



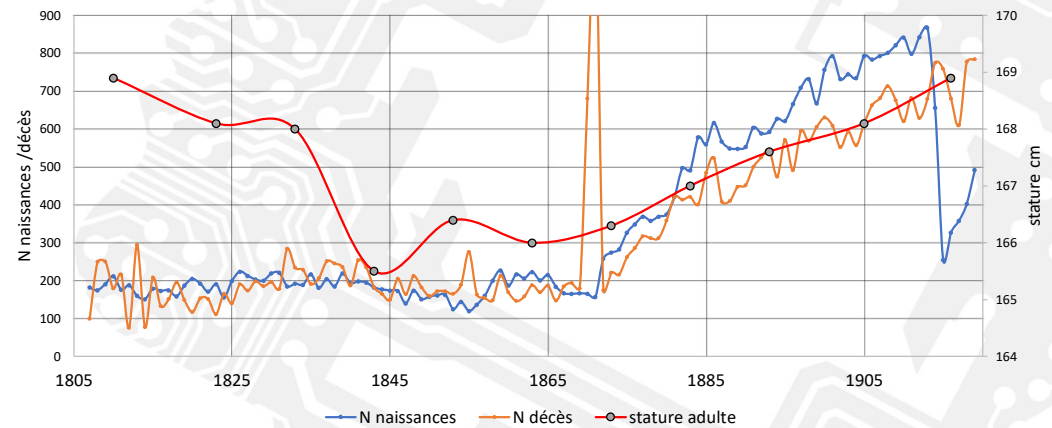
W. Al Kendi, G. Frossard, L. Heyberger

Belpop: a history-computer project to study the population of a town during early industrialization



N of births (blue) and deaths (red) and stature of conscripts (red with points) in Belfort during early industrialization (Heyberger 2013 & Belpop project)

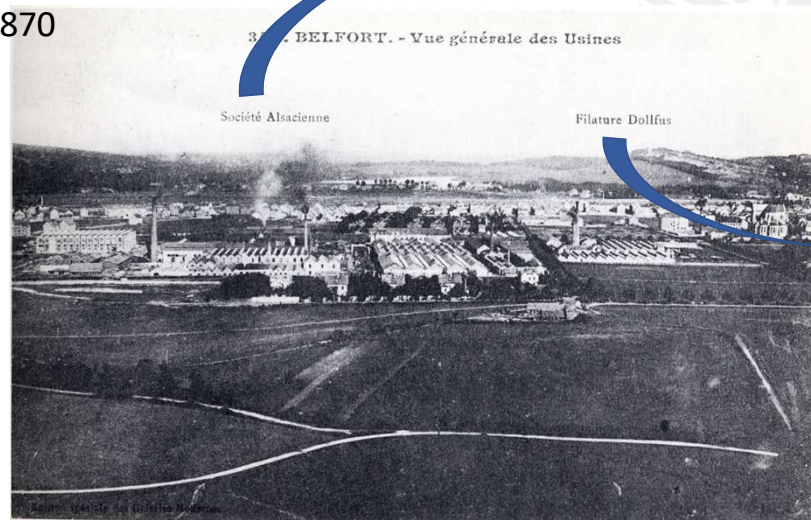
naissances & reconnaissances, décès (état civil) et stature des conscrits à Belfort (1807-1919)



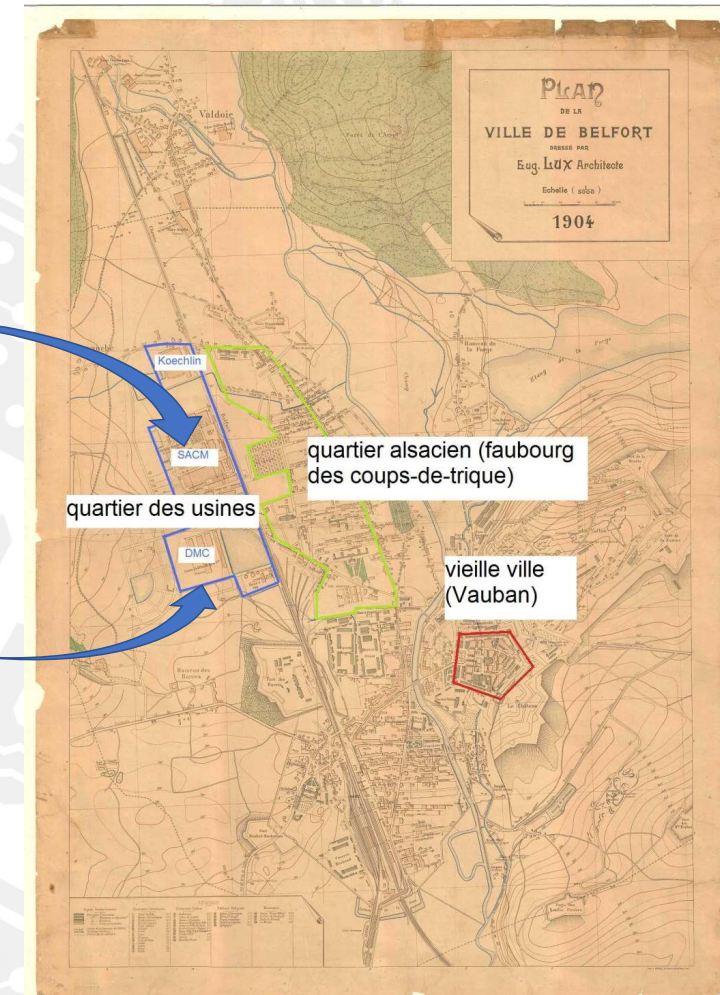
Introduction



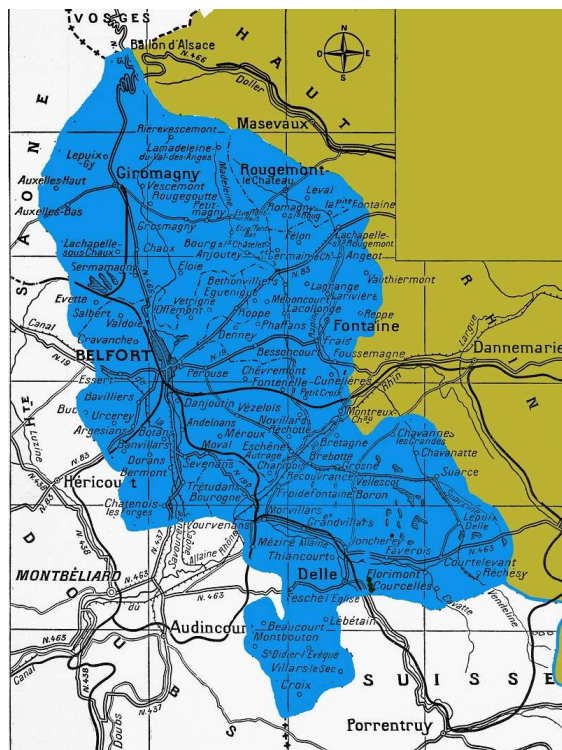
one of the many barracks built after 1870



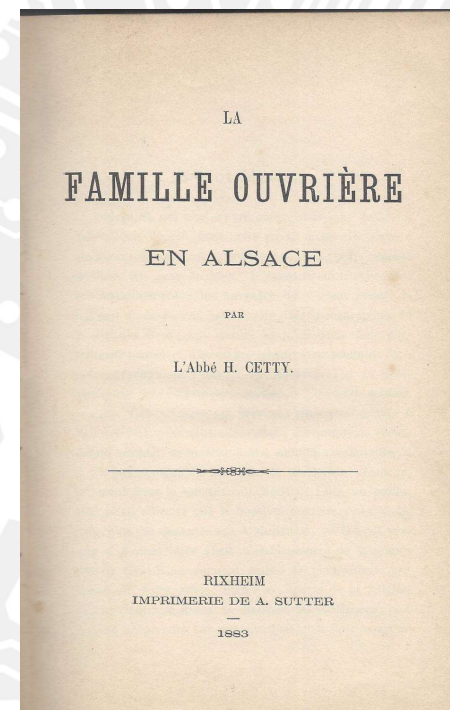
new factories (blue) and new working-class district (green) in Belfort (old town pre-dating 1870 in red)



New border between Territoire de Belfort (blue) and Alsace (green)

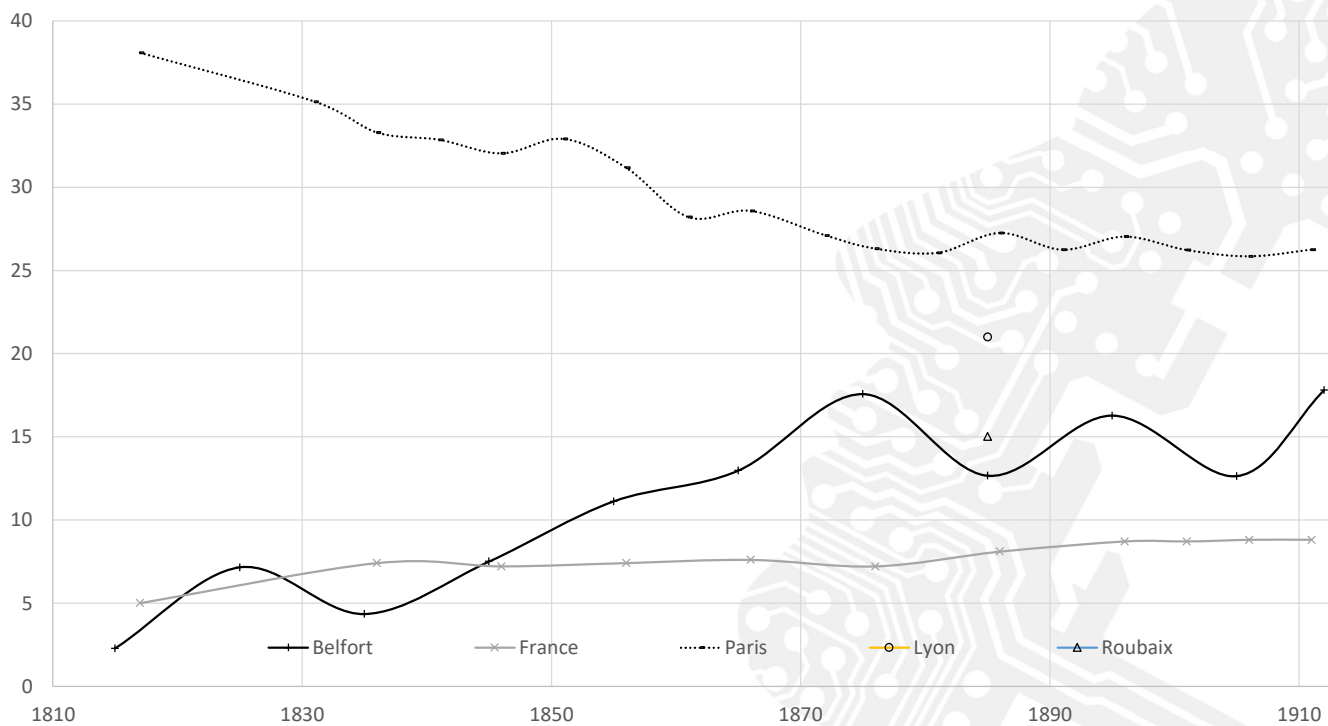


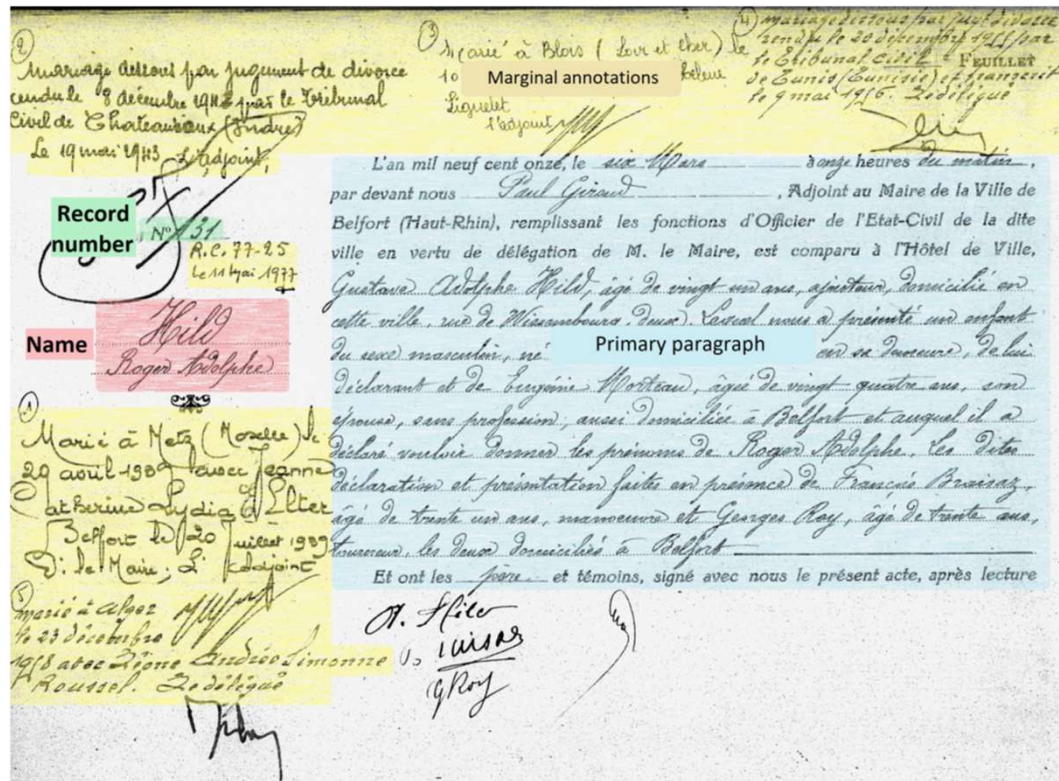
Workers « parisiens »
(Cetty 1881)



- birth and death certificates already scanned by the Belfort municipal archives for the years 1807 to 1919
- approximately 40,000 records for births and 38,000 records for deaths
- 2 random samples of just over 1,000 acts for each of the sets considered

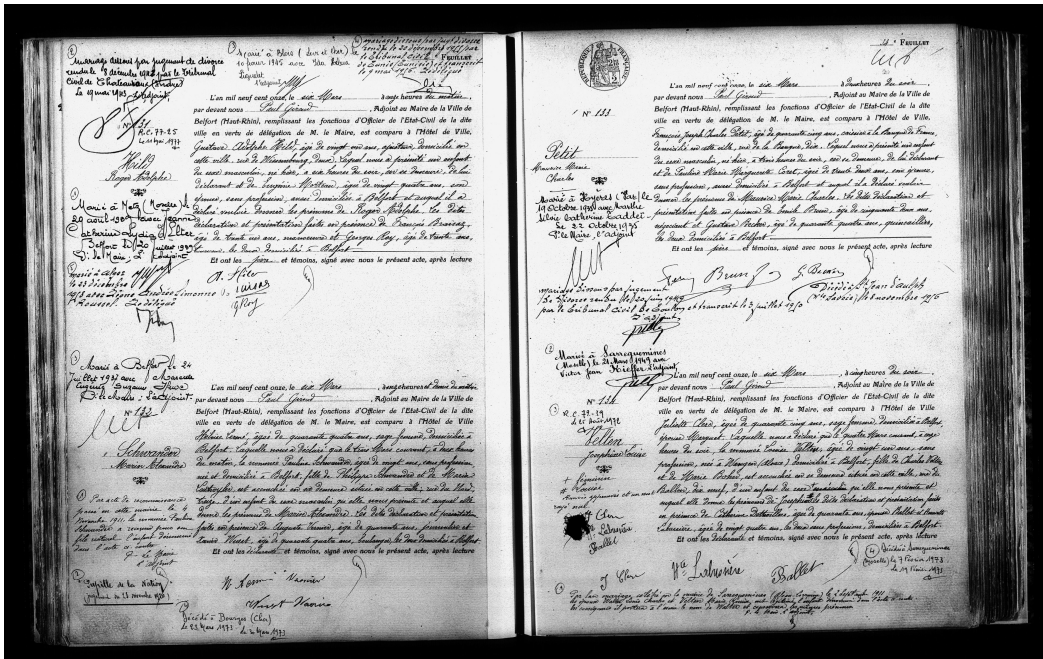
Out-of-wedlock births (%) in France & Belfort, 1815-1914
N = 1010 acts of birth for Belfort (Belpop project 2024, France & Paris: Brée 2022, Lyon: Brunet 2008, Roubaix: Petillon 2006)



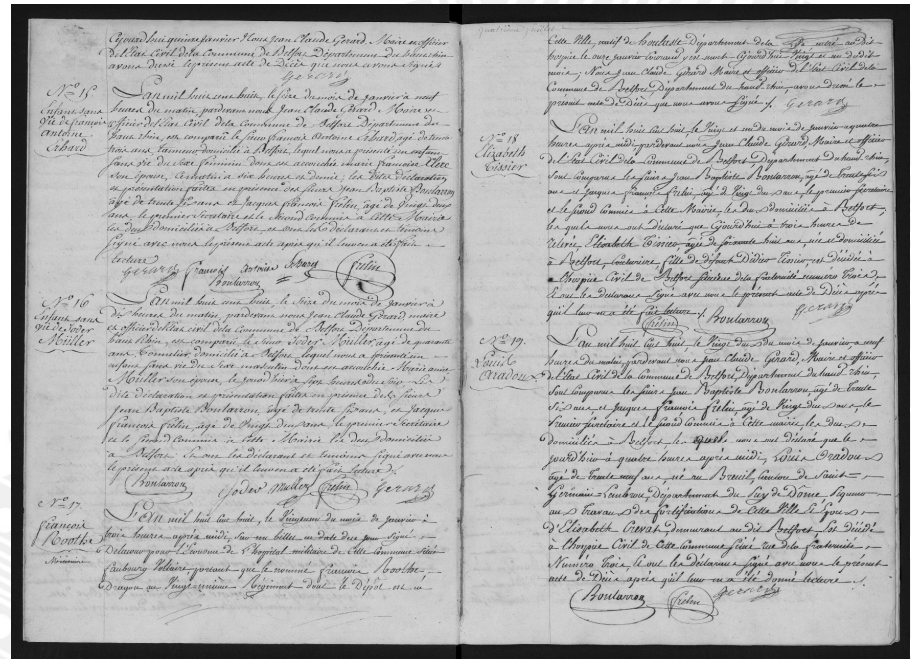


A sample record from Belfort civil registers of birth with annotated components [1]

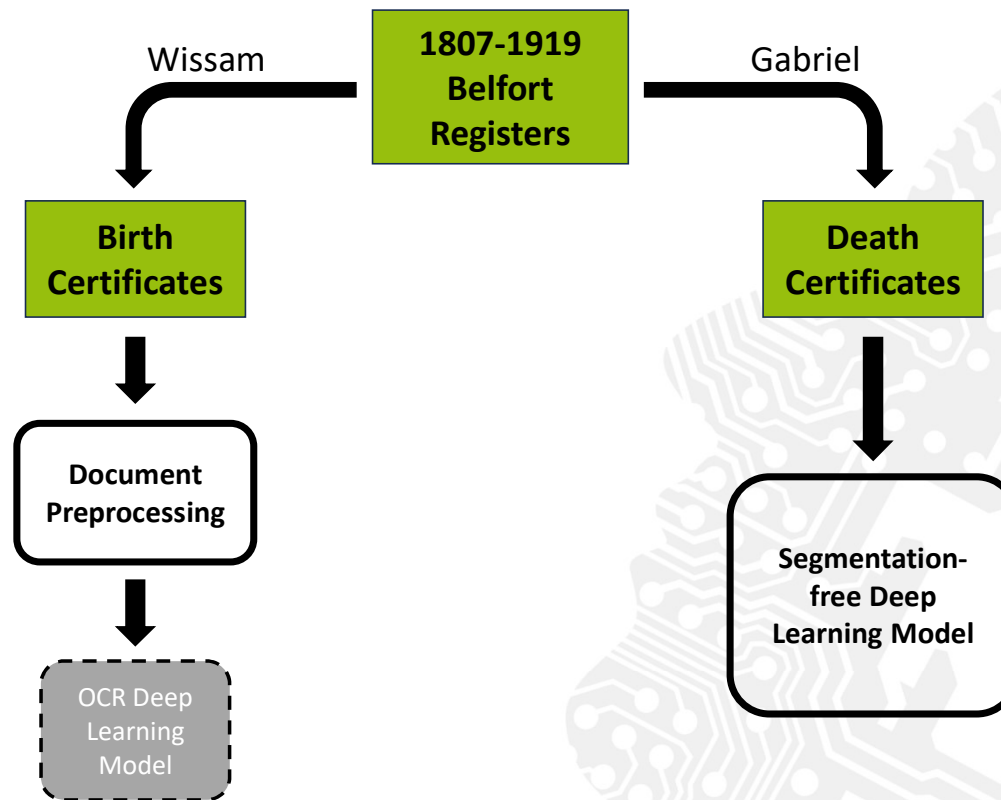
- The complexity of our dataset lies in different aspects



Sample birth record from 1911



Sample death record from 1808



UN BEAU CALENDRIER **Plonk & Replonk**

2022

UNE ANNÉE BERCÉE DE PROMESSES

AVEC UN 13^e MOIS OFFERT

JOURS SANS-CULOTTIDES INCLUS

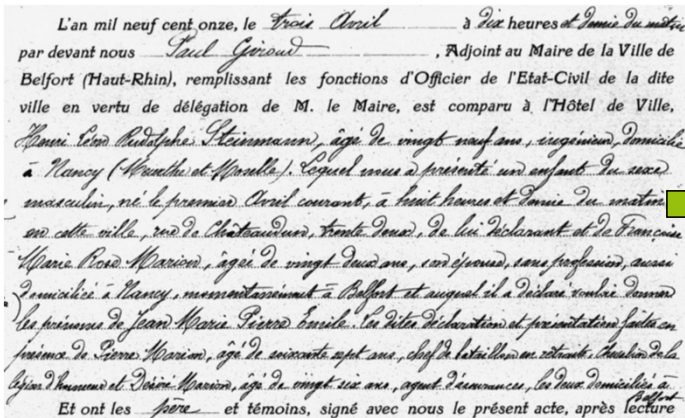
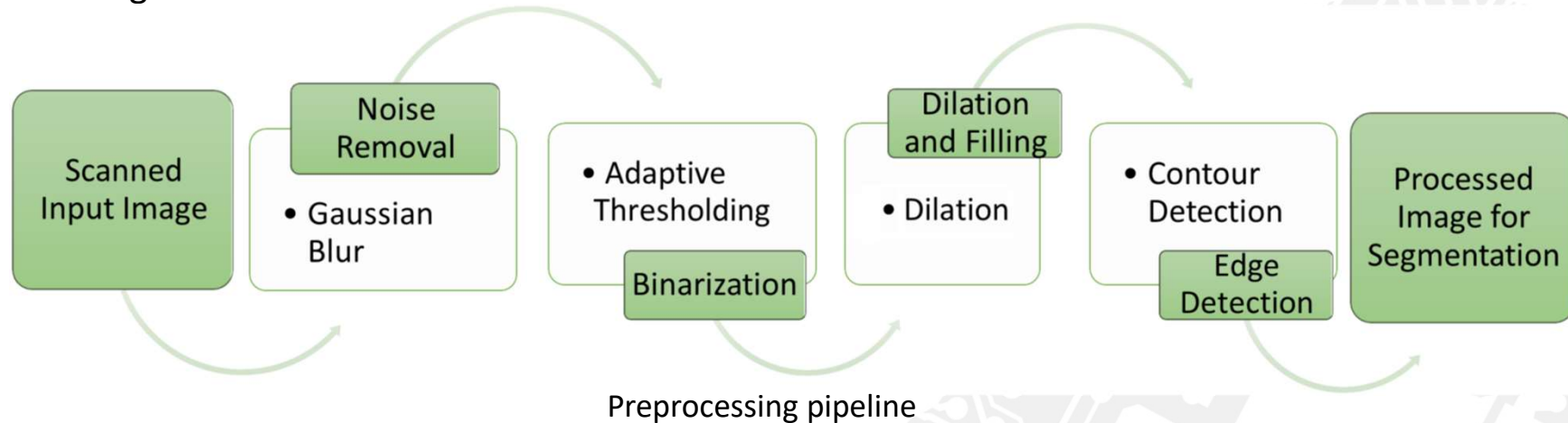
2022

MODÈLE « C'ÉTAIT MIEUX APRÈS » - VALABLE AUSSI EN 2023

1^{er} PRIX EX AEQUO DU CALENDRIER 2022 & 2023

Birth Certificates Transcription

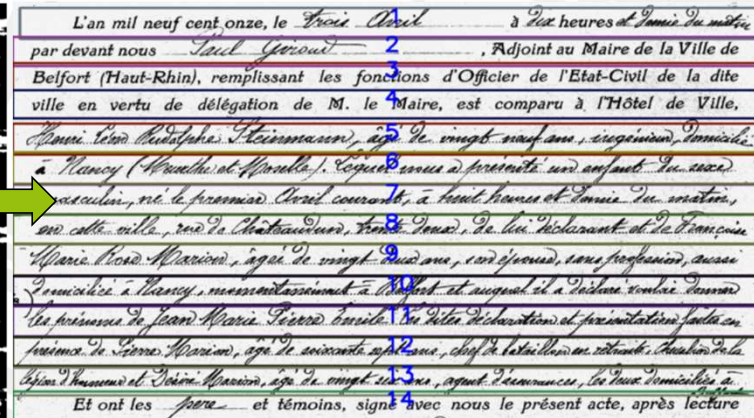
• Preprocessing



Grayscale record



Blurring & Binarization



Contour detection for text line extraction

- Skew correction

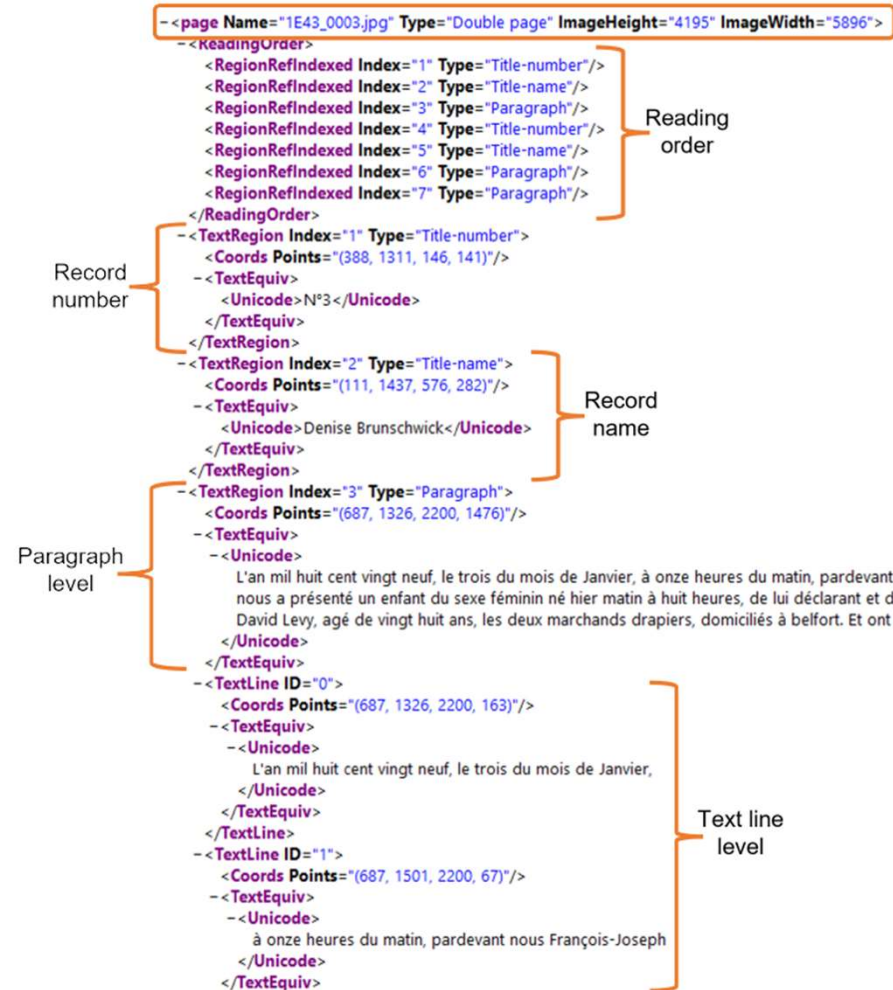
✓ Belfort et auquel il a déclaré vouloir donner le prénom de



✓ Belfort et auquel il a déclaré vouloir donner le prénom de

Skewness correction example [1]

- XML File Generator



Example of a generated XML file

- Metrics and results :

$$IoU = \frac{|\text{predicted mask} \cap \text{ground truth mask}|}{|\text{predicted mask} \cup \text{ground truth mask}|}$$

$$DR = \frac{O2O}{N_{gt}}$$

$$RA = \frac{O2O}{N_{pred}}$$

$$FM = 2 \times \frac{DR \times RA}{DR + RA}$$

| Component | Accuracy | | | |
|----------------------|----------|-----|-------|--------|
| | IoU | DR | RA | FM |
| Primary paragraphs | 93.3% | 97% | 95.2% | 95.9% |
| Marginal annotations | 77% | 83% | 82% | 82.5% |
| Saint Gall | 98.3% | 98% | 98% | 98.04% |
| RIMES | 96.6% | 95% | 94% | 94.4% |

Accuracy results on different datasets

| System | Accuracy (%) |
|-------------|--------------|
| DOCSUMO | 97 |
| Ocelus | 99 |
| Transkribus | 97 |
| Our method | 95.9 |

Accuracy comparison of this method with commercial AI systems

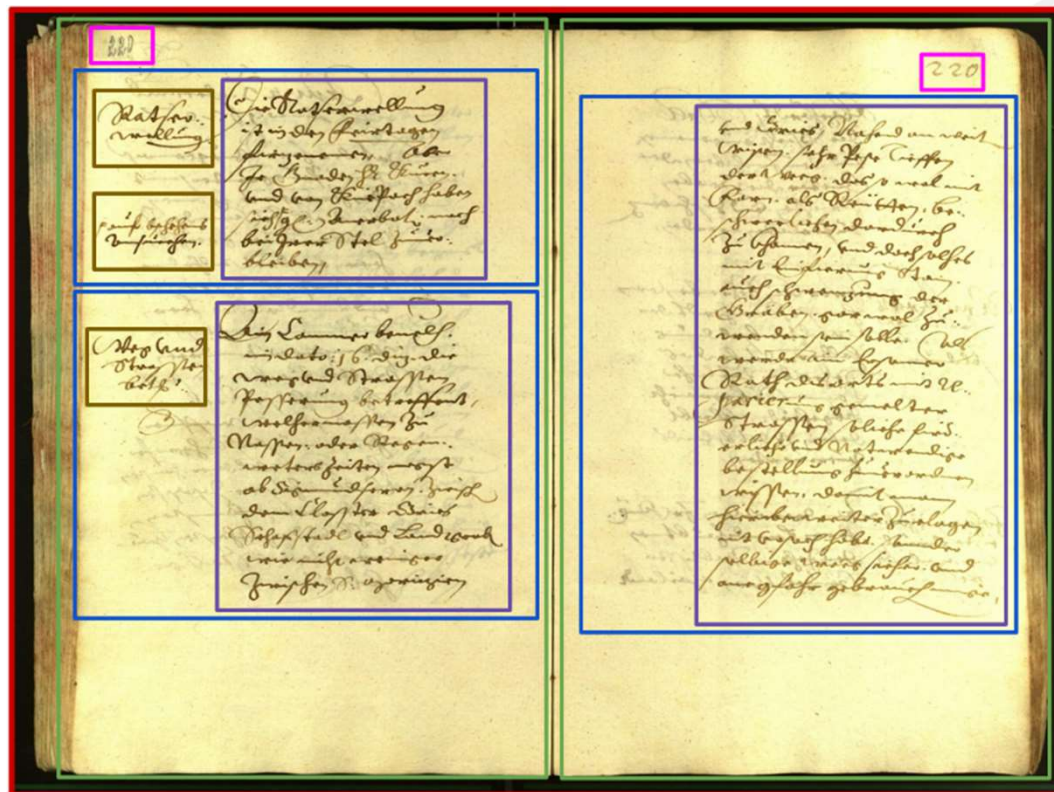
Death Certificates Transcription



Passage en revue du moral des troupes.

Death Certificates Transcription

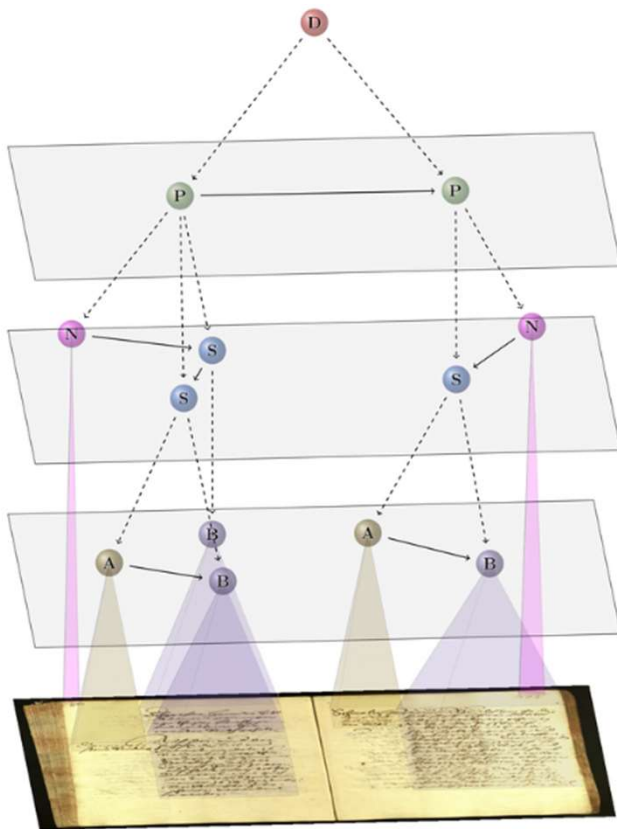
- The Document Attention Network [2] :



DAN segmentation results from READ 2016

Death Certificates Transcription

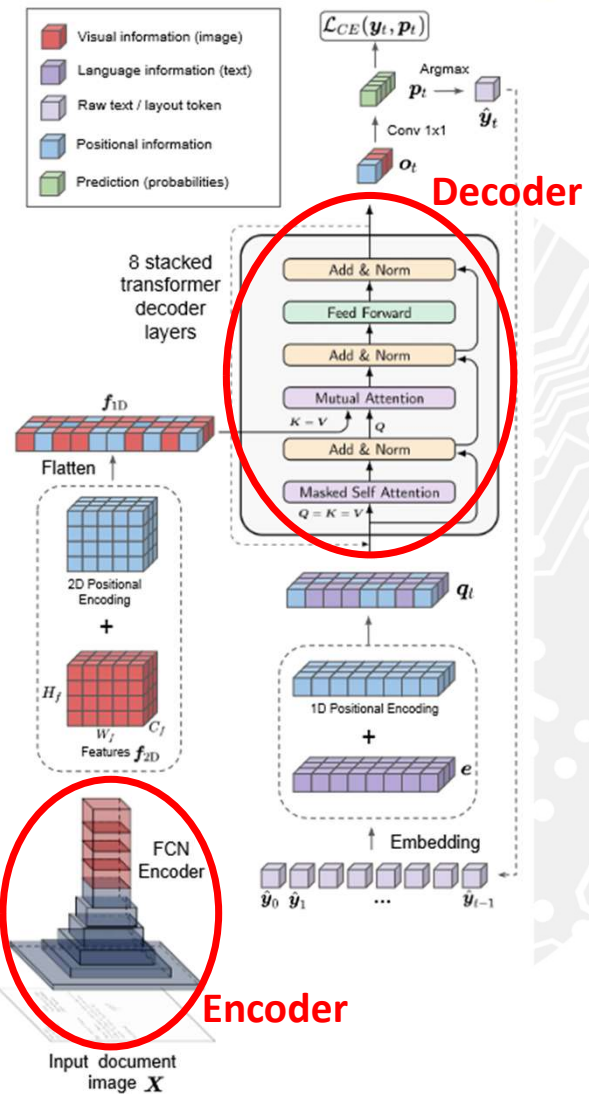
- The Document Attention Network [2]:



```

<document>
<page>
<page_number>
204
</page_number>
<section>
<body>
Schgrafer, [...] gehal
</body>
</section>
<section>
<annotation>
General [...] Raitung
</annotation>
<body>
Aüf den: [...] werden,
</body>
</section>
</page>
<page>
<page_number>
204
</page_number>
<section>
<annotation>
Schmalz. [...] bet:
</annotation>
<body>
Verer [...] dar
</body>
</section>
</page>
</document>
    
```

Ground truth

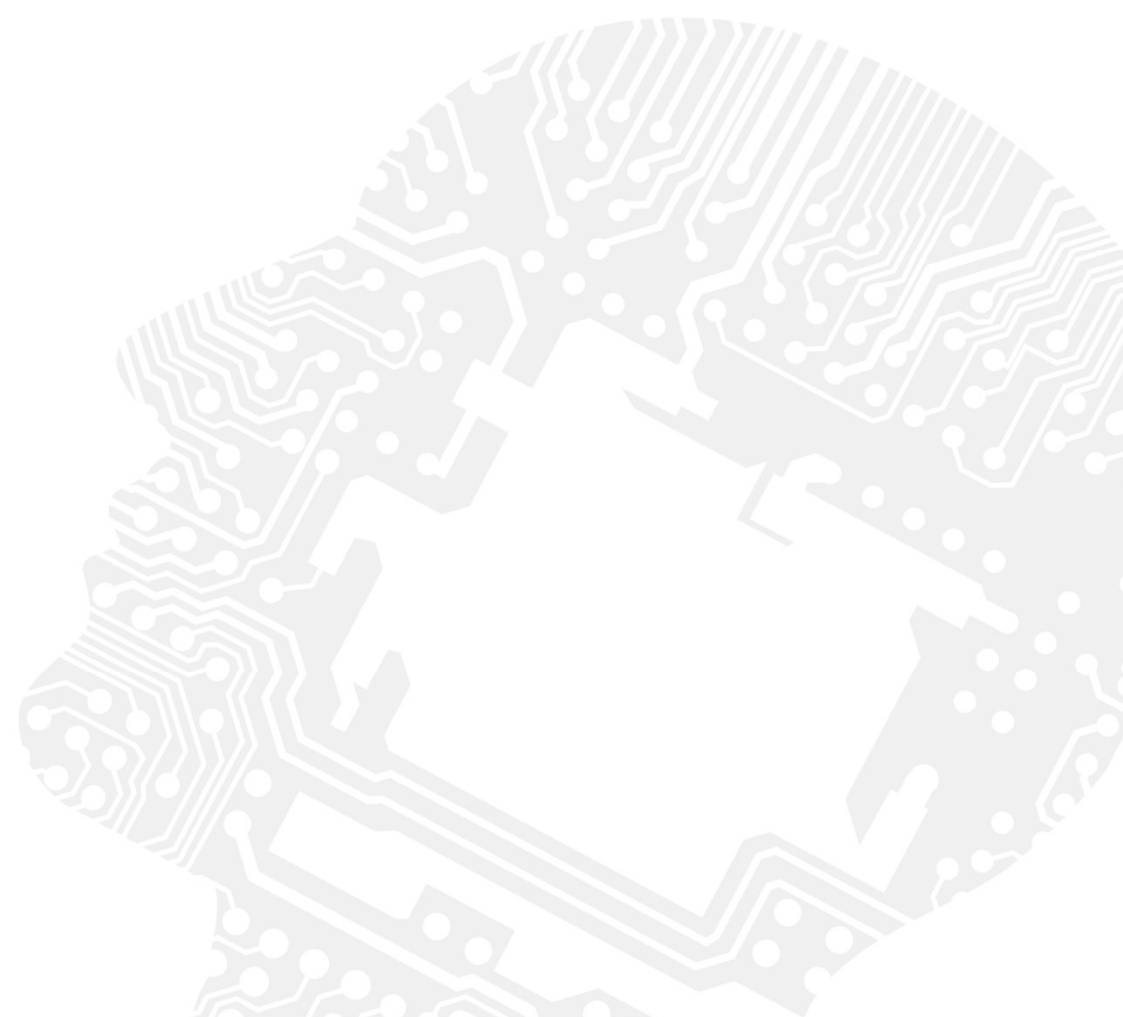
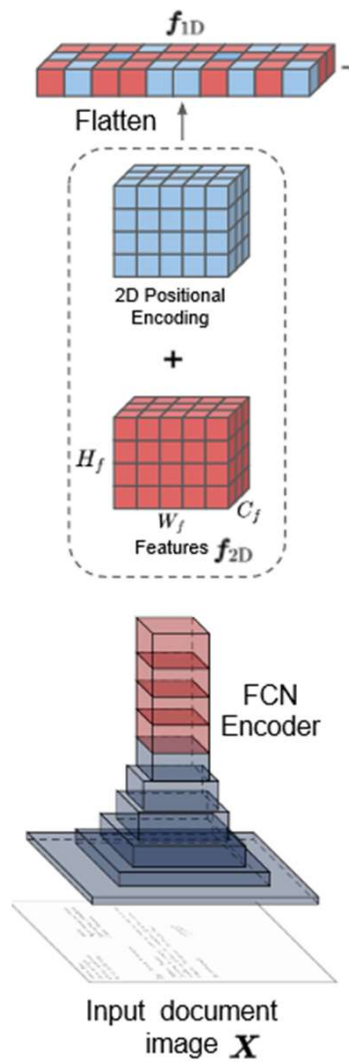


Decoder

Encoder

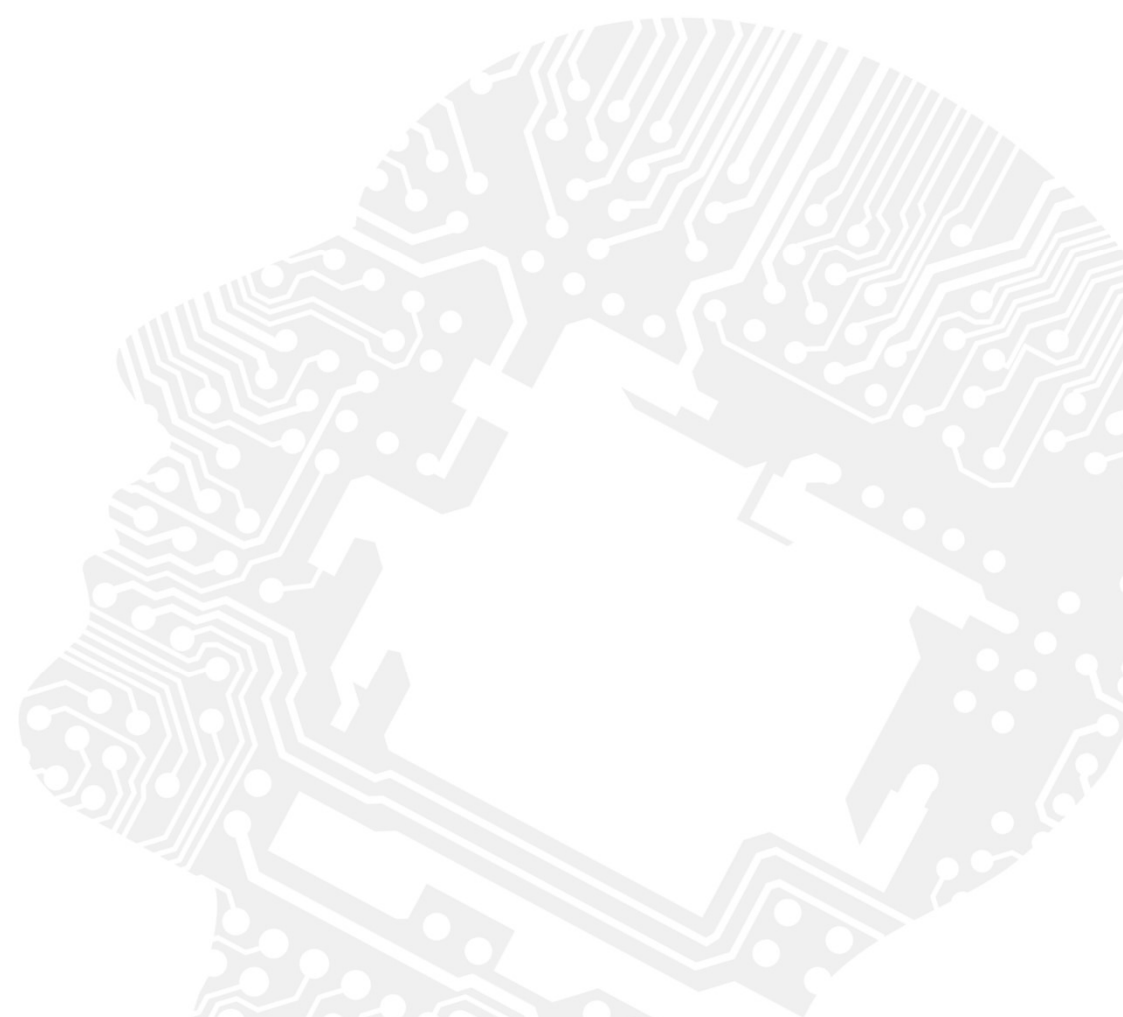
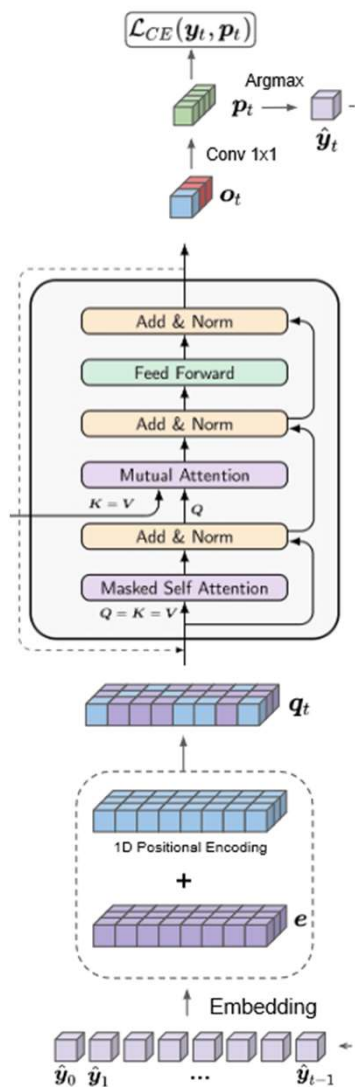
Death Certificates Transcription

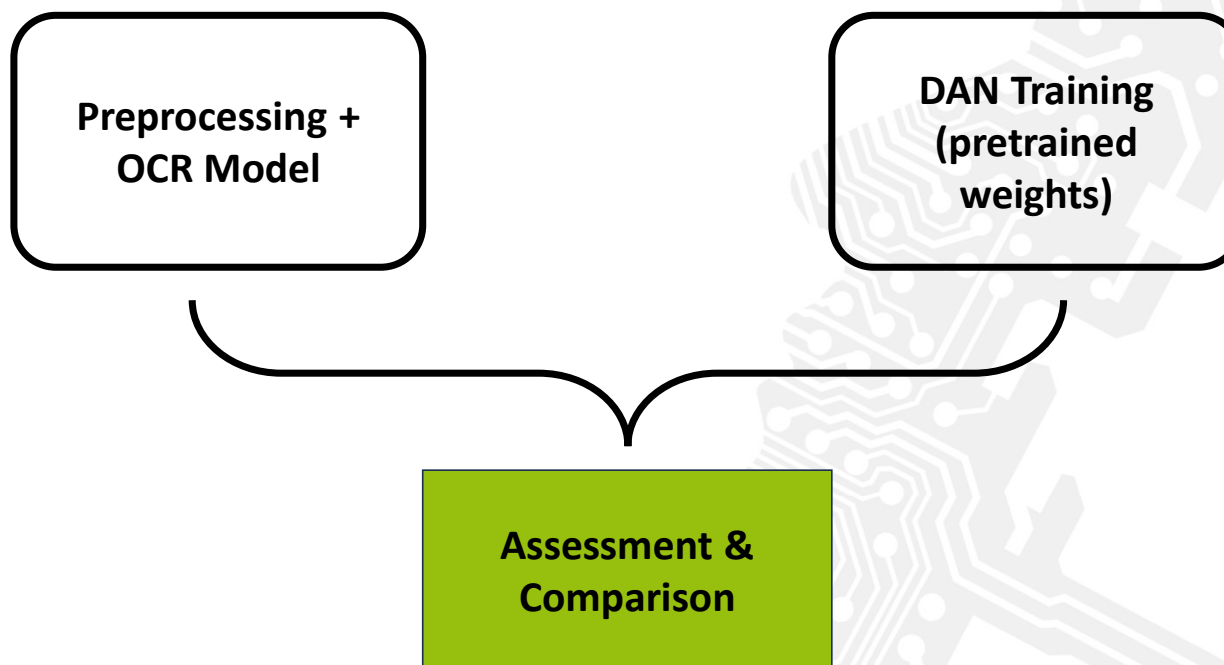
- The FCN Encoder:



Death Certificates Transcription

- The Transformer :





- [1] Preprocessing and Data Structuring Technique for Belfort Birth Records Transcription, W AlKendi, Franck Gechter, Laurent Heyberger, Christophe Guyeux, 2024
- [2] DAN : a Segmentation-free Document Attention Network for Handwritten Document Recognition, D. Coquenot, C. Chatelain, T. Paquet, 2022

Thank you for your attention

Any questions ?

