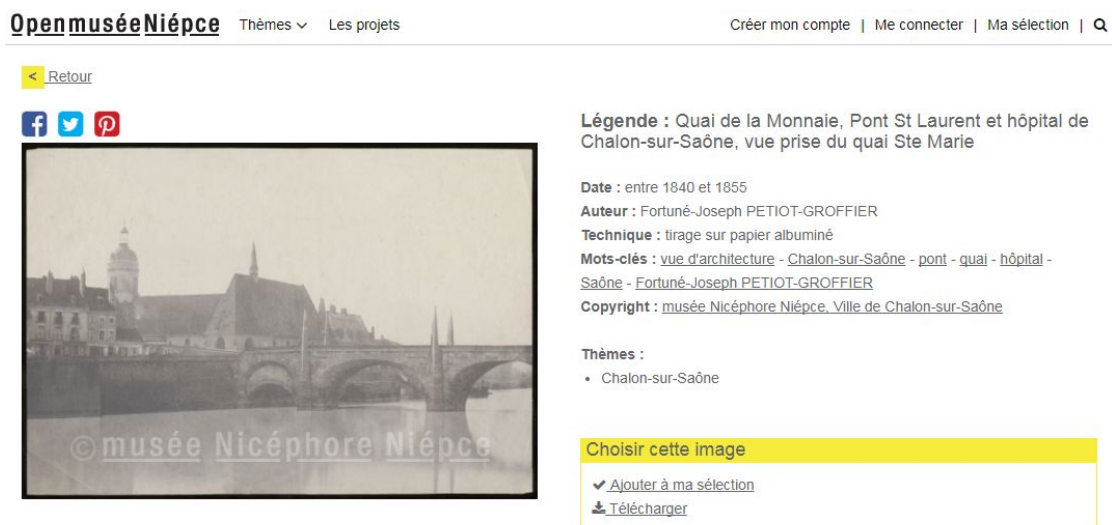

Géoréférencement des métadonnées de ressources patrimoniales et culturelles

A. Contexte

De plus en plus d'institutions patrimoniales et culturelles publient les catalogues des collections dont elles ont la garde sur le Web afin de faciliter la découverte de ces collections. Ces catalogues se composent, pour chaque ressource patrimoniale ou culturelle, de métadonnées permettant d'en découvrir le contenu même lorsque aucune copie numérique de la ressource n'est disponible: titre, auteur, éditeur, date, mots-clés, etc. Désenclaver ces jeux de métadonnées suppose de les interconnecter avec des référentiels communs, comme des thésaurus sur les types d'oeuvres architecturales, des index d'auteurs ou des jeux de données géographiques. Ceci permet de désambiguïser les chaînes de caractères utilisées pour désigner les auteurs, les types d'oeuvres ou les lieux en leur substituant les identifiants de ressources décrites par les référentiels et faisant autorité. En outre, ceci permet de naviguer aisément au travers des collections et de leur référentiels associés en suivant les liens ainsi explicités.

Le projet ALEGORIA vise à faciliter la valorisation de fonds iconographiques institutionnels décrivant le territoire français à différentes époques allant de l'entre-deux-guerres à nos jours. Ces clichés s'accompagnent de métadonnées, et lorsqu'une photographie représente une entité géographique, sa légende peut contenir des indications sur la localisation de cette entité ou sur le lieu de prise de vue. On retrouve ce type d'informations dans les métadonnées de la photographie présentée en figure 1.



The screenshot shows the OpenmuséeNiépce website interface. At the top, there is a navigation bar with the site name, a dropdown menu for 'Thèmes', and a link for 'Les projets'. On the right side of the header, there are links for 'Créer mon compte', 'Me connecter', 'Ma sélection', and a search icon. Below the header, there is a yellow button labeled 'Retour' with a left-pointing arrow. To the left of the main content area, there are social media icons for Facebook, Twitter, and Pinterest. The central focus is a large black-and-white photograph of a bridge over a river, with buildings in the background. A watermark '© musée Nicéphore Niépce' is visible at the bottom of the photo. To the right of the photo, the metadata is displayed in a structured format: 'Légende : Quai de la Monnaie, Pont St Laurent et hôpital de Chalon-sur-Saône, vue prise du quai Ste Marie', 'Date : entre 1840 et 1855', 'Auteur : Fortuné-Joseph PETIOT-GROFFIER', 'Technique : tirage sur papier albuminé', 'Mots-clés : vue d'architecture - Chalon-sur-Saône - pont - quai - hôpital - Saône - Fortuné-Joseph PETIOT-GROFFIER', and 'Copyright : musée Nicéphore Niépce, Ville de Chalon-sur-Saône'. Below the metadata, there is a 'Thèmes' section with a bullet point for 'Chalon-sur-Saône'. At the bottom right, there is a yellow box containing the text 'Choisir cette image' and two links: 'Ajouter à ma sélection' and 'Télécharger'.

Figure 1: Une photographie conservée au musée Nicéphore Niépce et ses métadonnées (<http://www.open-museeniepce.com/recherche-photos/photo,2611>)

B. Principales hypothèses et objectifs

Une des réalisations prévues par le projet est un moteur d'indexation et de recherche multimodal et à grande échelle, couplant recherche par contenu et par métadonnées dans les fonds d'images numérisées et documentées. Ce moteur s'appuiera, entre autres, sur les indications de localisation fournies par les métadonnées des images. Pour ce faire, une première étape consistera à extraire des métadonnées disponibles les indications de localisation, représentées sous la forme d'entités spatiales nommées. Il conviendra en outre de les typer afin de préciser si ces entités spatiales nommées désignent le contenu de l'image ou le lieu de sa prise de vue. Enfin, une dernière étape consistera à résoudre ces entités spatiales nommées, c'est-à-dire à associer à chaque mention d'entité nommée l'identifiant de l'entité du monde réel à laquelle elle fait référence, décrite par une ressource d'autorité dans un jeu de données de référence. L'utilisation d'un jeu de données de référence comportant des indications de localisation directes (coordonnées ou géométrie), permettra de localiser précisément les entités spatiales mentionnées dans les métadonnées. Les différentes étapes de l'approche de géoréférencement proposée pourront s'appuyer sur des travaux antérieurs du LaSTIG/COGIT [M15][PAB17].

C. Verrous scientifiques

Une première difficulté réside dans le manque d'informations de contexte dans les métadonnées qui restent des textes relativement courts. Les approches d'extraction et de résolution des entités spatiales choisies devront s'adapter à cette contrainte. Le traitement des entités spatiales relatives constitue une seconde difficulté majeure. En effet, il s'agit d'entités spatiales identifiées et localisées par référence à une autre entité spatiale nommée, comme "la rive droite de la Saône" par exemple. L'approche d'extraction proposée devra être en mesure de détecter ces entités spatiales relatives. En outre, dans la mesure où elles ne figurent pas dans les référentiels géographiques, il conviendra de proposer une approche de résolution adaptée pour dériver leur localisation à partir des données disponibles.

D. Compétences particulières et formation requises

Compétences et connaissances: Informatique (programmation Java), extraction d'information à partir de textes, données géographiques vectorielles, Web de données.

Formation: Master 2 ou troisième année d'école d'ingénieur en informatique ou en géomatique avec une forte composante informatique.

Durée et période de stage : 5 mois, au cours du printemps et de l'été 2018.

Lieu de stage : Equipe LaSTIG/COGIT de l'Institut National de l'Information Géographique et Forestière (IGN), à Saint-Mandé (métro 1, station Saint Mandé). Le stage se déroulera dans l'équipe COGIT menant des recherches en géomatique sur l'intégration, la diffusion, la représentation et l'analyse de données géographiques sous forme de référentiels vectorisés et à grande échelle.

Indemnités de stage: Stage gratifié.

Modalités de candidature : Envoyer par email et au format PDF en un seul fichier, votre curriculum

vitae, une lettre de motivation ciblée sur le sujet, vos relevés de notes des deux dernières années d'études.

Encadrement de stage :

Nathalie Abadie (IGN/COGIT): nathalie-f.abadie@ign.fr

Carmen Brando (EHESS): carmen.brandot@ehess.fr

E. Bibliographie

[M15] Moncla, L. (2015) Automatic Reconstruction of Itineraries from Descriptive Texts. Thèse de doctorat de l'Université de Pau et des Pays de l'Adour et de l'Université de Saragosse.

[PAB17] Paris, P-H, Abadie, N., Brando C. (2017). Linking spatial named entities to the Web of Data for geographical analysis of historical texts. *Journal of Map & Geography Libraries*. Volume 13, 2017 - Issue 1: Semantic Historical Gazetteers: A Place for Places - Papers from the DH2016 GeoHumanities SIG Workshop.