La diffusion des archives à l’ère des humanités numériques

Sous la direction de :

Lionel Bartolini

Francesco Beretta

Dans le cadre de l’enseignement :

MA - pilier patrimoine régionale et humanités numériques

Université de Neuchâtel, semestres d’automne 2024-2025

Faculté  
Lettres et sciences humaines

Mémoire de Master réalisé par :   
Raphaël ROLLINET

Résumé

Le résumé ne doit pas dépasser une page. Il est rédigé dans le style « **Corps de texte** ». Il est suivi d’une liste de mots-clefs.

Table des matières

[Résumé i](#_Toc203427381)

[Liste des tableaux iii](#_Toc203427382)

[Liste des figures iv](#_Toc203427383)

[1. Introduction 1](#_Toc203427384)

[2. Méthodologie 3](#_Toc203427385)

[2.1 Délimitation du sujet 3](#_Toc203427386)

[2.2 Problématique et questions de recherche 3](#_Toc203427387)

[2.3 Méthode 3](#_Toc203427388)

[2.3.1 Revue de la littérature technique et professionnelle 3](#_Toc203427389)

[2.3.2 Utilisation de données en open data 3](#_Toc203427390)

[2.3.3 Enquête 3](#_Toc203427391)

[2.3.4 Création d’une base de données 3](#_Toc203427392)

[2.3.5 Visualisation de données 3](#_Toc203427393)

[2.3.6 Cas pratique 3](#_Toc203427394)

[3. Traitement du sujet 4](#_Toc203427395)

[3.1 Les archives 4](#_Toc203427396)

[3.1.1 Institutions 4](#_Toc203427397)

[3.1.2 Fonctions de l’archiviste 4](#_Toc203427398)

[3.1.3 Fonction de diffusions 4](#_Toc203427399)

[3.1.4 Stratégie numérique suisse 2025 4](#_Toc203427400)

[3.1.5 Livre blanc 4](#_Toc203427401)

[3.1.6 Enquête 4](#_Toc203427402)

[3.2 Les humanités numériques 4](#_Toc203427403)

[3.2.1 Domaine 4](#_Toc203427404)

[3.2.2 Formation 4](#_Toc203427405)

[3.2.3 Projets applicatifs 4](#_Toc203427406)

[3.2.3.1 Indexation des images 4](#_Toc203427407)

[3.3 Perspective 7](#_Toc203427408)

[4. Conclusion 8](#_Toc203427409)

[Bibliographie 9](#_Toc203427410)

[Annexe 1 : Titre de l’annexe 10](#_Toc203427411)

[Annexe 2 : Titre de l’annexe 2 11](#_Toc203427412)

Liste des tableaux

[Tableau 1 : Titre du tableau 2](#_Toc338749023)

Liste des figures

[Figure 1 : Titre de la figure 2](file:///\\heg-03\eid\_LaboDoc\1-Cours\160-travail_de_bachelor\Modele_TDB_HEG_2012.dot#_Toc338748353)

# Introduction

Quel est le rôle du patrimoine, si ce n’est de conserver et transmettre l’information et les connaissances amassées à travers le temps, non seulement à un public actuel, mais également futur. À cette fin, les archives tiennent un rôle central tant dans son soutien à la recherche comme passeur de sources auprès des chercheurs qu’auprès du citoyen comme garant de la transparence de l’état.

La profession d’archiviste tient sur des fonctions dont la finalité est celle de la diffusion, encadrée par une pratique professionnelle bien établie, des normes, une législation et une déontologie professionnelle. La diffusion auprès d’un public est donc au cœur de la réflexion de l’archiviste tant pour les besoins actuels que futur.

Nos sociétés modernes actuelles sont en évolution constante au sein de ce que nous pouvons décrire comme la 4e révolution de l’information, soit la communication informatique. Dans ce contexte les archives tiennent un rôle de médiateur dans cette diffusion et s’adaptent à de nouveaux médiums documentaires numériques.

En suisse les archives sont compris dans la stratégie numérique 2025 de la Confédération (Chancellerie fédérale 2025) visant directement à permettre l’accès numérique des archives indépendamment du temps et du lieu. Cela deviendra la norme d'ici 2025, tandis que l'accès analogique deviendra l'exception. Cette stratégie numérique met également l’accent sur l’utilisation de l’IA, l’utilisation de logiciel ouvert et la sécurité de l’information.

Dans ce contexte, cette recherche tend à montrer les évolutions de cette fonction archivistique de diffusion à l’ère numérique, par un regard d’un archiviste humaniste. Cette recherche vise également à comprendre les évolutions actuelles qui impactent les archives, les tendances des institutions locales, les innovations en la matière, d’en donner un panorama à l’aune du domaine des humanités numériques.

En effet, les humanités numériques sont une discipline relativement jeune, aux croisements des sciences humaines et de l’informatique. Elle vise comme transdiscipline à ouvrir de nouvelles perspectives de recherches grâce aux développements de méthode numérique. Il s’agit d’une discipline particulièrement pertinente dans le cadre de cette recherche, car les archives sont souvent un passage inévitable pour les chercheurs de SHS et un grand nombre d’archivistes sorte de formation initiale en science humaine ou sociale, particulièrement l’histoire dont l’archivistique en est une science connexe.

Les humanités numériques sont donc pertinentes pour fonder un discours sur les évolutions numérique touchant l’ensemble des GLAM notamment les archives tant du point de vue du chercheur en SHS que du professionnel des archives. De plus les archives ont toujours constitué un laboratoire pour les humanités numériques pour appliqué des méthodes tant du point de vue du traitement automatique du langage, de l’indexation automatique qu’aujourd’hui avec le web sémantique auquel les archives s’adaptent notamment avec la nouvelle norme Record in Context.

Mais il s’agit au sens large d’un ensemble de méthode computationnelle tel que l’intelligence artificielle générative, le machine learning, le deep learning pratiqué depuis longtemps en humanités numériques que l’on peut aujourd’hui retrouver cité comme objectif prioritaire de développement pour les archives. L’on peut évoquer récemment le « Livre blanc de l’apprentissage automatique dans les archives » rédigé par l’Association suisse des archivistes, qui tend à montrer des synergies importantes entre les humanités numériques et les objectifs stratégiques des archives en Suisse.

Ces méthodes numériques ont un point commun, elles permettent l’exploitation et l’enrichissement documentaires pour offrir une meilleure diffusion à l’ère numérique que nous connaissons actuellement. Dans ce cadre cette recherche menée par l’axe d’évolutions numériques touchant les archives, s’intéressera à la fonction de diffusion dans un contexte institutionnel régional.

Cette recherche intègrera une réflexion sur la place des humanités numériques tant sur le plan disciplinaire que de l’importance de la formation, aussi bien pour la pratique de la recherche que de celle des professions du patrimoine. Mais également l’importance du point de vue de l’archiviste d’avoir une convergence entre la technique ou pratique professionnelle et celle des enjeux liés aux technologies dans le cadre de son rôle de gardien d’une mémoire commune.

Enfin cette recherche questionnera ses évolutions pour résoudre des problèmes pratiques actuels dans le cadre de la diffusion archivistique en motivant ses réflexions à l’aune d’un nouveau paradigme tel que défini par Thomas Kuhn. L’idée de cette recherche n’est pas d’apporter des certitudes en termes de prospectives professionnelles, mais de structurer ce travail par une approche structurale de la diffusion archivistique, à la fois épistémologique par sa définition conceptuelle, sociologique de par son apport à la société et enfin historique du point de vue d’une étude de l’évolution de la pratique de cette diffusion archivistique ses dernières années à un niveau régional.

Finalement, ces réflexions mêlant patrimoine régional et humanités numériques, provenant d’un archiviste de tradition humaniste, tendent à valoriser à la fois la place du patrimoine dans la société, son rôle pour la recherche que celui d’inscrire les sciences humaines dans un débat légitime et contemporain sur le numérique. Mais également sur notre consommation intelligente des technologies de l’information en tant qu’outils aux services tant des humanités que de l’Humanité.

# Méthodologie

## Délimitation du sujet

## Problématique et questions de recherche

**Problématique principale**

* Quel est l’impact du numérique dans la diffusion archivistique en suisse romande ?

**Problématique secondaire**

* Les stratégies institutionnelles du domaine des archives sont-elles en adéquation avec les évolutions numériques actuelles ?
* Quels sont les apports du domaine et des formations en humanités numériques pour ces évolutions technologiques aux domaines du patrimoine ?

## Méthode

### Revue de la littérature technique et professionnelle

### Utilisation de données en open data

### Enquête

### Création d’une base de données

### Visualisation de données

### Cas pratique

* Traitement de fonds d’images
* Indexation automatique
* Web sémantique

# Traitement du sujet

## Les archives

### Institutions

### Fonctions de l’archiviste

### Fonction de diffusions

### Stratégie numérique suisse 2025

### Livre blanc

### Enquête

## Les humanités numériques

### Domaine

### Formation

### Projets applicatifs

Maintenant que nous avons identifié les besoins numériques métiers des archivistes notamment dans le livre blanc de l’association des archivistes, il est intéressant de montrer des exemples d’application des humanités à celle-ci. Etant archiviste de profession et étudiant en humanités numériques il m’a paru propice d’illustrer quelques applications apporter par des compétences en humanités numériques vu dans des cours dans les universités en Suisse romandes sous la forme de mini-projet représentatif.

#### Indexation des images

Les images sont très présente en archives qu’elles soient physiques ou numériques. En effet, le métier d'archiviste évolue vers une numérisation et automatisation croissante, passant obligatoirement par une rétronumérisation des archives et donc par la gestion d'images numérisées. L'arrivée actuelle d'outil performant de traitement automatique de l'image fait de l'automatisation une priorité pour la profession symbolisée par de nombreux articles comme "Du Machine Learning aux archives" [Source] (Reyrodriguez 2021). Mais encore au niveau stratégique l'écriture d'un livre blanc sur l'apprentissage automatique dans les archives par l'Association des Archiviste Suisse (Arnold et al. 2024) dont il décrit la numérisation comme un préalable.

L’utilisation de méthode permettant l’indexation ou la description automatique des images est parfaitement représentatif des perspectives des humanités numériques aux archives. Le code utilisé dans le cadre de ce mini-dossier est adapté en partie du code vu dans le cours « Du Pixel aux images : une Introduction au traitement numérique en 2D » à l’Université de Genève. Il vise à effectuer une détection d'objet sur des images, tout en répondant à des contraintes métiers qui sont par exemple un besoin d'automatiser le traitement à partir d'un workflow. Mais également en termes de rendu, un archiviste aura besoin d'extraire les détections sous format CSV afin qu'elles soient intégrées aux notices d'un système d'information archivistique par exemple sous la forme d'une indexation. Le code permet à n'importe quel archiviste avec des notions informatiques minimales de l'exécuter.

Les méthodes utilisées dans le cadre de ce projet sont un mélange de méthodes de deep learning ou apprentissage profond pour le traitement d'image. Pour être exactes, les méthodes fonctionnent en mode inférence uniquement, et exploitent des poids figés issus de l’entraînement supervisé de grands corpus d’images. Leurs spécifications sont les suivantes :

[Tableau]

Les modèles de détections et classifications des images seront appliqués à un fonds d'images datant des années 1900 environ et provenant de l'inventaire des archives de la commune de Milvignes, membre du service intercommunal d'archives de Neuchâtel. La plateforme de diffusion des inventaires est gérée par docuteam SA, entreprise de gestion d'archives qui m'a employé de 2016 à 2021. Les images ont été numérisées par mes soins en format de conservation (TIFF) et archivées électroniquement dans un repository de préservation numérique et sont également disponibles librement en téléchargement dans un format dégradé de diffusion (JPEG).

**Figure 1: Images numérisées de la commune de Milvignes**

Une image contenant arbre, Papier photographique, collage, maison

Le contenu généré par l’IA peut être incorrect.

Les résultats appliqués à un fonds d'images anciennes ont d'abord été mitigés. Avec des résultats à seulement 33% de détection correcte les modèles SSD MobileNet v2 (TensorFlow Hub) et ViT (Vision Transformer) ne sont pas adapté a mon fonds d'images. Le code a parfaitement fonctionné et le traitement de détection et classification a parfaitement été appliqué sur le fonds d'images. Cependant les images utilisées mettent en évidence certaines limitations de cette méthode.

La première provient probablement des images elles-mêmes, il s'agit de JPEG utilisé pour la diffusion sur le web avec une faible résolution. Les modèles se basant sur les pixels de l'image pour détecter les objets, la qualité de l'image influence le résultat final.

Deuxièmement les modèles utilisés ont été entrainés sur des images récentes, cela peut-être observé notamment avec les classes "COCO". Le modèle fonctionnant par inférence peu importe l'image, il y a un taux d'erreur plus important sur des images anciennes. Une première solution a été de pallier ce biais en enlevant des classes sans amélioration probante. Néanmoins cela reste un biais inhérent à ces modèles. En effet, une étude datant de 2021 utilisant le modèle SSD MobileNet v2 effectué sur une base de données de 50000 images confirme un taux de précisions variant entre 32 et 47% environ (Cohen et al. 2021) soit dans la marge d'erreur de mon résultat de 33% de précision.

À ce stade avec les résultats de ces 2 premiers modèles très modestes, un test de modèle plus récent avec un LLM multimodal "llama-4-maverick" m’a paru intéressant. Ce modèle d'un tout autre type que les deux précédents est adapté tant au texte qu'aux images. Son point fort est ça polyvalence tant sur le plan des supports qu'elles soient textuelles ou sous forme d'images, mais également pour sa compréhension multilingue comprenant le français.

Comme il s'agit d'un LLM fonctionnant via une API importée directement dans le code, cela permet une plus grande liberté d'interaction avec l'utilisation d'un prompt. Le prompt est capital pour la qualité du résultat, je me suis ici limité à une indexation des images soit une liste de mots-clés représentant les détections d'objets.

Le prompt utilisé est le suivant :

"Vous êtes expert en indexation d'archives photographiques. Votre tâche consiste à effectuer une indexation descriptive de 3 à 5 mots-clés en français pour chaque image. Basez votre analyse uniquement sur le contenu visuel de l'image ; ne formulez aucune hypothèse et n'incluez aucune information supplémentaire. Formulez votre réponse par des mots-clés d'indexation. Exemple : mot-clé1, mot-clé2, mot-clé3. Analysez maintenant l'image fournie et générez les mots-clés d'indexation."

Après analyse des résultats, Llama-4 obtient d'excellents résultats avec 88% de détection parfaitement correcte. Il faut également faire ressortir la nature des erreurs par Llama-4, soit la détection d'un "banc" pour décrire des enfants assis sur un escalier et une "église", le bâtiment en question est en fait un château avec un clocher de chapelle. Il est probable qu'un opérateur humain aurait également pu commettre ces erreurs.

Une image contenant texte, capture d’écran, diagramme, Rectangle

Le contenu généré par l’IA peut être incorrect.

Ces résultats sont significatifs dans le sens où ils sont en mesure d'égaler un opérateur humain sans vérification systématique des résultats des détections d'objets. Il s’agit d’imprécision qui ne sont pas nécessairement rédhibitoire selon l’objectif de l’utilisation. En effet pour les images sont très présentes dans les fonds d’archives mais ne sont pas forcément décris en détail à la « pièce », souvent la description s’arrête à un titre relativement descriptif, parfois plus détailler au niveau du « dossier ».

L’utilisation d’une indexation ou d’une description automatique par un LLM d’un grand nombre d’images permettrait donc d’offrir une description minimale au niveau de la pièce ou d’enrichir celle de l’archiviste afin d’augmenter les points d’entrés de l’information aux notices à des fin de recherche.

Il s’agit dans ce cas d’un mini projet appliqué à des images d’archives, mais qui est intéressant dans l’optique de considérer les applications des humanités numériques au domaine du patrimoine et particulièrement aux archives. De plus il permet d’exercer un avis critique sur les différents modèles et de constater l’évolution rapide et le gain de performance entre des modèles entrainés il y a quelques années et un LLM « grand public » récent.

## Perspective

# Conclusion

Bibliographie

CHANCELLERIE FÉDÉRALE, 2025. Suisse numérique. *Suisse numérique* [en ligne]. 2025. Disponible à l’adresse : https://digital.swiss/fr/plan-d-action/mesures/acces-en-ligne-aux-archives [consulté le 14 mars 2025].

1. Titre de l’annexe
2. Titre de l’annexe 2