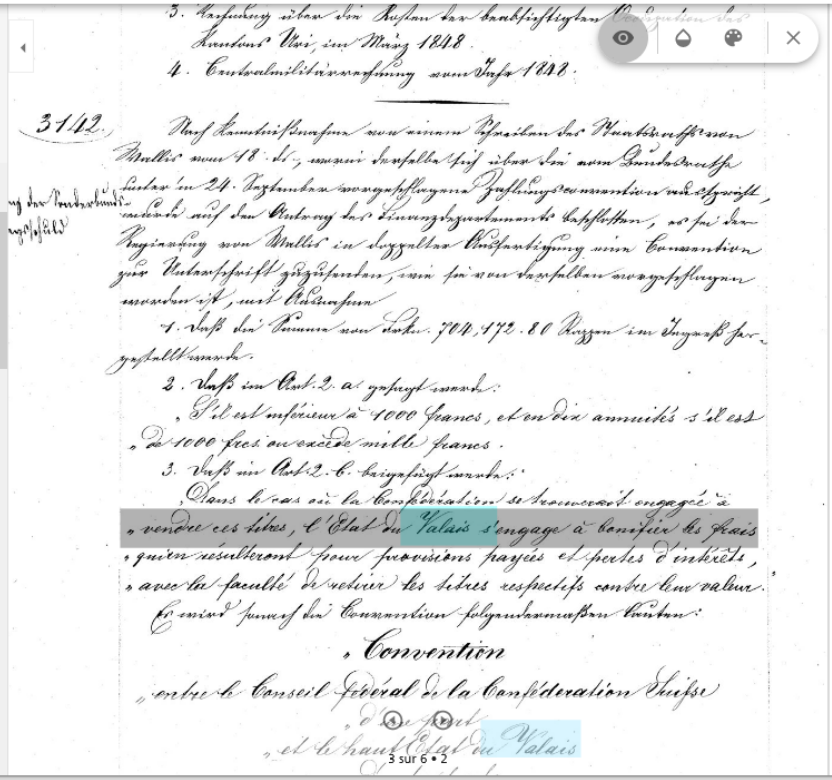
entrete Conseil fédéral de la Confederation Suisse d une hart et behaut Estat der **Valais** d autre part, concernant le payement de la quote hart de ce dernier Stat aus frai

3 2

ügt werde. Dans le cas ou la Confederation se trouverait engagée a „vendre ces titres, l Etat de **Valais** s'engage a bonifier des frais quien resulteront pour provisions payées et pertes d interst avec

plus...

4 3



Abtragung der Sonderen unde auf den Antrag des Finanzdepartements beschlossen, es sei der

Regierung von Wallis in doppelter Ausfertigung eine Convention

zur Unterschrift zuzusenden, wie sie von derselben vorgeschlagen

worden ist, mit Ausnahme

s

1. Daß die Summe von Frkn. 704, 172. 80 Rappen im Ingreß her¬-

gestellt werde.

2. Daß im Art. 2. a. gesagt werde:

Sil est inférieur à 1000 Francs, et en die ammités s il est

de 1000 fres ou exelde mille francs

3. Daß im Art. 2. b. beigefügt werde.

Dans le cas ou la Confederation se trouverait engagée a

„vendre ces titres, l Etat de Valais s'engage a bonifier des frais

quien resulteront pour provisions payées et pertes d interst

avec le faculte dereterer les titres respectifs contre leur valeur.

Es wird sonach die Convention folgendermaßen Bauten:

Convention

entrete Conseil fédéral de la Confederation Suisse

d une hart

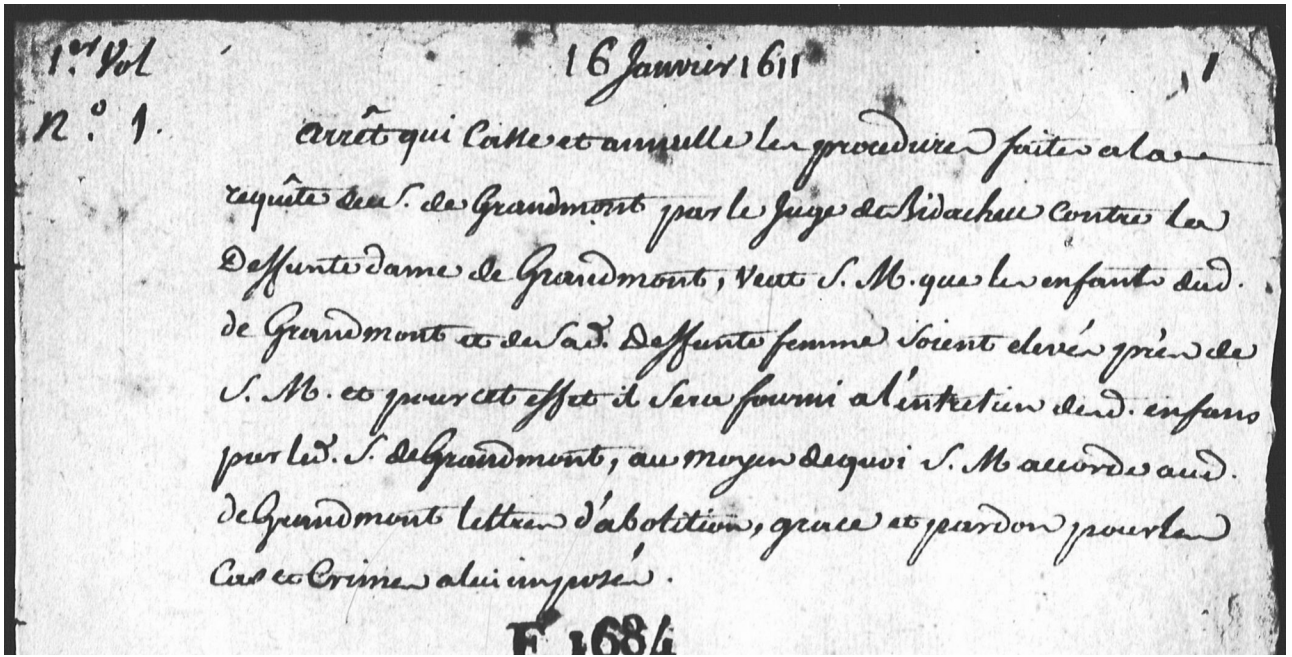
ARCHIVES NATIONALES
Paris – Pierrefitte-sur-Seine



Un projet autour des inventaires anciens des Archives nationales : SIMARA

Le cas des inventaires anciens

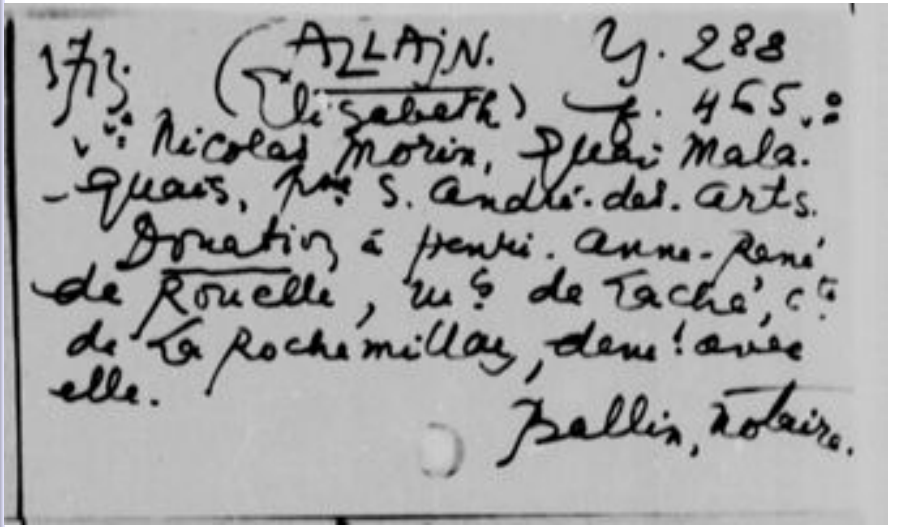
- Des inventaires manuscrits du XVIII[e] au XX[e] siècle pas encore « rétroconvertis » (seulement accessibles en mode image).

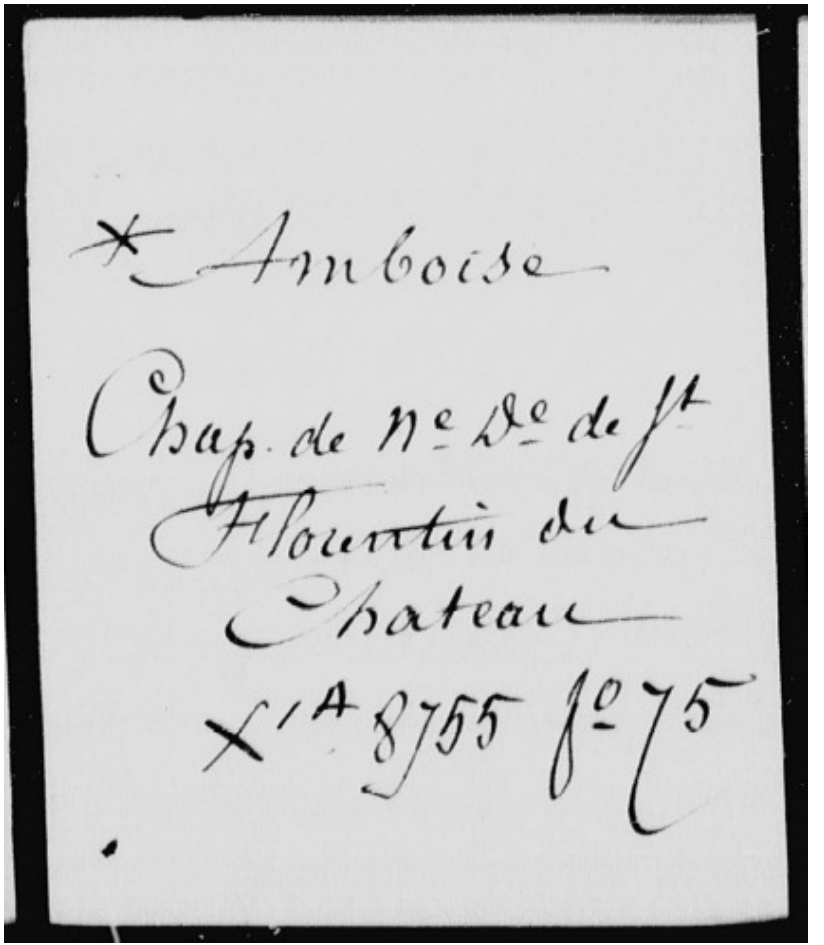

1.er Vol
16 Janvier 1611
N.° 1.
Arrêt qui Casse et annulle les procedures faites a la requête de S.r de Grandmont par le Juge de Sidaches Contre les Deffuntes dame de Grandmont, Veuve S.r M.e que les enfants dud. de Grandmont et dud. Deffunte femme soient elevés près de S.r M.e et pour cet effet il Sera fourni a l'entretien dud. enfans par les H.s S.rs de Grandmont, au moyen dequoi S.r M.e accorde aud. de Grandmont lettres d'abolition, grace et pardon pour les Cas et Crimes a lui imputés.
E 1684

Série E : arrêts en commandement du Conseil (XVIII[e] s.)

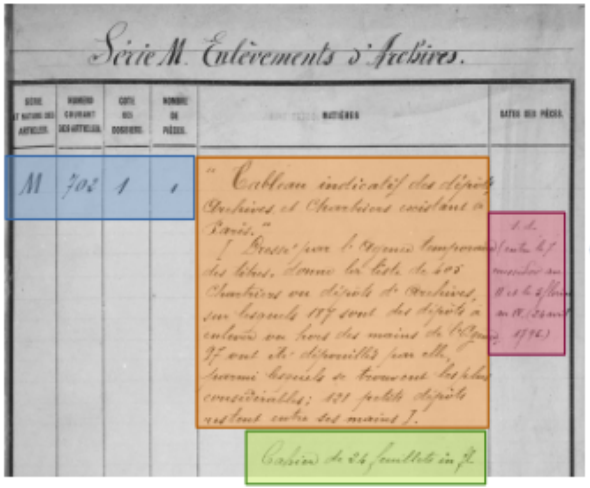
Le cas des inventaires anciens

- Seules clés d'accès aux fonds, mais encore inaccessibles aux publics (non présents en salle de lecture ou seulement sous forme de microfiches).

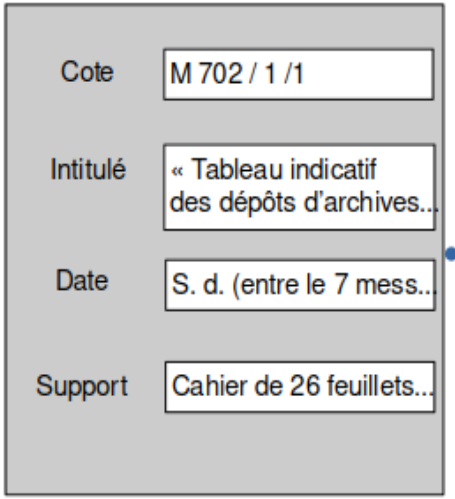




Postulats de départ



1. Segmentation : détection et typage des blocs de texte (les même catégories d'information sont en règle générale toujours aux mêmes emplacements).



2. Transcription dans l'interface SIMARA. Le texte reconnu est transcrit dans les champs d'informations définis par le responsable de fonds (adéquation entre les types de blocs de texte segmentés et le champ intellectuel).

```
<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
>
```

3. Génération d'un fichier en XML par la plate-forme (les champs correspondent à des éléments XML EAD).

\+ 4. Possibilité d'indexation à partir des référentiels AN.

Publication Prodoc/SLV

Relecture d'une transcription automatisée

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

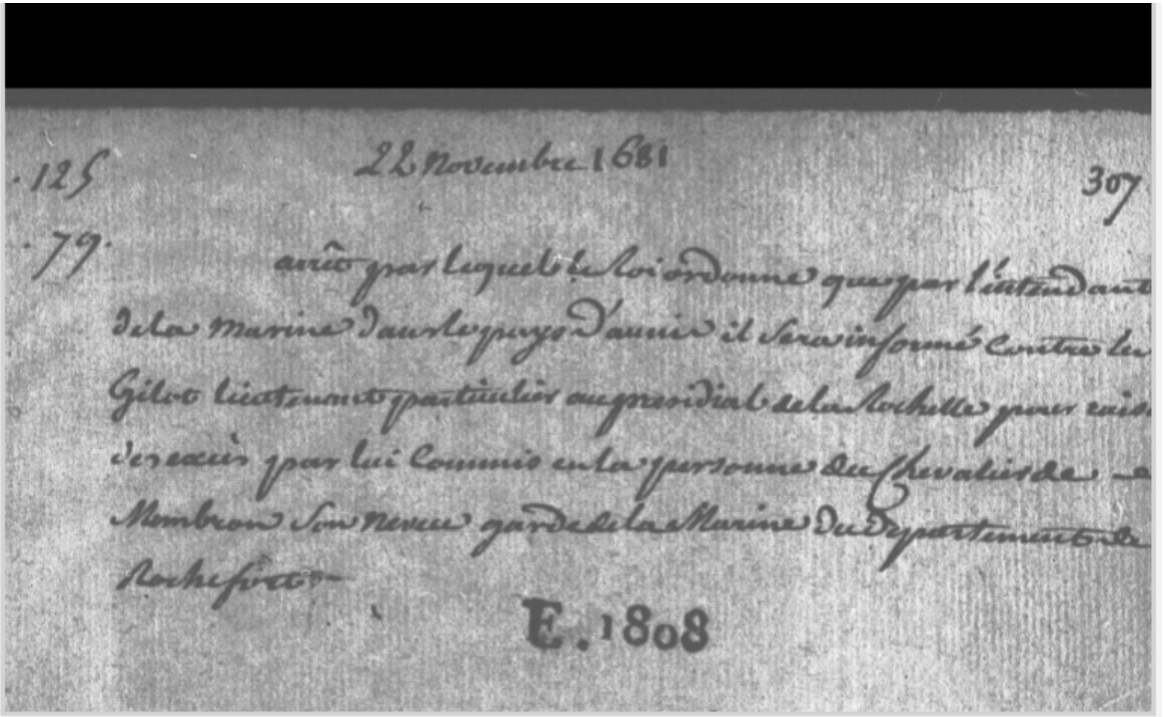
307

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



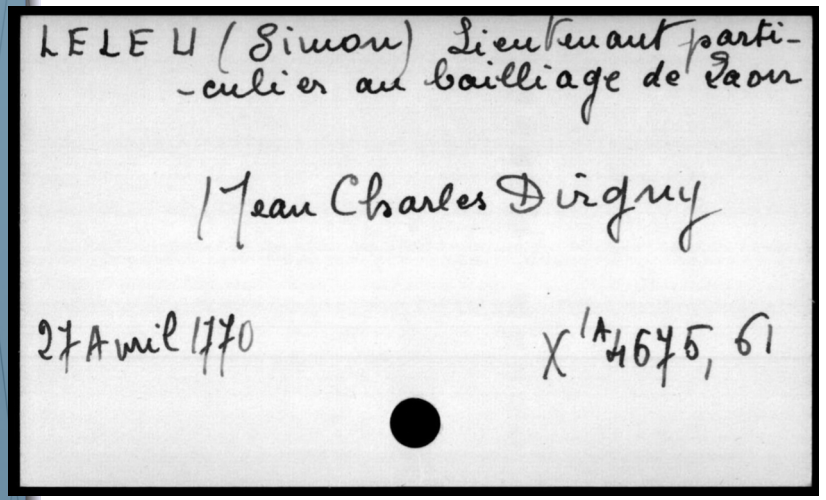
Transcription automatisée à corriger.



Annotateur

Changement d'approche

Processus d'HTR + NER **sans segmentation** (modèle génératif DAN – *Document Attention Network*)



Created by **Simara (HTR + NER)**

Simara (HTR + NER) da92…

100%

INTITULE LELEU (Simon) Lieutenant particulier au bailliage de Laon ANALYSE_COMPL Contre Jean Charles Durgny DATE 27 avril 1770 COTE_SERIE X1A COTE_ARTICLE 4675 PRECISIONS_SUR_COTE 61



ARCHIVES NATIONALES
Paris – Pierrefitte-sur-Seine

Un exemple à l'échelle du réseau des archives

SocFace

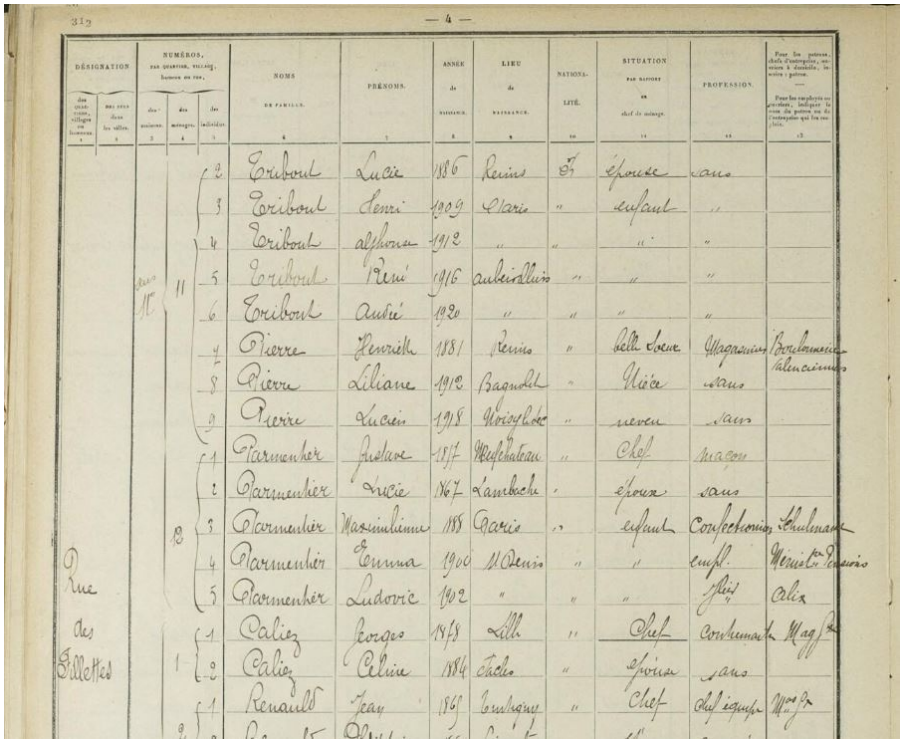
Les archives replacées au cœur de la recherche



SOCFACE : conversion en base de données des recensements de population.

Le projet SocFace

- Objectif : analyse de la demographie du XIXe siècle (1836-1936).
- Utilisation des recensements de population répartis dans les 100 services d'archives départementales.



https://socface.site.ined.fr/

Le projet SocFace

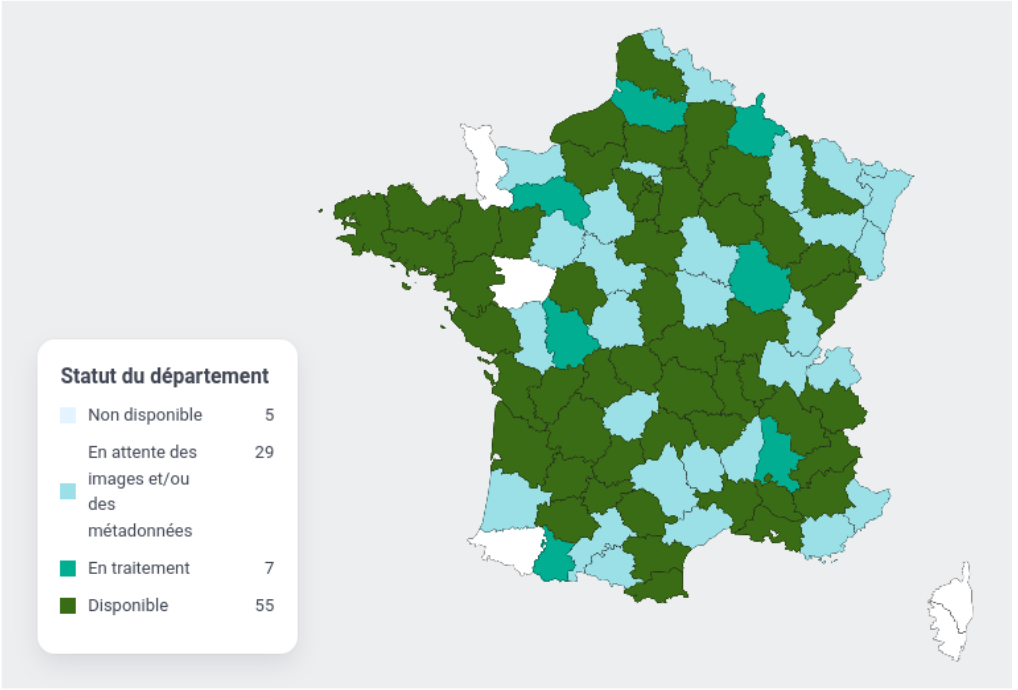
Départements
96

Communes
22 168

Pages
13 841 592

Ménages
75 684 297

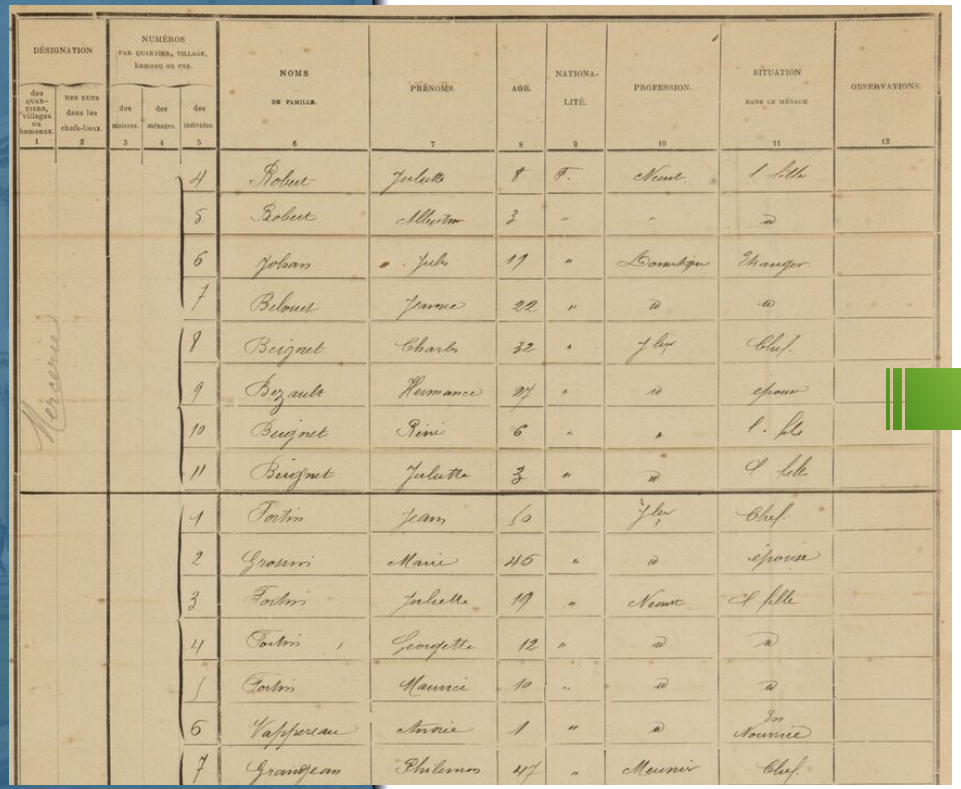
Mentions d'individus
306 145 801



- Approximativement :
  - 12.000.000 d'images
  - 580.000.000 noms de personnes

https://socface.teklia.com/

Le projet SocFace

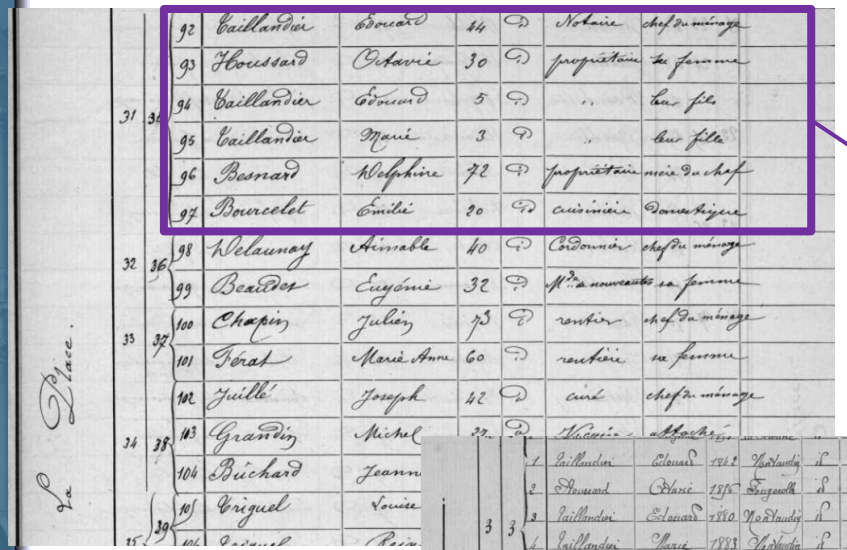


surname Robert firstname Juliette occupation néant link l. bille age 8 nationality française
surname Robert firstname Albertine occupation idem link idem age 3 nationality idem
surname Johan firstname Jules occupation domestique link chauger age 19 nationality idem
surname Belouet firstname Jeanne occupation idem link idem age 22 nationality idem
surname Beignet firstname Charles occupation j les link chef age 32 nationality idem
surname Bezault firstname Hermance occupation idem link épouse age 27 nationality idem
surname Bugnet firstname René occupation idem link l. fils age 6 nationality idem
surname Biignet firstname Juliette occupation idem link l. bille age 3 nationality idem
surname Portin firstname Jean occupation j leer link chef age 50 nationality idem
surname Gronni firstname Marie occupation idem link épouse age 46 nationality idem
surname Fortin firstname Juliette occupation néant link sa fille age 19 nationality idem
surname Portin firstname Georgette occupation idem link idem age 12 nationality idem
surname Portin firstname Maurice occupation idem link idem age 10 nationality idem
surname Vappereau firstname Andrée occupation idem link idem age 1 nationality idem
surname Grangeau firstname Philomèn occupation meunier link chef age 47 nationality idem

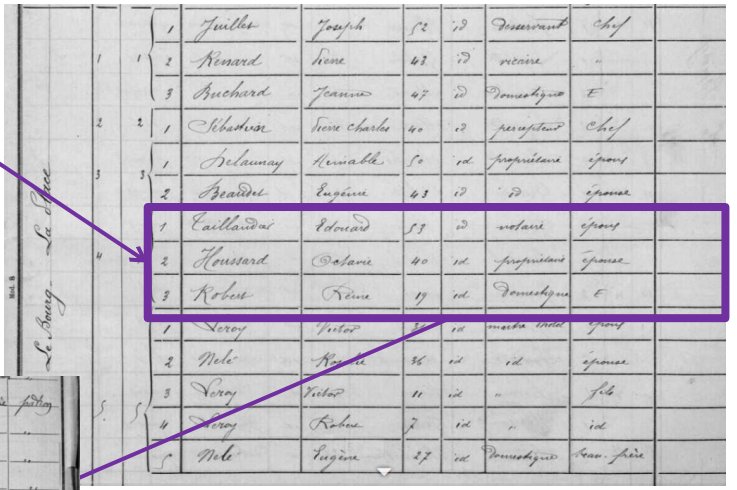
HTR+NER

Le projet SocFace

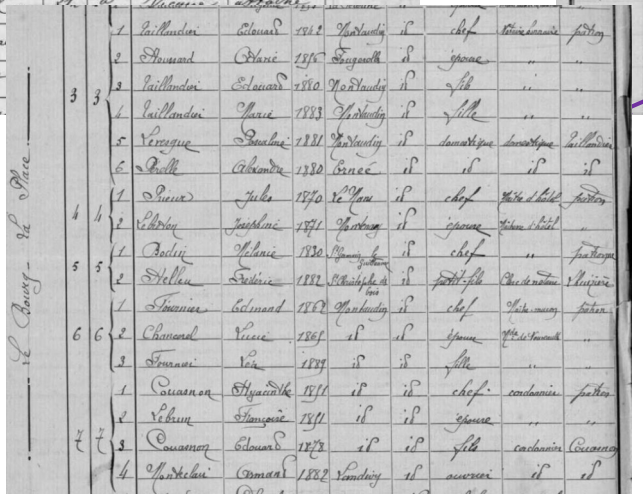
1886



1896



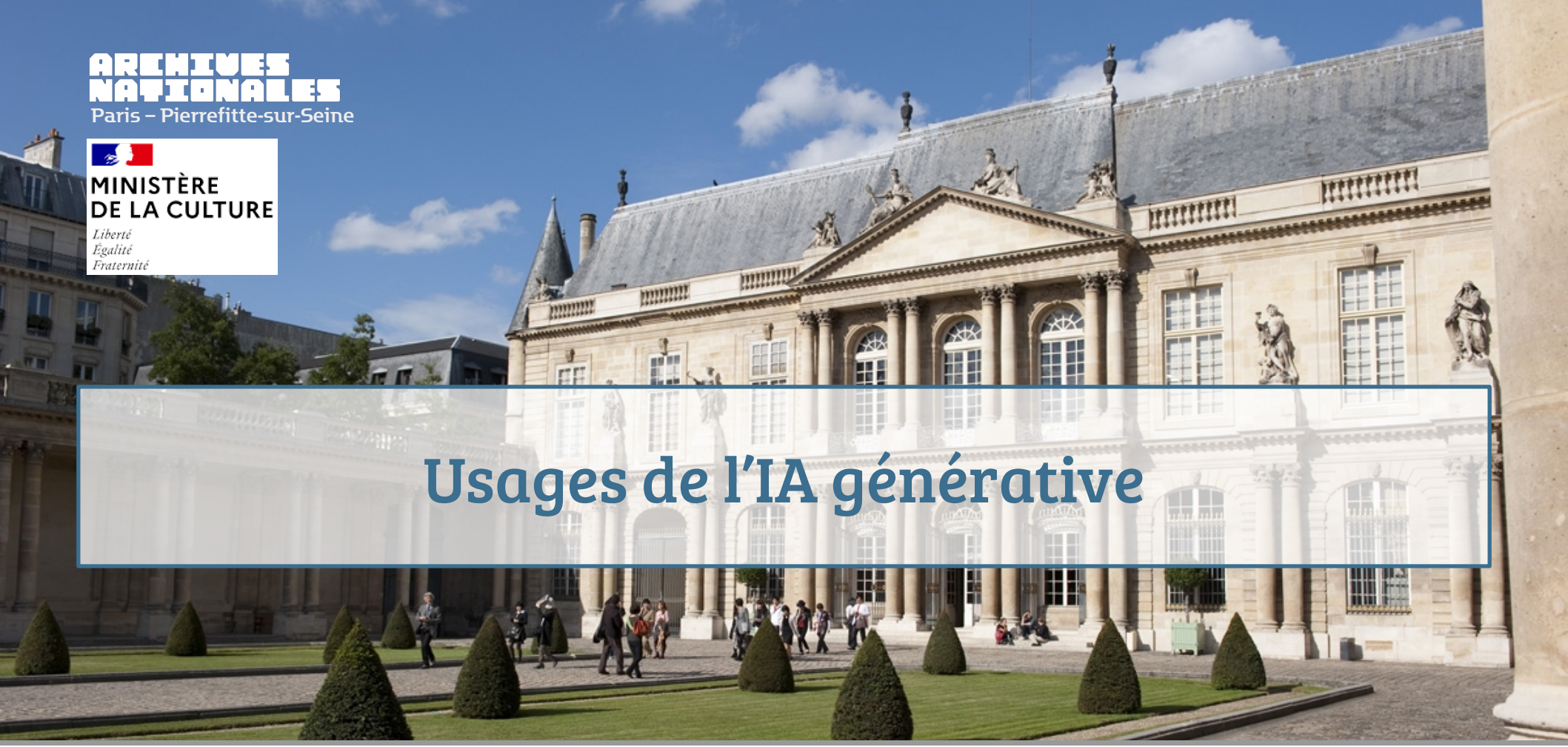
1906



Relier les personnes entre les recensements.

ARCHIVES NATIONALES
Paris – Pierrefitte-sur-Seine

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

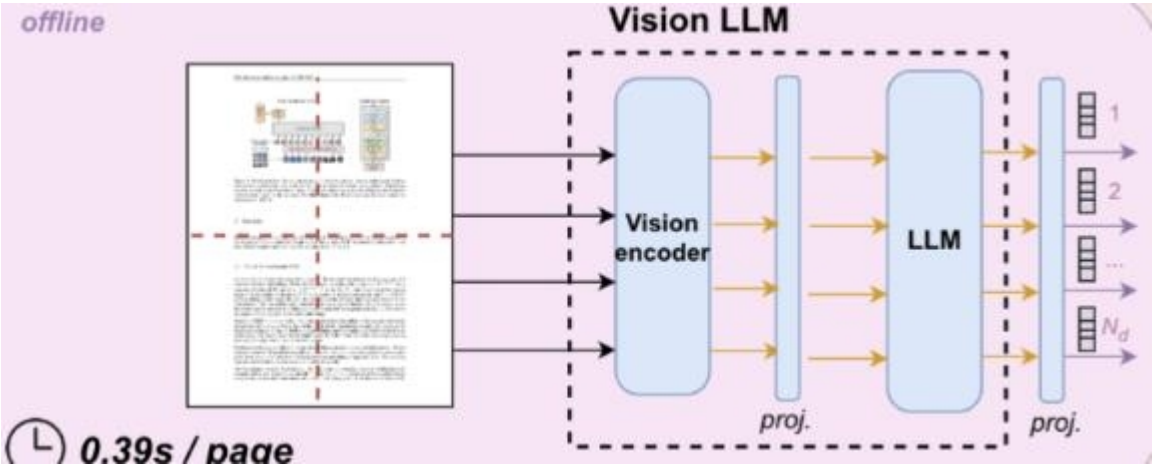


Usages de l'IA générative

L'usage de l'IA générative

IA générative / LLM

(dep. 2022)



- Création d'une vision d'ensemble d'une donnée (ici, une page) => *embedding*.
- Génération d'une réponse en langage humain en réponse à un prompt.

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

ARCHIVES NATIONALES

PictorIA

pictoria
CONSORTIUM HUMA-NUM

PROGRAMME D'IDENTIFICATION, CLASSIFICATION, TRAITEMENT D'IMAGES, OBSERVATION ET RECONNAISSANCE DES FORMES PAR L'INTELLIGENCE ARTIFICIELLE

PARTENARIATS ET COLLABORATIONS

AN – Archives nationales
https://www.archives-nationales.culture.gouv.fr/

AOrOc – Archéologie & Philologie d'Orient et d'Occident
https://archeo.ens.fr/

APP – Université Rennes 2, laboratoire Arts Pratiques et Poétiques
https://sites-recherche.univ-rennes2.fr/arts-pratiques-poetiques/

ArScan – Archéologies et Sciences de l'Antiquité
http://www.arscan.fr/

Ariane – Analyses, Recherches, Intelligence Artificielle et Nouvelles Editions numériques
https://cst-ariane.huma-num.fr/

BHVP – Bibliothèque historique de la Ville de Paris
https://www.paris.fr/lieux/bibliotheque-historique-de-la-ville-de-paris-bhvp-16

Biblissima+ – Observatoire des cultures écrites anciennes, de l'argile à l'imprimé
https://projet.biblissima.fr/fr

BnF – Bibliothèque nationale de France
https://www.bnf.fr/fr

CANAL – Cultural Analytics Lab
http://canal-lab.uva.nl/

03/11/2026

Webinaire AIAF

Autre exemple

- Consortium Pictor'IA (Nanterre), dont les AN sont partenaires.
- Utilisation de l'IA pour traiter les documents iconographiques (ex. d'Hikaria/Guimet).



LLM
Prompt : décrire une photographie

Cette photographie met en scène un groupe de huit hommes japonais en tenue traditionnelle, probablement des samouraïs, reconnaissables à leur port de sabres (katana et wakizashi) et à
leurs coiffures caractéristiques. …

PictorIA

Consortium regroupant laboratoires, institutions culturelles, sociétés privées pour développer des solutions autour des documents iconographiques.



numer	legende_transcrite	legende_relue_corrigée_manuellement_FH	description_visuelle
1	Route détruite	Manque titre planche : Corse	Un pont en bois enjambant un cours d'eau asséché ou peu profond, avec une route endommagée. Le paysage est montagneux avec de la végétation et des rochers.
2	Pâturage incendié après incendie	Pâturage excessif après incendie	Un vaste pâturage ou une zone de végétation basse, montrant des signes d'un incendie récent, avec des montagnes rocheuses en arrière-plan sous un ciel nuageux.

ARCHIVES NATIONALES
Pierrefitte-sur-Seine – Paris

MINISTÈRE
DE LA CULTURE
Liberté
Égalité
Fraternité

L'IA générative et la description des archives

Des pistes prometteuses : l'exemple des archives numériques natives – le projet InventAIre

Mémoire de Master 2 TNAH (École des chartes)



Sarah Marcq (Prix AAF 2025)

ÉCOLE NATIONALE DES CHARTES
UNIVERSITÉ PARIS, SCIENCES & LETTRES

Sarah Marcq
licenciée ès histoire et histoire de l'art

DES USAGES ARCHIVISTIQUES POUR L'INTELLIGENCE ARTIFICIELLE

LE CAS DE L'AUTOMATISATION DE L'INVENTAIRE DES ARCHIVES À LA CHAMBRE DES DÉPUTÉS DU GRAND-DUCHÉ DE LUXEMBOURG

Mémoire pour le diplôme de master
« Technologies numériques appliquées à l'histoire »
2024

https://dumas.ccsd.cnrs.fr/dumas-04765262

Cas d'étude d'InventAIre

Archives de la Chambre des députés du Luxembourg

- Utilisation d'une IA générative pour le signalement et la description d'archives (description, tri, repérage des délais de communication)



Objectifs d'InventAIre

- Décrire des fonds
- Déterminer leurs délais de communicabilité
- Modèle : tableur Excel de 17 colonnes (ANL)

- Prototype pour automatiser la rédaction.

- Cote
- Localisation
- Identification de la série du plan de classement
- Code série du tableau de tri
- Titre
- Description
- Période de création : de
- Période de création : à
- Soumis au droit d'auteur
- Données à caractère personnel
- Acte d'état civil
- Acte notarié
- Atteinte aux relations extérieures, à la sécurité du Grand-Duché ou à l'ordre public
- Affaires portées devant les instances juridictionnelles, extrajudiciaires ou disciplinaires
- Prévention, recherche de faits punissables
- Données commerciales et industrielles
- Secret fiscal

Délais (loi de 2018)

Type de donnée	Délai de communicabilité
Données à caractère personnel	25 ans après le décès de la personne concernée ou 75 ans à compter de la date du document le plus récent inclus dans le dossier
Actes d'état civil	100 ans à partir de la date de l'acte
Actes notariés	75 ans à partir de la date de l'acte
Atteinte aux relations extérieures, à la sécurité du Grand-Duché ou à l'ordre public	50 ans à compter de la date du document le plus récent inclus dans le dossier
Affaires portées devant les instances juridictionnelles, extrajudiciaires ou disciplinaires	50 ans à compter de la date du document le plus récent inclus dans le dossier
Prévention, recherche de faits punissables	50 ans à compter de la date du document le plus récent inclus dans le dossier
Données commerciales et industrielles	50 ans à compter de la date du document le plus récent inclus dans le dossier
Secret fiscal	100 ans à compter de la date du document le plus récent inclus dans le dossier

Délais contraints et gestion de projet

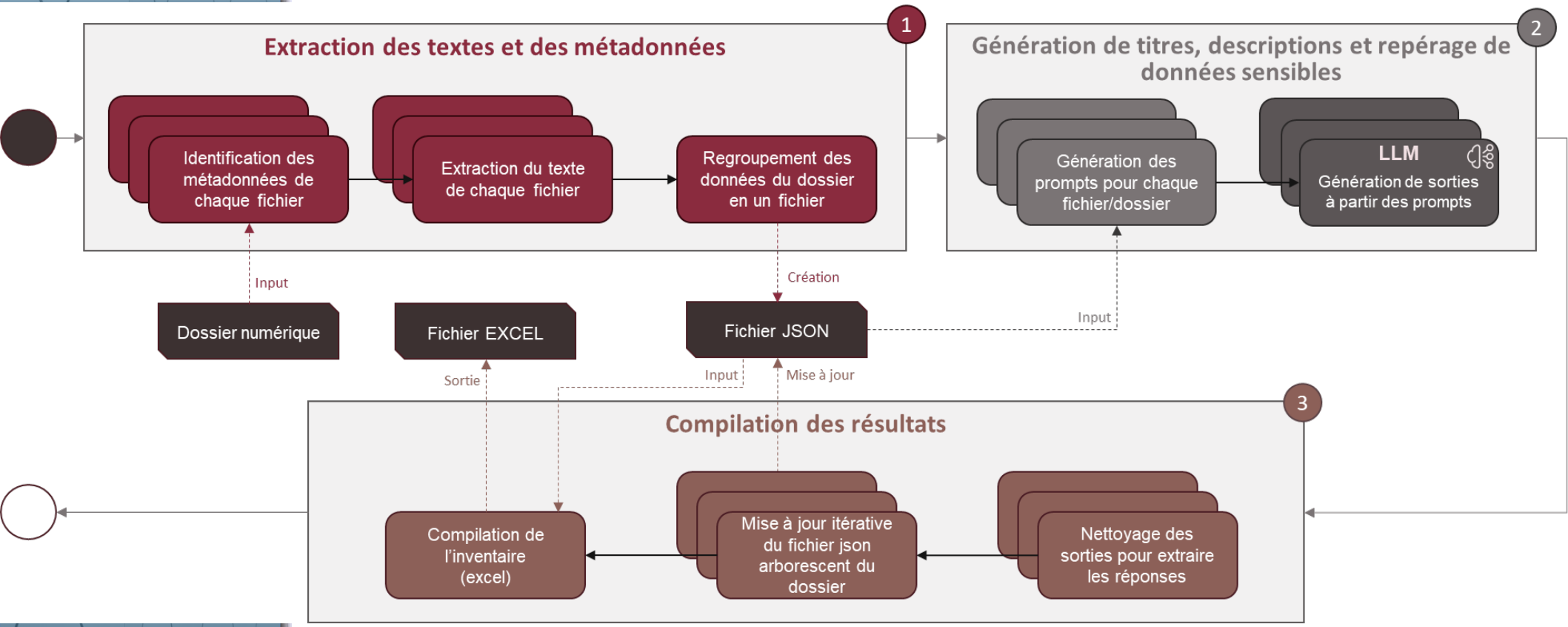
- Stage de 4 mois…
- Difficile d'annoter et de faire un apprentissage surpervisé.
- Choix d'un modèle pré-entraîné ou d'un apprentissage non supervisé.
- Mais l'investissement en temps pour évaluer les modèles est important.
- Investissement dans un ordinateur avec GPU et 64G de ram…

Tâche	**Estimation de temps**
Définition du périmètre et de l'approche	15%
Tests pour le choix d'un modèle	40%
Code sur l'intégration du modèle	10%
Prompt-engineering	15%
Évaluation de l'outil*	15%
Autre	5%

Questions éthiques et sécurité

- Pas de transmission des données à des tiers.
  - Entraînement en local
- Biais et hallucinations.
- Invisibilisation de données importantes mais non jugées comme telles.

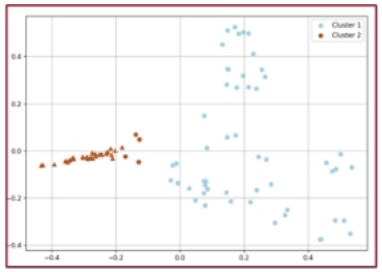
Traitements

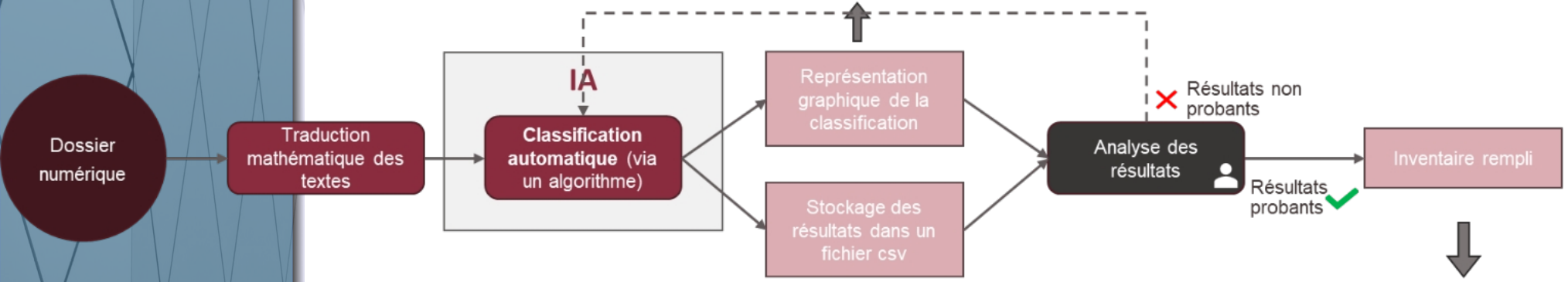


Traitements

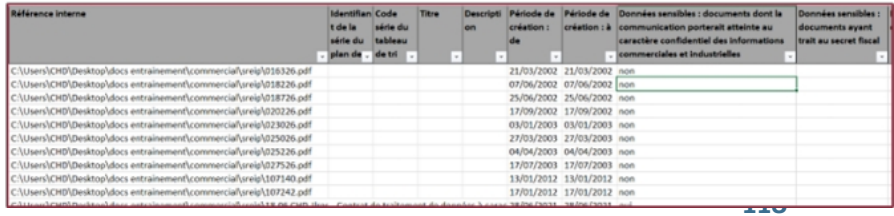
- Algorithmes de classification automatique (NLP/TAL)
  - Regroupement des documents par vectorisation (*embedding*)

Traitements





chemin	cluster	etiquette
C:\Users\CHD\Desktop\docs entrainement\commercial\sreip\016326.pdf	pas d'info commerciale	pas d'info commerciale
C:\Users\CHD\Desktop\docs entrainement\commercial\sreip\018226.pdf	pas d'info commerciale	pas d'info commerciale
C:\Users\CHD\Desktop\docs entrainement\commercial\sreip\018726.pdf	pas d'info commerciale	pas d'info commerciale
C:\Users\CHD\Desktop\docs entrainement\commercial\sreip\020226.pdf	pas d'info commerciale	pas d'info commerciale
C:\Users\CHD\Desktop\docs entrainement\commercial\sreip\023026.pdf	pas d'info commerciale	pas d'info commerciale
C:\Users\CHD\Desktop\docs entrainement\commercial\sreip\025026.pdf	pas d'info commerciale	pas d'info commerciale
C:\Users\CHD\Desktop\docs entrainement\commercial\sreip\025226.pdf	pas d'info commerciale	pas d'info commerciale



Traitements

- Algorithmes de classification automatique (NLP/TAL)
  - Regroupement des documents par vectorisation (*embedding*)
- Reconnaissance d'entités nommées
  - Identification de données à caractère personnel
  - Anonymisation (remplacement des noms, mails, téléphones, etc.)
- OCR/HTR

Utilisation d'un LLM

- Choix de Llama 3
- "Nous posions un prompt par colonne à remplir dans l'inventaire en fournissant du contexte et le texte d'un document ou d'une partie de document en entrée."



Pour 420 répertoires et documents…

Qualité de la réponse	Ensemble des colonnes	Titre et description	Données sensibles
Correcte	77%	71%	79%
Incomplète/sujette à débat	10%	22%	6%
Incorrecte	13%	6%	15%

Étape 1 : Extraction des textes et des métadonnées des documents

Exemple de fichier JSON arborescent pour un dossier contenant un document


```
{
   "path": "C:\User\Chemin\du\dossier_exemple",
   "folderName": "dossier_exemple",
   "type": "directory",
   "content": [
   {
       "path": "C:\User\Chemin\du\dossier_exemple\document.txt",
       "type": "file",
       "fileName": "document.txt",
       "creationDate": "Wed May 20 10:26:26 2024",
       "modificationDate": "Wed May 29 10:35:06 2024",
       "mimeType": "text/plain; charset=UTF-16LE",
       "texte": "Ceci est le texte d'un document txt.",
       "generatedTitle": "to be defined"
   }
   ],
   "generatedTitle": "to be defined"
}
```

Étape 2 : Génération de titres, descriptions et repérage de données sensibles

Exemple de prompt

Input :


```
{
    "path": "C:\User\Chemin\du\dossier_exemple\document.txt",
    "type": "file",
    "fileName": "document.txt",
    "creationDate": "Wed May 20 10:26:26 2024",
    "modificationDate": "Wed May 29 10:35:06 2024",
    "mimeType": "text/plain; charset=UTF-16LE",
    "texte": "Ceci est le texte d'un document txt.",
    "generatedTitle": "to be defined"
}
```


Context : You are an achivist at the Chambre des Députés of Luxembourg.

Question : You have been given a full document and some metadata formatted as json as input. You are working on an archival inventory project. Please generate one title (max 20 words) in french for this document for the inventory.
In your title, avoid the word 'fonds'. Avoid generic words such as 'divers' or 'variés'. Also avoid the word 'documents' without further precisions : add the types or names of the documents being exhaustive (all of the major types). It would be better if you could format it this way : 'subject, topic - action/type of document', only do it if there is a clear topic or subject and a clear action or only one clear types/names of documents. If there is a chronological indication in your title (not madatory, only if you are sure and if it is pertinent, must be with the topic if it is the date of the event that constitutes the topic) the format is day (number) month (letters, no abbreviation) year (number). If there are several parts in your title, separate them with commas. Be precise and exhaustive in your title (remember : topic/subject - action(s)/precise type(s) of documents if you can) but do not put information you are not sure about. If the document's text does not give you information, do not make it up. Make sure to mention the important people, organizations, locations, dates, ... if you identify some.
Examine well the whole content for this. Please only write the title in french, nothing else, so I can directly reuse it in my inventory. Do not explain your answer or tell me about alternatives in your answer. I only need the title. Do not invent information, only use the ones you have. It's ok if your answer is minimalist because you do not have enough info. Be careful, sometimes a file contains several types of document.
Make sure the syntax and spelling in french in your answer are correct.
title in french :

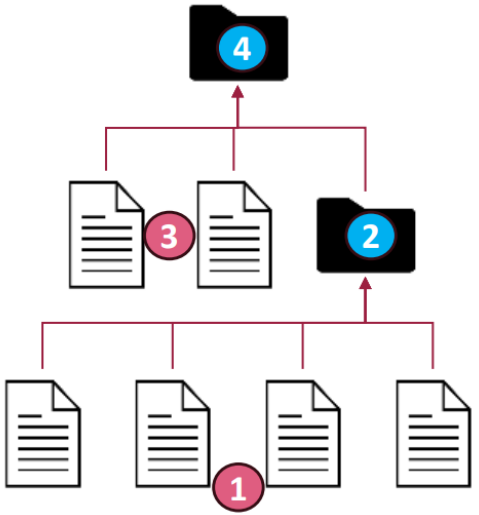
MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

ARCHIVES NATIONALES

Étape 3 : Compilation des résultats obtenus par le *LLM*

Path	Type	Titre	Description	Période de création : de	Période de création : à	Données à caractère personnel	Atteinte au relations extérieures, à la sécurité ou à l'ordre public	Affaires portées devant les instances juridictionnelles, extrajudiciaires, disciplinaires	Prévention, recherche, poursuite de faits punissables	Données commerciales et industrielles	Documents ayant trait au secret fiscal
2005 CANADA	directory	Christos Sirros	Ce dossier cor	08/06/2006	18/07/2024	non	non	non	non	non	non
2006 Québec Sirros Christons représentant	directory	"Christos Sirrc	Ce dossier cor	19/06/2006	18/07/2024	non	oui	oui	non	non	non
2007 Canada ambassadrice Glasgow	directory	"L'ambassadri	Ce dossier cor	20/03/2007	18/07/2024	non	oui	oui	non	non	non
2008 Visite ambassadeur de la pêche SEM SU	directory	"Visite d'amba	Dossier conte	06/10/2008	18/07/2024	non	non	non	non	non	non
2008 Visite ambassadeur de la pêche SEM SU	directory	"Visite de l'am	Dossier conte	30/10/2013	18/07/2024	non	non	oui	non	non	non
2012 Visite parlementaire\CV	directory	"Visite parlem	Ce dossier cor	11/01/2012	18/07/2024	non	oui	non	non	non	non
2012 Visite parlementaire\Documentation	directory	"Politique can	Dossier conte	25/01/2012	18/07/2024	non	oui	non	non	oui	non
2012 Visite parlementaire\Notes MAE	directory	"Relations cor	Ce dossier cor	19/01/2012	18/07/2024	non	non	non	non	non	non
2012 Visite parlementaire\Programme	directory	"Visite parlem	Ce dossier cor	25/01/2012	18/07/2024	non	oui	oui	non	non	non
2012 Visite parlementaire	directory	"Visite parlem	Ce dossier cor	25/01/2012	18/07/2024	non	oui	oui	non	oui	oui
2015 1617 04 comité exécutif de l'Associatio	directory	"Biographies e	Dossier conte	08/04/2015	18/07/2024	non	non	non	non	non	non
2015 1617 04 comité exécutif de l'Associatio	directory	"Relations Ca	Ce dossier cor	14/04/2015	18/07/2024	non	oui	non	non	oui	non
2015 1617 04 comité exécutif de l'Associatio	directory	"Comité exéc	Ce dossier cor	05/05/2015	18/07/2024	non	non	non	non	non	non
2015 1617 04 comité exécutif de l'Associatio	directory	"Procès-verba	Ce dossier cor	15/12/2016	18/07/2024	non	non	oui	non	non	non
2015 1617 04 comité exécutif de l'Associatio	directory	"Reconnaissa	Ce dossier cor	16/06/2015	18/07/2024	non	non	non	non	non	non
2015 1617 04 comité exécutif de l'Associatio	directory	"Relations Ca	Ce dossier cor	08/04/2015	18/07/2024	non	non	non	non	non	non
2015 1617 04 comité exécutif de l'Associatio	directory	"Comité exéc	Ce dossier cor	15/04/2015	18/07/2024	non	oui	oui	non	oui	oui
2016 AUDET MIchel Visite de courtoisie	directory	"Visite de cou	Dossier conte	28/09/2016	18/07/2024	non	oui	non	non	oui	non
2016 Pierre Marc Johnson\Revue de presse	directory	"Étude de pre	Ce dossier cor	21/01/2016	18/07/2024	non	non	non	non	non	non
2016 Pierre Marc Johnson	directory	"Pierre Marc J	Ce dossier cor	13/01/2016	18/07/2024	non	oui	oui	oui	oui	non
2017 Visite Assemblée du Québec	directory	"Visite de la d	Ce dossier cor	24/05/2017	18/07/2024	non	oui	oui	non	oui	oui
2018 Audet entretien Angel 18 janvier 2018	directory	"Biographie et	Dossier conte	15/01/2018	18/07/2024	non	non	non	non	non	non
2019 Québec Pierre-Luc Desgagné	directory	"Pierre-Luc De	Ce dossier cor	26/09/2019	18/07/2024	non	oui	non	non	non	non
2022 Orateur Penny QUébec	directory	"Politique scie	Ce dossier cor	26/10/2022	18/07/2024	non	non	oui	non	non	non
Ambassade\2009 Canada ambassadeur De L	directory	Accréditation	Dossier conte	15/06/2011	18/07/2024	non	oui	oui	non	non	oui

Traitement itératif…



1. Génération d'un titre et d'une description pour les **fichiers** en bas de l'arborescence
2. Génération d' un titre et d'une description du **dossier**, à partir des titres, descriptions et métadonnées des fichiers qu'il contient
3. Génération d'un titre et d'une description pour les **fichiers** du niveau supérieur
4. Génération d' un titre et d'une description du **dossier** du niveau supérieur à partir des titres, descriptions et métadonnées des fichiers et dossiers qu'il contient

ARCHIVES NATIONALES
Pierrefitte-sur-Seine – Paris

MINISTÈRE
DE LA CULTURE
Liberté
Égalité
Fraternité

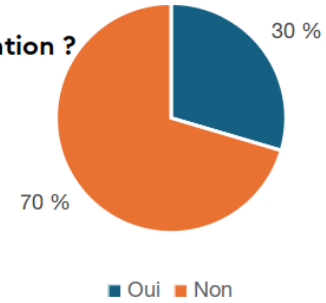
Les attentes des archivistes pour l'avenir…

MINISTÈRE DE LA CULTURE
Liberté Égalité Fraternité

Pratiques liées à l'IA dans le service d'archives (2/5)

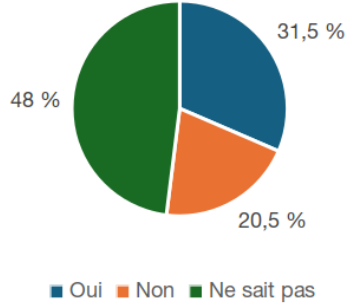
Au sein de votre service, utilisez-vous des outils commerciaux d'intelligence artificielle générative (Mistral IA, Chat GPT, ...) dans le cadre de vos missions d'archiviste ?
Si oui, pouvez-vous préciser votre utilisation ?

Oui	50
Non	119
TOTAL	**169**



Si non, envisagez-vous de les utiliser ?

Oui	40
Non	26
Ne sait pas	61
TOTAL	**127**



Les répondants qui ont utilisé des outils d'IA dans le cadre de leur missions archivistiques l'ont expérimenté sur l'ensemble des missions d'un archiviste.

On constate une grande prise en main de l'IA sur les sujets d'archivage électronique, notamment afin de générer des scripts de traitement automatique.

Les recherches juridiques ou historiques sont également très courantes, bien que les archivistes mentionnent le manque de fiabilité de l'IA et des sources qui ne sont pas systématiquement citées.

Plus de détail dans les trois diapositives suivantes.

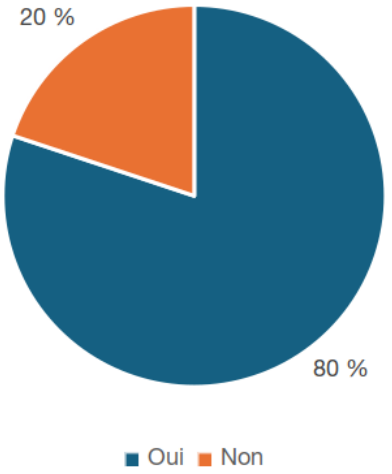


Enquête conduite par le service interministériel des Archives de France (M. Cournil) – avril 2026

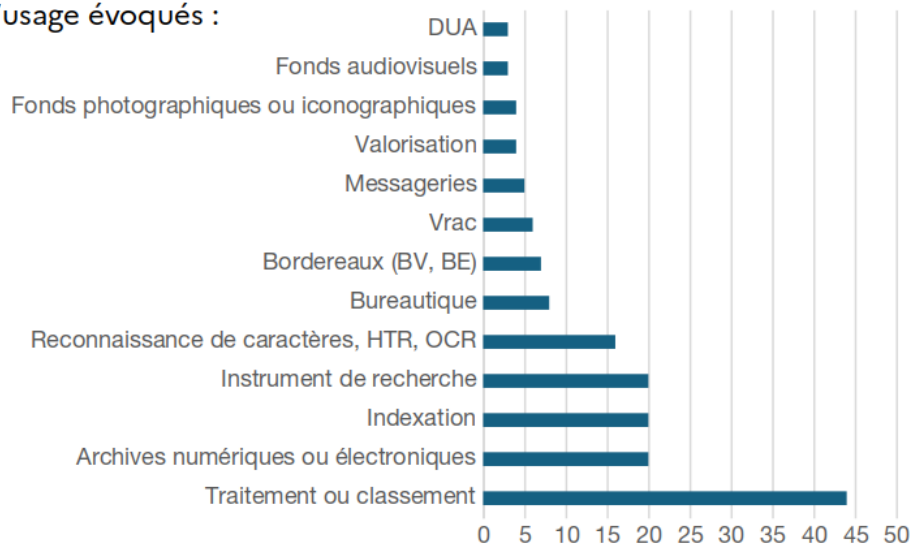
Attentes liées à l'IA dans notre métier

Avez-vous des attentes relatives à l'utilisation de l'intelligence artificielle par les archivistes ?

Oui	134
Non	34
TOTAL	**168**



Cas d'usage évoqués :



Enquête conduite par le service interministériel des Archives de France (M. Cournil) – avril 2026

ARCHIVES NATIONALES
Paris – Pierrefitte-sur-Seine

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité



Pour conclure…

Rôle primordial des structures de recherche dans le développement des projets

- Intérêt scientifique de redécouvrir certaines séries d'archives.
- Garantie d'une meilleure diffusion des données en l'état actuel.
- Maîtrise technique de l'IA par certains laboratoires universitaires.

Merci pour votre attention !

jean-francois.moufflet@culture.gouv.fr