



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

Transforming Historical Archives into Searchable Open Data with AI

WIPO ICT Leadership Dialogue (2nd Session) – 15th April 2026

Samir Ghamri-Doudane – INPI France

inpi
FRANCE

OPEN DATA PORTAL

Simple and seamless access to 48.8 million records

Welcome to DATA INPI

Search in all databases (Companies, Trademarks, Patents and Designs)

Name, applicant, SIREN, etc.

FIND

> Advanced search

OPEN DATA ▾

Licences for reuse

Access to APIs - Companies

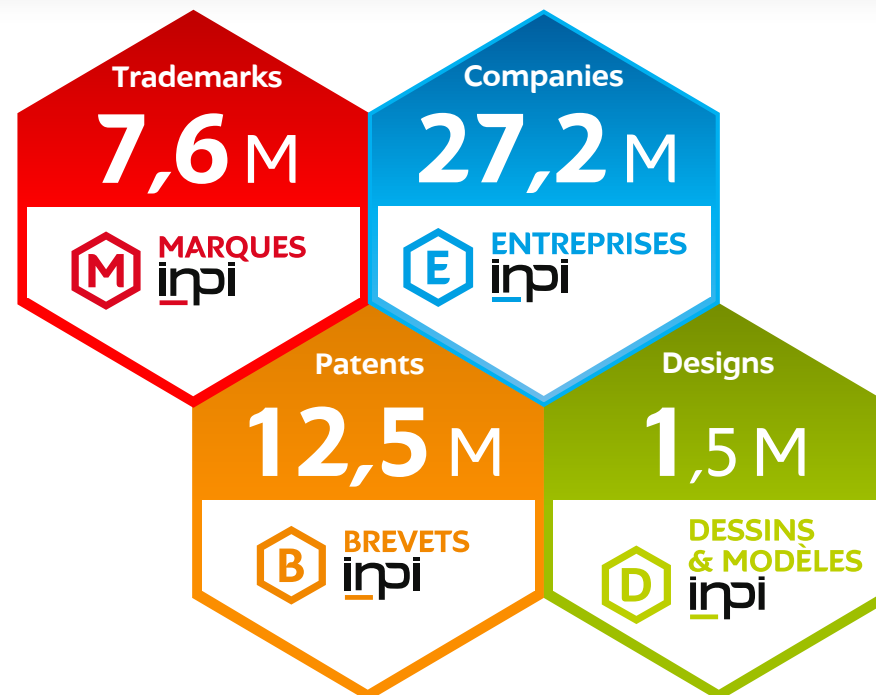
Access to the FTP server - Companies

Access to APIs - Industrial property - Swagger

Access to APIs - Industrial property

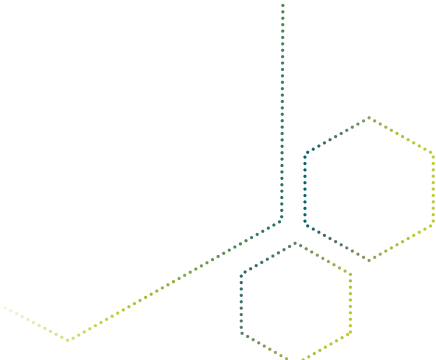
Access to the FTP server - Industrial property

Multiple data access methods



OPEN DATA PORTAL

A dedicated section for historical data



▼ Free-text search *i*

Operator And ▼

▶ Applicant or representative

▶ Patent number (from 1844 onwards) *i*

▶ Municipality *i*

▶ Applicant profession

▼ Years of filing

1791 1910

Search options for historical patents



INPI's historical archives

Explore two hundred years of historical patent and trade mark archives. Discover the outstanding results of an inventive spirit, revolutionary innovations and inspiring characters. Explore the exclusive collections preserved by INPI!

Search patents
from 1791 to 1901

Search trade marks
from 1857 to 1920

But limited search options for historical trademark (only class and year) !

INDEX HISTORICAL TRADEMARK GAZETTES

Build a searchable database

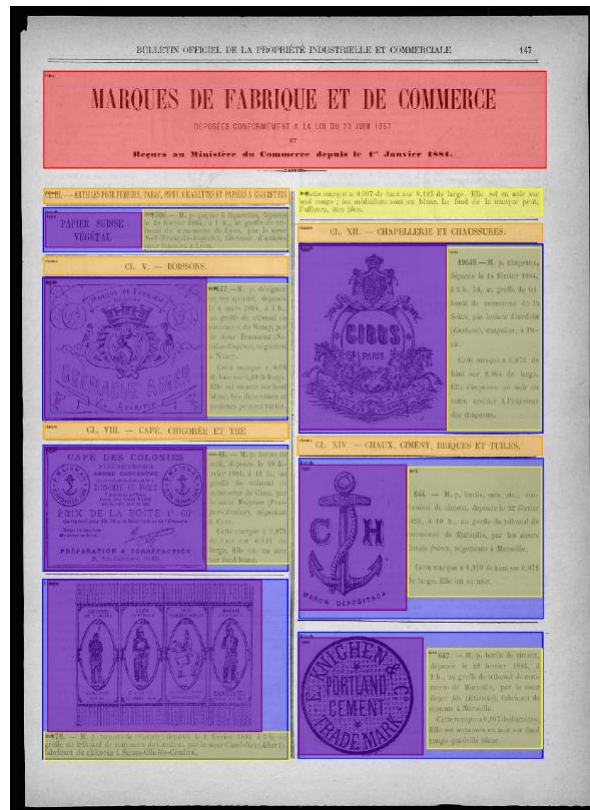


- ▶ **1884 to 1920 official gazettes and application forms** available in paper and scanned formats
- ▶ **With the objective to enrich** our historical open data portal with advanced search capabilities
- ▶ **Automated data extraction and indexing** are needed to enable multi-criteria searches
- ▶ **Ongoing work** implementing a set of models and functions to extract figures and structured data

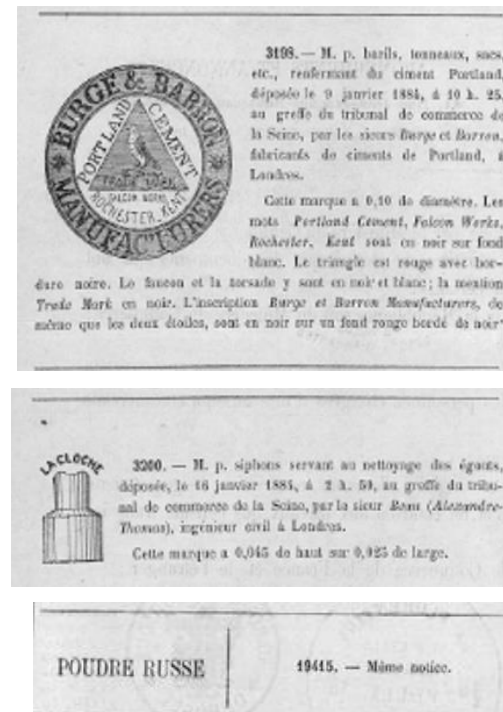
INDEX HISTORICAL TRADEMARK GAZETTES

Step 1 – Segmentation and labelling

- ▶ **Supervised model** to detect and label the different components of a page (trained in-house)
- ▶ **Automated functions** to verify, potentially correct and validate the model output



Model output



Complex task due to format variations

CL. XXIII. — EAUX-DE-VIE.

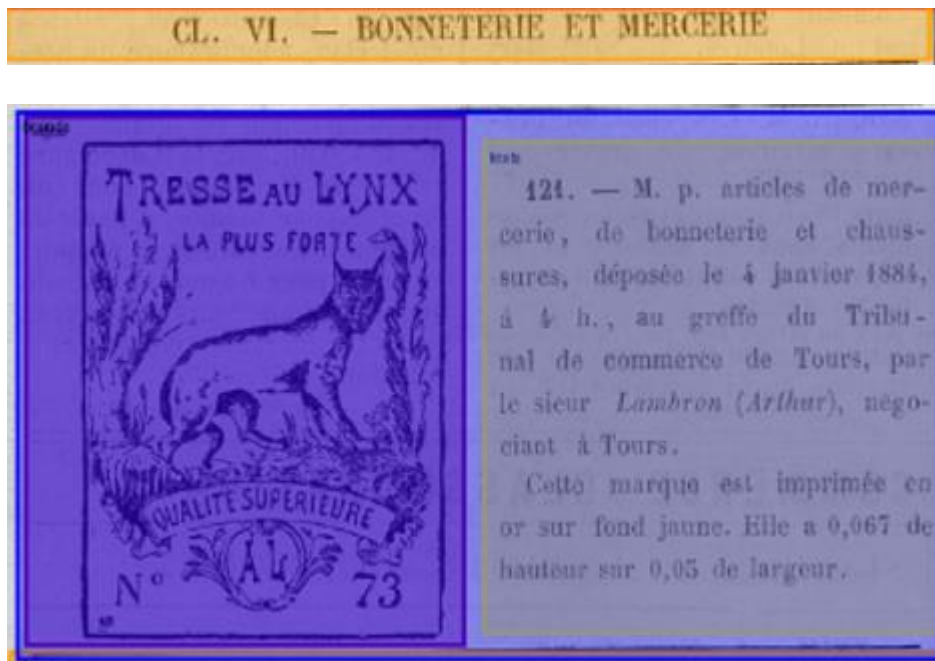
CL. IX. — VINS.


MARQUES ÉTRANGÈRES

INDEX HISTORICAL TRADEMARK GAZETTES

Step 2 – Information extraction and indexation

- ▶ **OCR tools** to get machine readable text (using Mistral LLM-powered solution)
- ▶ **Simple rules** to extract structured data and to assign the class and figurative element



Figurative Element	
	Class: Bonneterie & Mercerie
	Registration number: 121
	Date: 04 January 1884
	Registration at : Tours
	Applicant: Arthur Lambron
	...

INDEX HISTORICAL TRADEMARK GAZETTES

Step 3 – Retrieving original application forms

- ▶ **Computer vision models** to identify the related form based on the figurative element using vector embeddings and similarity measures





RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

Thank you



www.inpi.fr



+33 (0)1 56 65 89 98



INPI France

inpi
FRANCE