



HAL
open science

Automatic reconstruction of the genealogy of plots and buildings from historical sources

Solenn Tual, Nathalie Abadie, Bertrand Duménieu, Joseph Chazalon, Julien Perret

► To cite this version:

Solenn Tual, Nathalie Abadie, Bertrand Duménieu, Joseph Chazalon, Julien Perret. Automatic reconstruction of the genealogy of plots and buildings from historical sources. International Semantic Web Research Summer School (ISWS 2024), Jun 2024, Bertinoro, Italy. . hal-04610537

HAL Id: hal-04610537

<https://hal.science/hal-04610537>

Submitted on 13 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

CONTEXT

- **Fast and consistent progress brought about by deep neural networks** in the field of extracting information from unstructured historical sources.
- The aim of the French National Mapping Agency (IGN) to complete its databases with information such as building construction dates.

MOTIVATIONS

The Napoleonic cadastre is the **first French land registry**, dating from the early 19th century to the mid-20th century. It is rich in **information on land use/cover evolution** provided for the whole territory before the generalization of aerial photography. It is complex to exploit due to its large corpus size, its nature and its content.

The targeted IGN application and the chosen archival sources lead to the following objectives:

- Propose an **automated processing pipeline to extract information from archival documents**.
- Structure the extracted data into a **geo-historical knowledge graph**.

THE NAPOLEONIC CADASTRE

Two main types of documents : **maps** and **registers**.

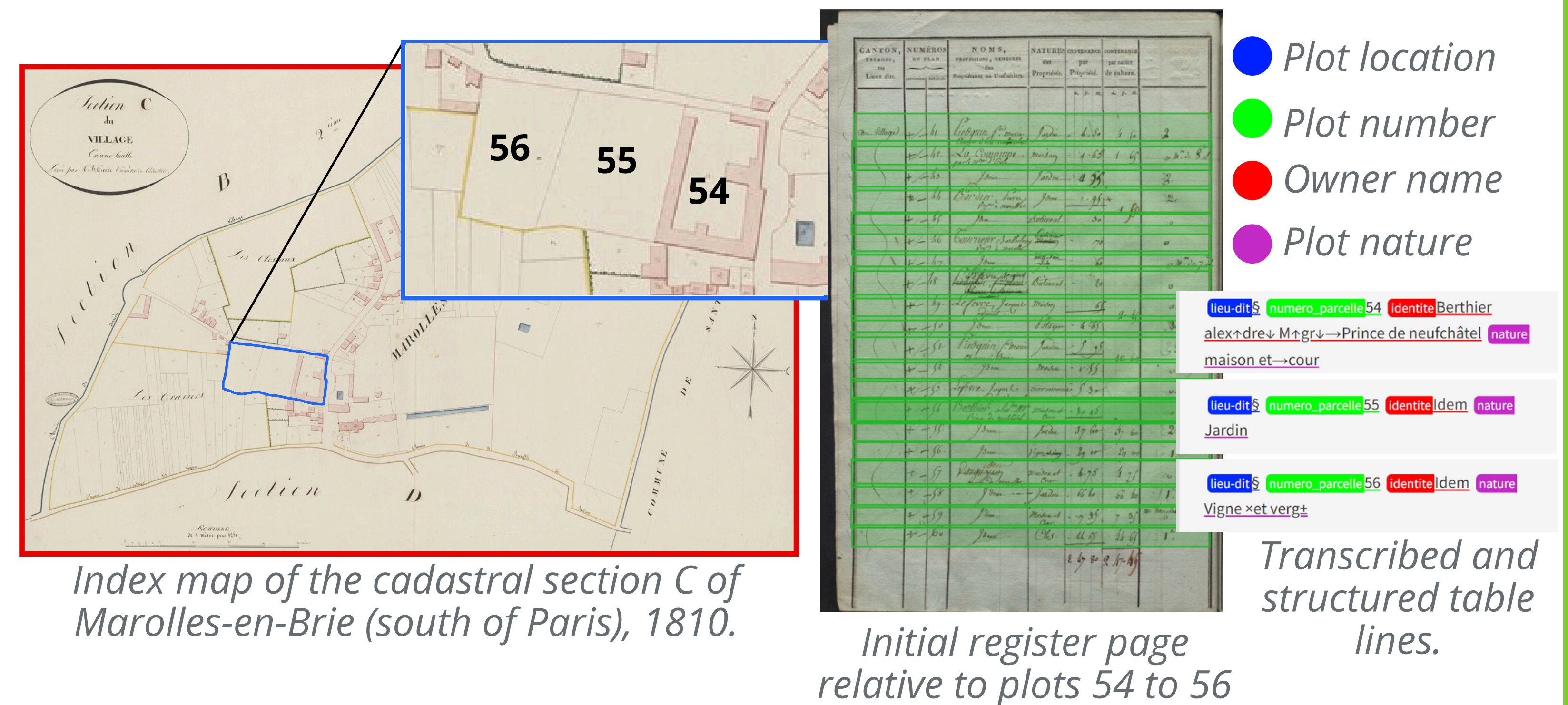
Maps include :

- **Cadastral index maps** represent the built and unbuilt properties
- The **assembly map** of the cadastral index maps shows the area represented by each index maps sheets of a given commune.

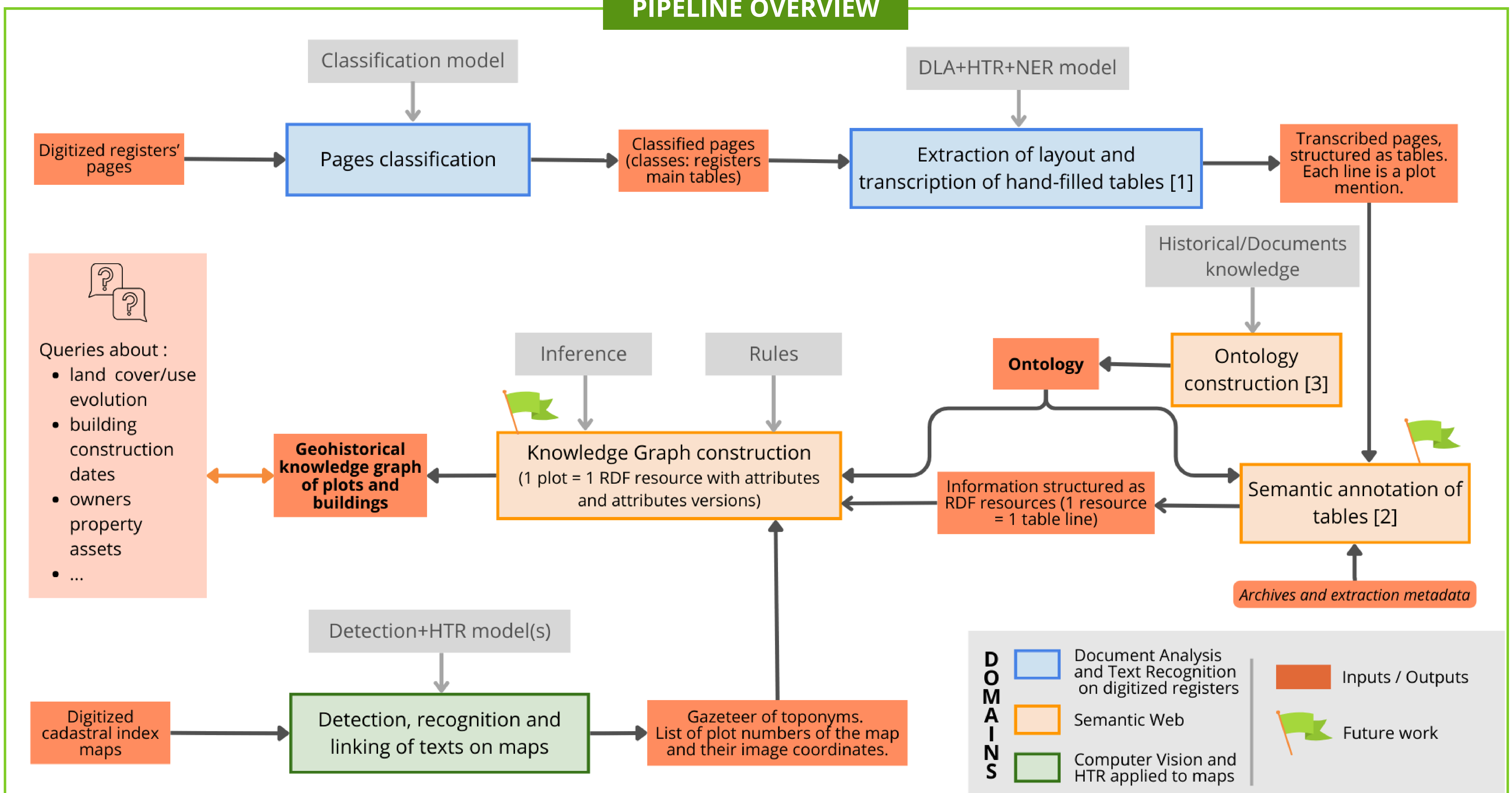
Maps were not updated until a new cadastre was initiated (several decades separate one version from another).

Registers include :

- **Initial registers** filled at the same time of the maps production and are considered as its legend. Gives first owner name, nature and initial taxes amount for each plot. They should not have been updated after creation.
- **Mutation registers** : created at the same time of the initial register. Updated each time a change happened to a plot.



PIPELINE OVERVIEW



TAKEAWAYS

- The proposed method consists of an **end-to-end pipeline to extract, structure and retrieve information contained in archival documents**.
- The annotation model designed for the extraction and transcription model is fully compatible with **semantic annotation of tables tasks**.
- The table annotation tasks are performed to create RDF resources for each plot mention extracted from the registers.
- After being linked and reordered, the resulting RDF resources can be used to create a geohistorical knowledge graph that allows us to follow the evolution of the plots in terms of each of their attributes (owner, nature, etc.).

ACKNOWLEDGEMENT

This work is supported by the French National Mapping Agency (IGN) and the Innovation Defence Lab (AID) of the French Ministry of Defense.

Digitized documents are provided by the Archives of the French department of Val-de-Marne (AD94)

REFERENCES

- [1] Coquenot et al, IEEE TPAMI 2023
- [2] Liu et al, Journal of Web Semantics 2023
- [3] Bernard et al, GAST 2024



SCAN ME !