

From field notes to digital archive: AI-assisted indexing of Bernard Bruyère's excavation journals

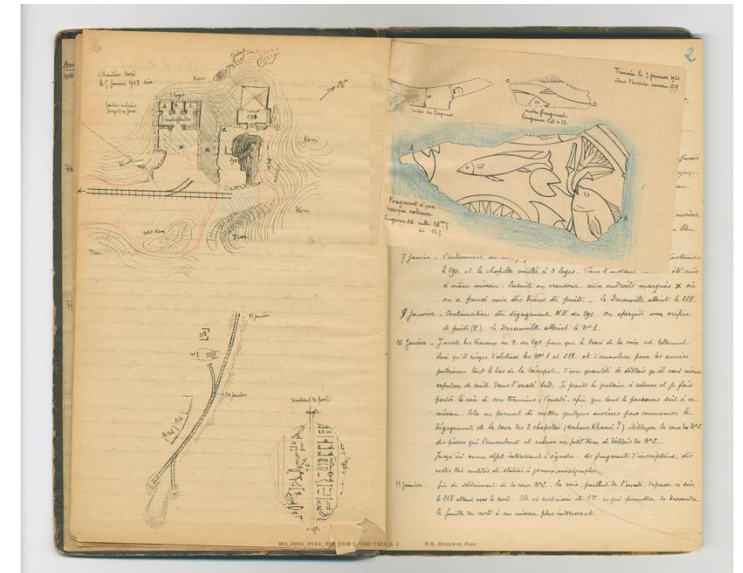
[Christopher Kermorvant](#), TEKLI A, Paris

Cédric Larcher, IFAO, Cairo, Egypte

Bernard Bruyère's handwritten excavation journals

Journals of excavations by Egyptologist Bernard Bruyère from 1922 to 1951 at Deir el-Medina, on the left bank of Thebes. These excavation journals are kept at the IFAO in Cairo.

A lot of information, drawings of objects and surveys of monuments were not included in the published reports and are only available in the handwritten journals.



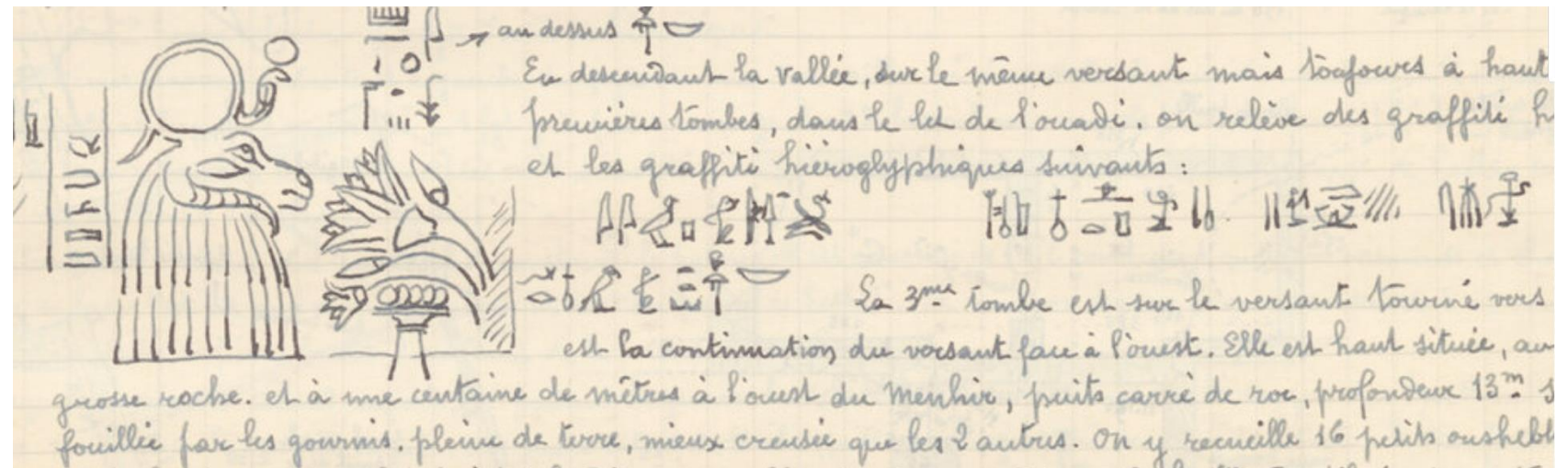
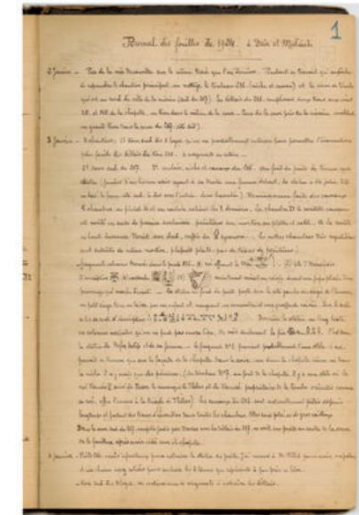
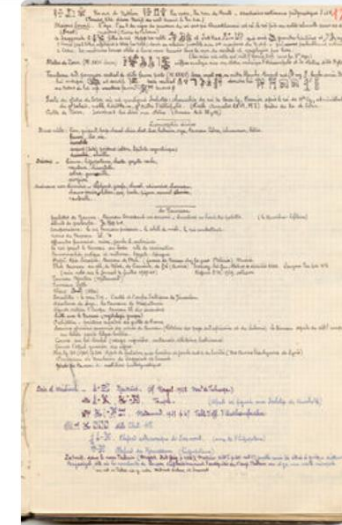
Bernard Bruyère's handwritten excavation journals

4 journals - 891 pages

Very carefully written

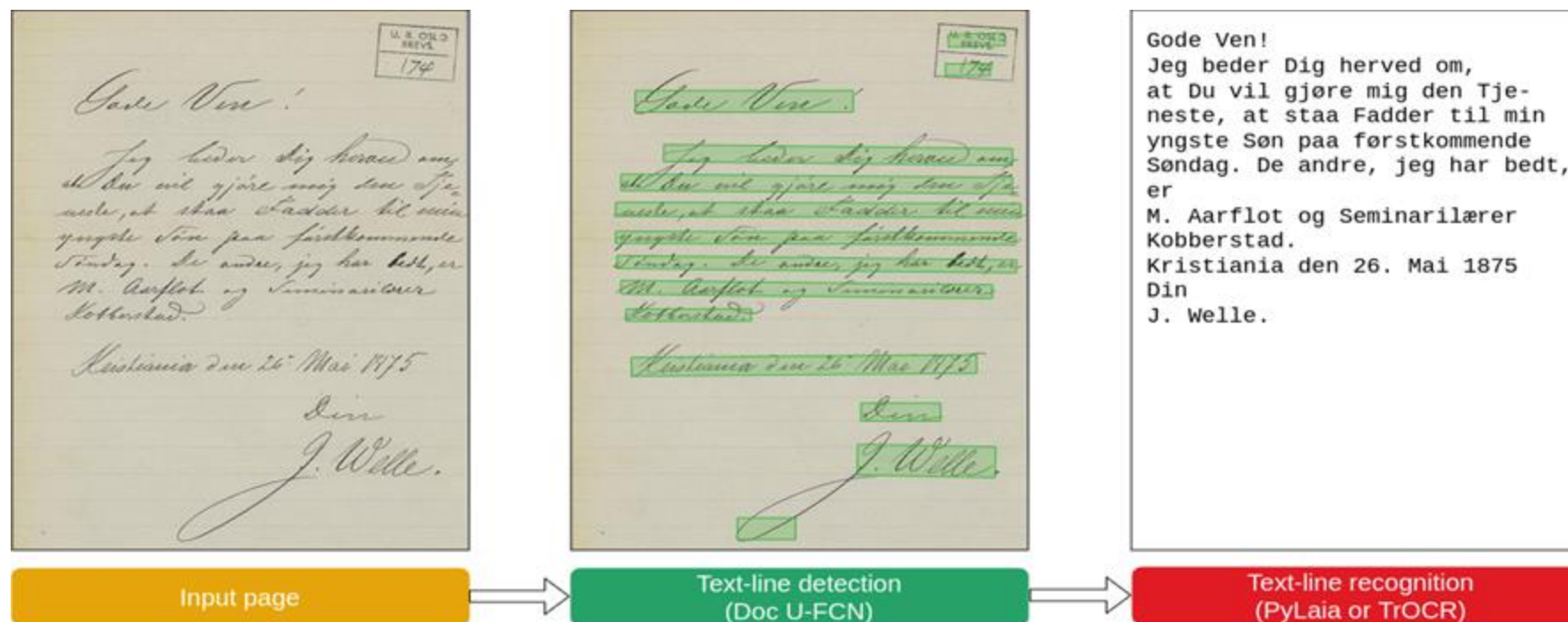
They contain:

- Text in French
- Hieroglyph
- Object drawing
- Maps



HTR Sequential pipeline (until 2021)

1. Text-line detection with Doc U-FCN or YOLO
2. Text-line recognition with PyLaia or TrOCR or Kraken



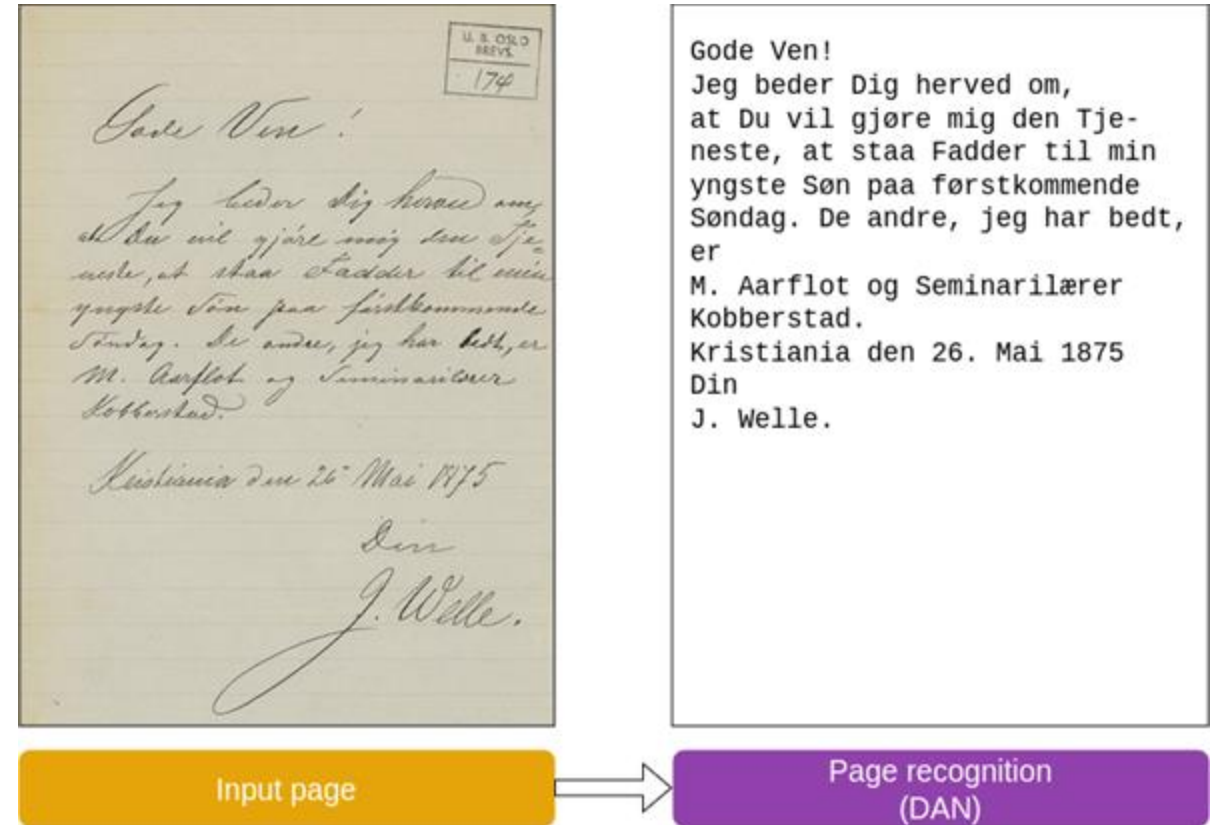
<https://gitlab.teklia.com/atr/pylaia>

End-to-end pipeline (since 2023)

End-to-end page recognition
with DAN (Document
Attention Network) or LLMs

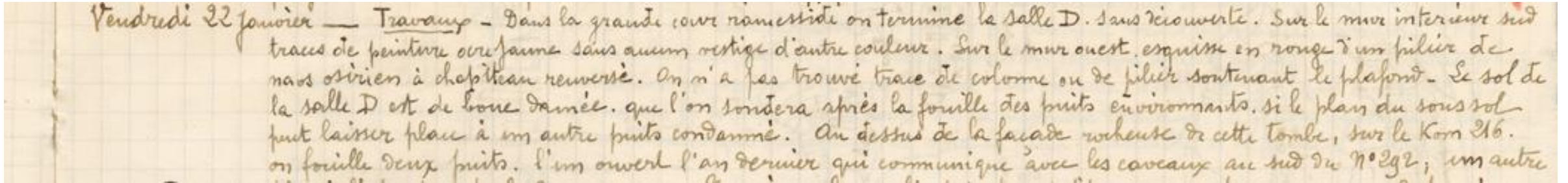
No explicit text detection

Combined detection and
recognition



<https://gitlab.teklia.com/atr/dan>

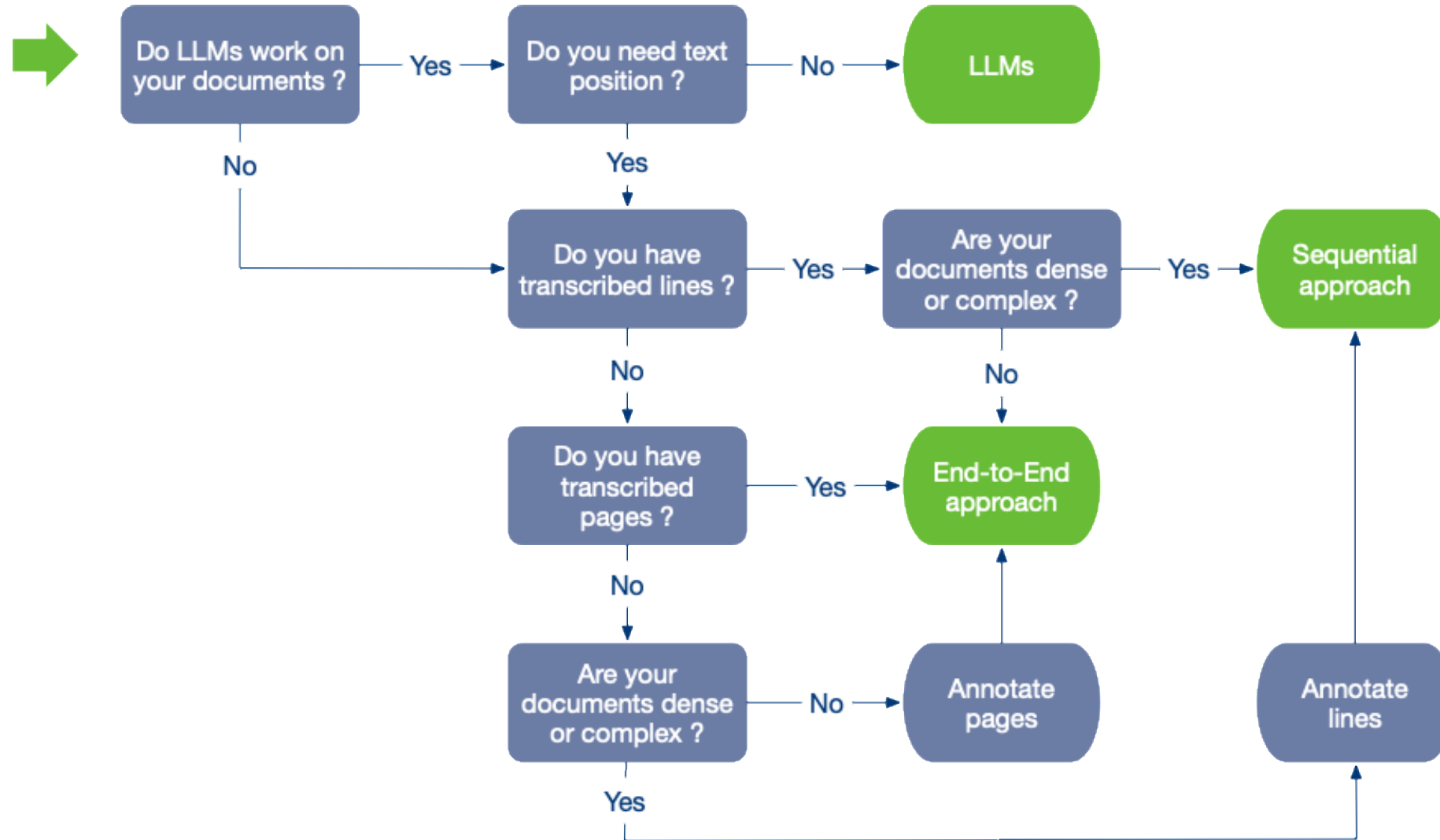
Transcription with LLMs ?



Vendredi 22 janvier - Travaux - Dans la grande cour ramassée on termine le salle D. sans découverte. Sur le mur intérieur des traces de peinture carminée sous un autre reste d'autre couleur. Sur le mur ouest, esquisse en rouge d'un pilier de marbre à chapiteau renversé. On n'a pas trouvé trace de colonne ou de pilier soutenant le plafond. Le sol de la salle D est de bonne hauteur, qu'on sondera après la fouille des puits environnants....

- Pre-trained LLMs (commercial or open-weights) are very good at reading legible handwriting
- They produce very realistic text when illegible (*hallucinations*)
- Better results if the text is split into paragraphs

HTR: which models to use ? (simple version)



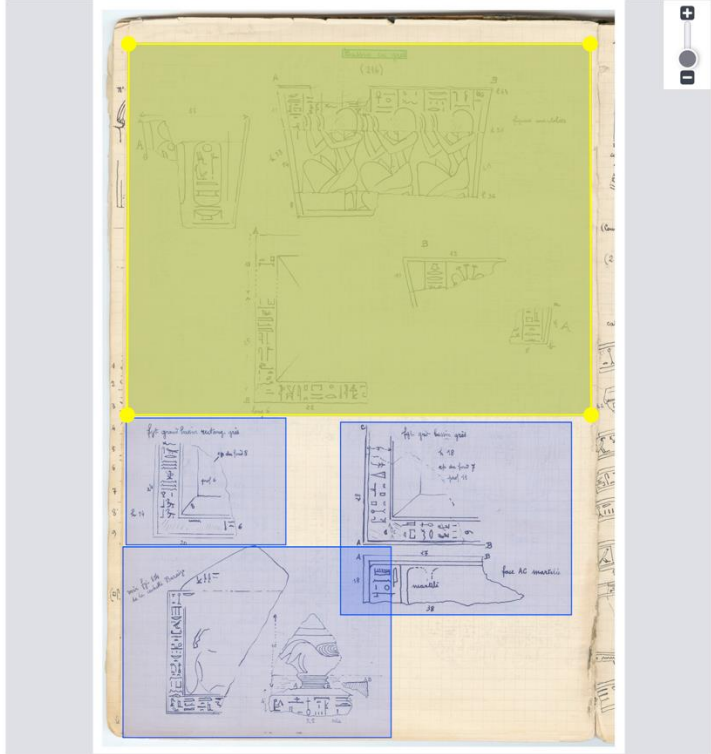
Manuel indexing of the illustrations

Illustration location

Arkindex Projects Process Search 0 kermorvant@tekli.com

IFAO | Journaux de fouilles de... / Folder Journal de fouilles 1939-1955 / Page ms_2019_03777

Filter by type... Hide
Filter by worker run...
Page ms_2019_03777
Text line 1
Illustration ms_2019_0...
Illustration ms_2019_03777_02
Illustration ms_2019_03777_03
Illustration ms_2019_03777_04



View source image List elements on this image

Illustration ms_2019_03777_01
Element ID: a15a68cb...
Created by ZCochin

TOOLS
MOVE TYPE EDITION DELETE
RECTANGLE POLYGON MEDIAN POINT

BATCH ANNOTATION
☐ Enable batch creation (no prompt)

Element type
Illustration

Version 1.8.2 Enterprise Edition Built by Teklia Support Documentation for Arkindex & CLI

Manuel indexing of the illustrations

Illustration meta-data

Arkindex

Projects

Process

Search

0 kermorvant@tekla.com

IFAO | Journaux de fouilles de... / Folder Journal de fouilles 1939-1955 / Page ms_2019_03777

Filter by type...

Hide

Filter by worker run...

Page ms_2019_03777

Text line 1

Illustration ms_2019_0...

Illustration ms_2019_03777_02

Illustration ms_2019_03777_03

Illustration ms_2019_03777_04

Illustration ms_2019_03777_01

Element ID: a15a68cb...

Created by ZCochin

ANNOTATE

CLASSIFICATIONS

Manual

Bassin

100%

Add a classification

METADATA

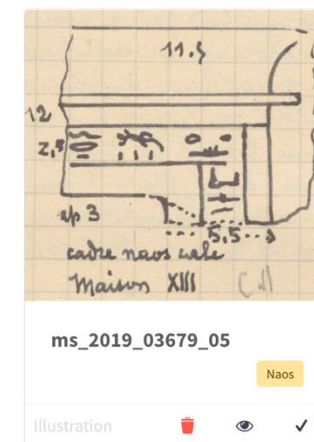
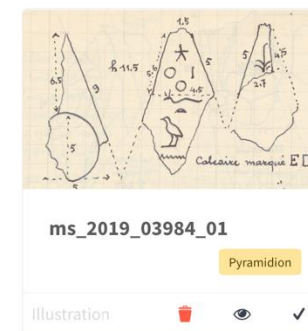
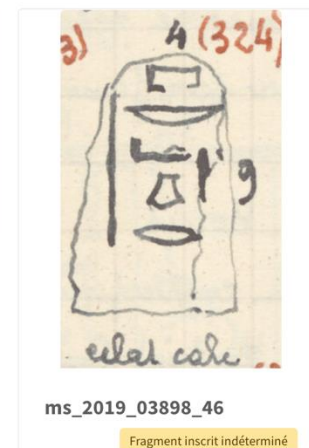
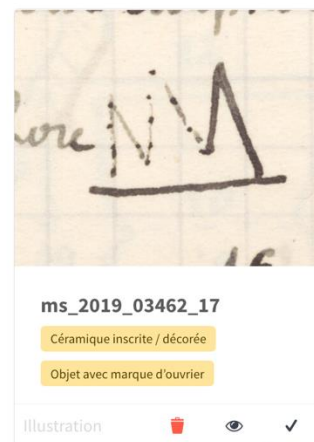
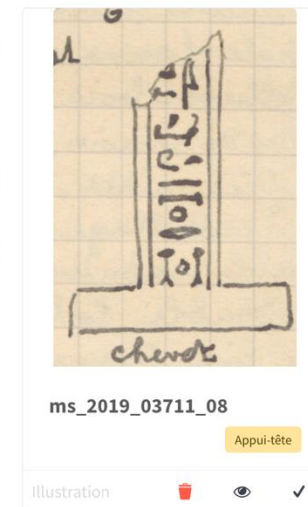
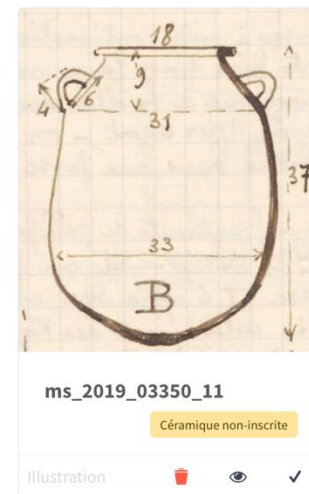
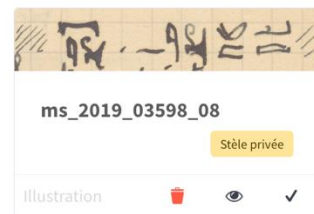
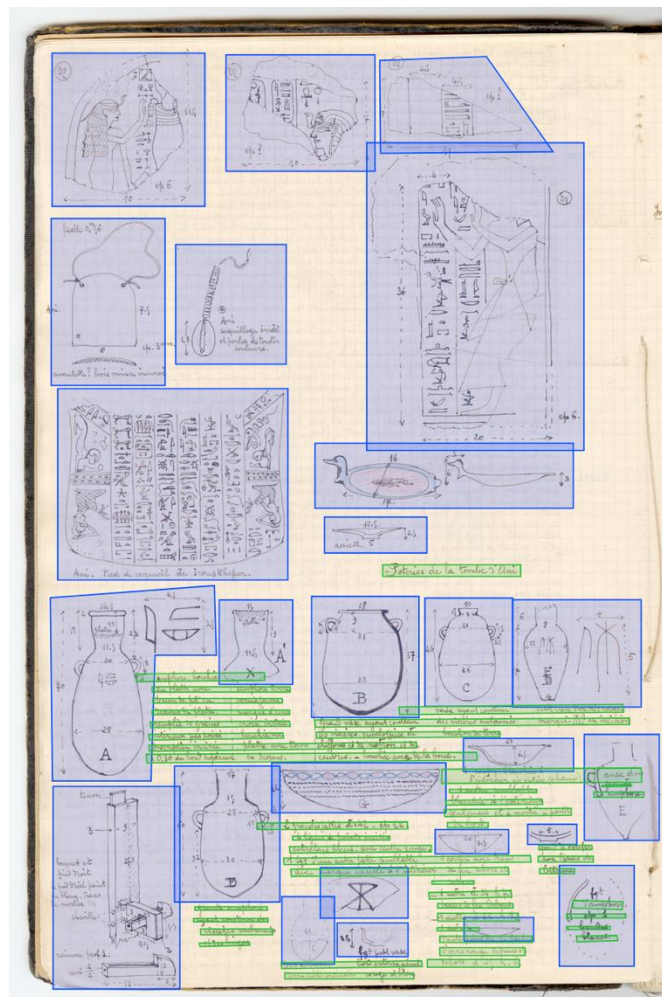
	(216)	
Bruyère_inventory_number		
date of find	1939, March 3	
found_location	Temple ptolémaïque, enceinte nord, chapelle ramesside	
illustration_name	(216)	
owner	Kanouro	
owner	Ouserhat	

DATASETS

[View source image](#)

[List elements on this image](#)

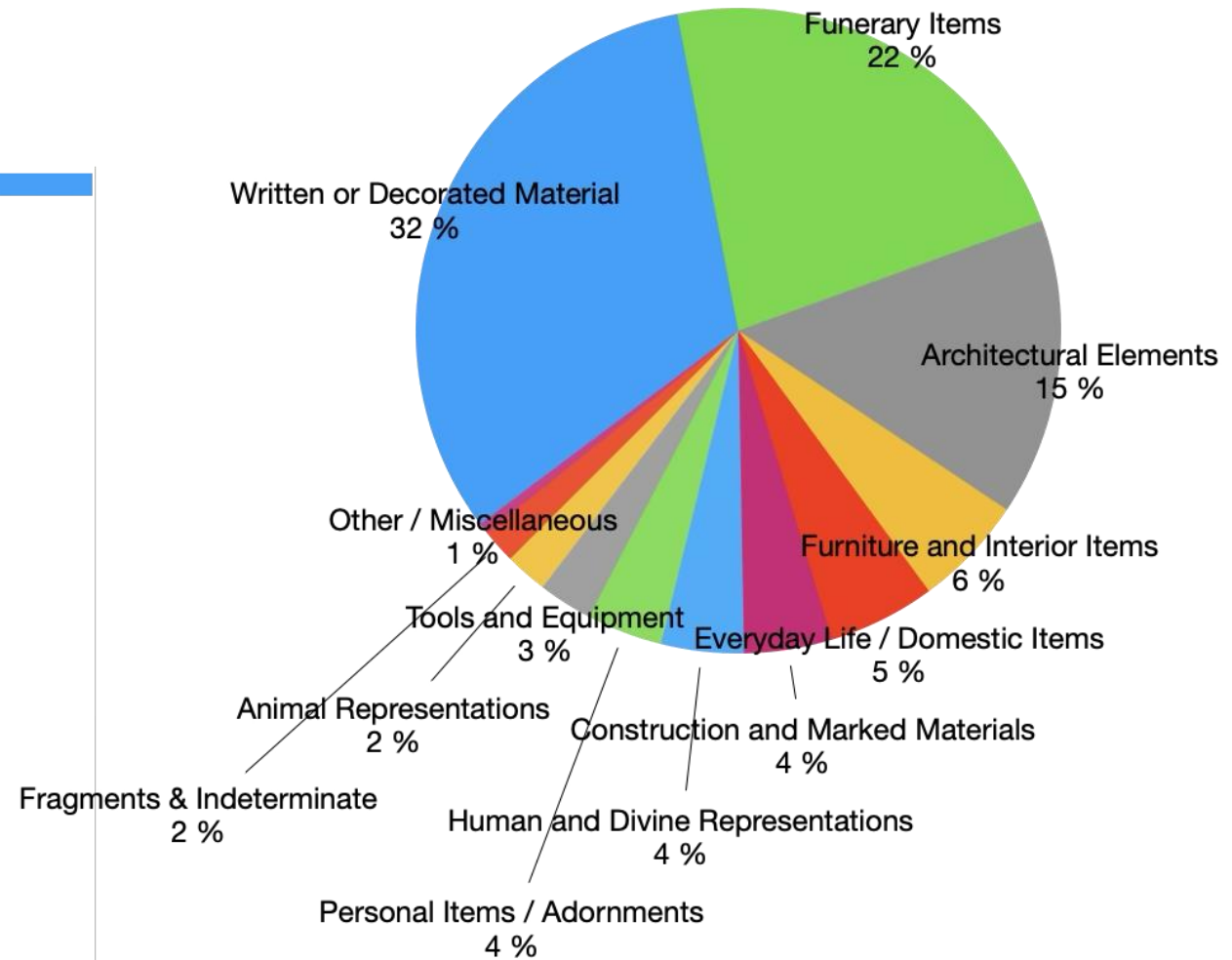
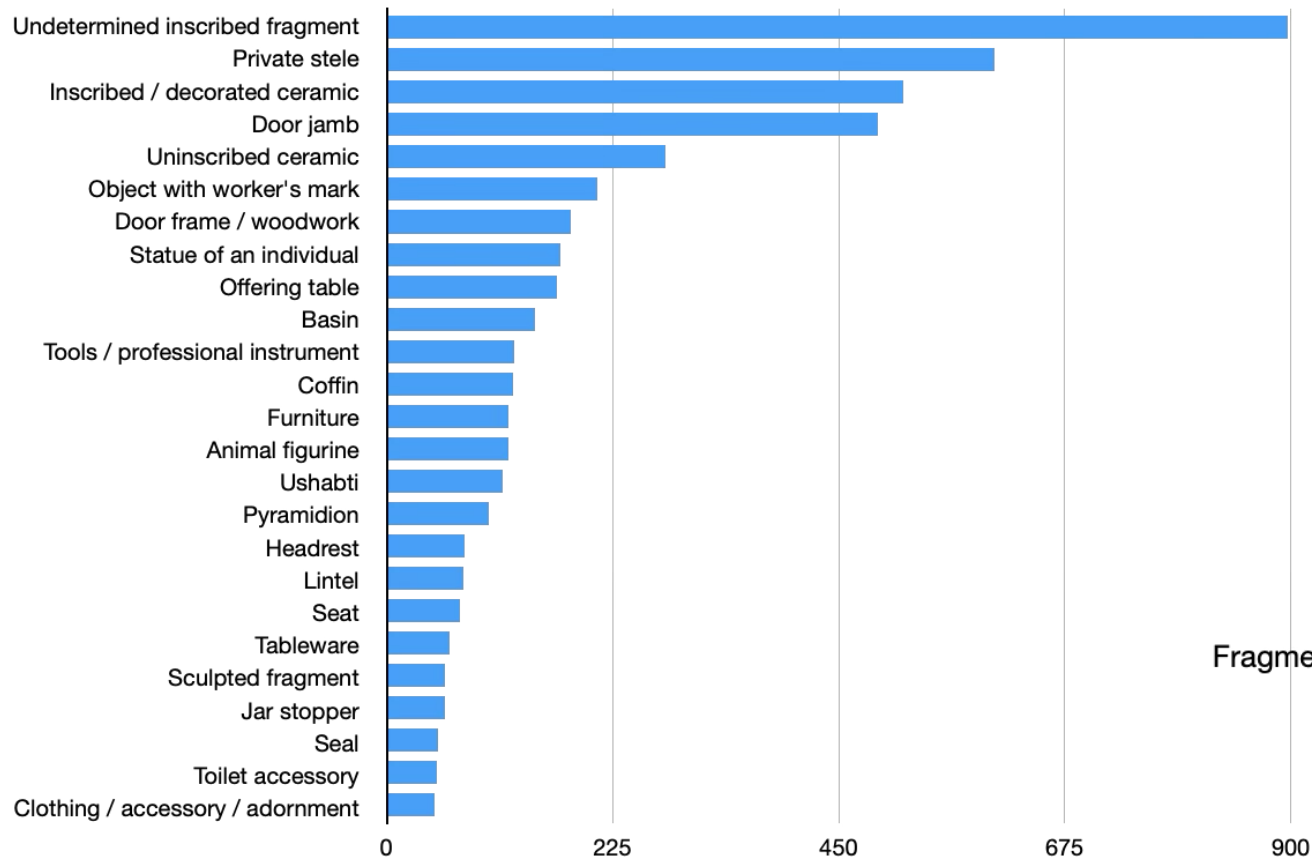
Analysis of the illustrations



Analysis of the illustrations

5479 illustrations

62 different classes of illustrations



Analysis of the metadata on illustrations

11551 Metadata :

- 748 inventory number
- 4186 dates of find
- 4838 found locations
- 1138 owner name
- 204 illustration names
- 438 other names

Arkindex Projects Process Search 0 kermorvant@tekla.com

IFAO | Journaux de fouilles de... / Folder Journal de fouilles 1939-1955 / Page ms_2019_03785
/ Illustration ms_2019_03785_02

← 4 → Elements Processes Display

Illustration ms_2019_03785_02 [ANNOTATE](#)
Element ID: 11acdafa...
Created by ZCochin

ORIENTATION

CLASSIFICATIONS
Manual
Stèle privée 100% [✓](#)

Add a classification

TRANSCRIPTIONS

NEW TRANSCRIPTION

METADATA

date of find	1940, January 5	✎ ✖
found_location	P_1422	✎ ✖
illustration_name	(339)	✎ ✖
owner	Amennakht	✎ ✖

[+](#)

DATASETS

View source image List elements on this image

Version 1.8.2 Enterprise Edition Built by Teklia [Support](#) [Documentation for Arkindex & CLI](#)

Combined search on text and meta-data

Arkindex

Projects

Process

Search

0

kermorvant@tekla.com

Search in project

IFAO | Journaux de fouilles de Bernard Bruyère

*1422

Element

Transcription

Metadata

Entity

Classification

Sort by

Relevance

Element Type

Sort by occurrences

Illustration (8)

Text line (8)

Element Worker

doc-ufcn inference (8)

Metadata Name

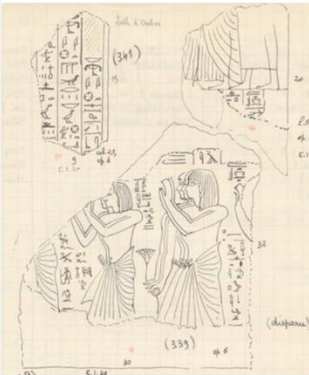
found_location (8)

Metadata Type

location (8)

Items 1 to 16 out of 16 results

Page ms_2019_03785 /
Illustration ms_2019_03785_02



found_location : P_1422

VIEW ELEMENT

Page ms_2019_03850 / Text line 11

sud du sanctuaire M. un petit groupe de maisons
sur lequel s' enfonce le puits funéraire P 1422 à

100%

VIEW ELEMENT

Page ms_2019_03788 / Text line 7

Dimanche 7 Janvier. Travaux. On achève la hutte
des ghafir, on continue la fouille devant le
temple [...]

100%

VIEW ELEMENT

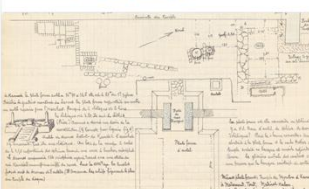
Page ms_2019_04006 / Text line 47

(1422) (4) 6 c. n. (75) : (6) (7)

100%

VIEW ELEMENT

Page ms_2019_03787 /
Illustration ms_2019_03787_08



found_location : P_1422

VIEW ELEMENT

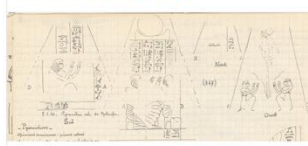
Page ms_2019_03786 / Text line 79

On dégage au dessus de 1422 des restes de
constructions de 2 époques superposées : murs ,
sol dam[...]

100%

VIEW ELEMENT

Page ms_2019_03787 /
Illustration ms_2019_03787_01



found_location : P_1422, au nord de

VIEW ELEMENT

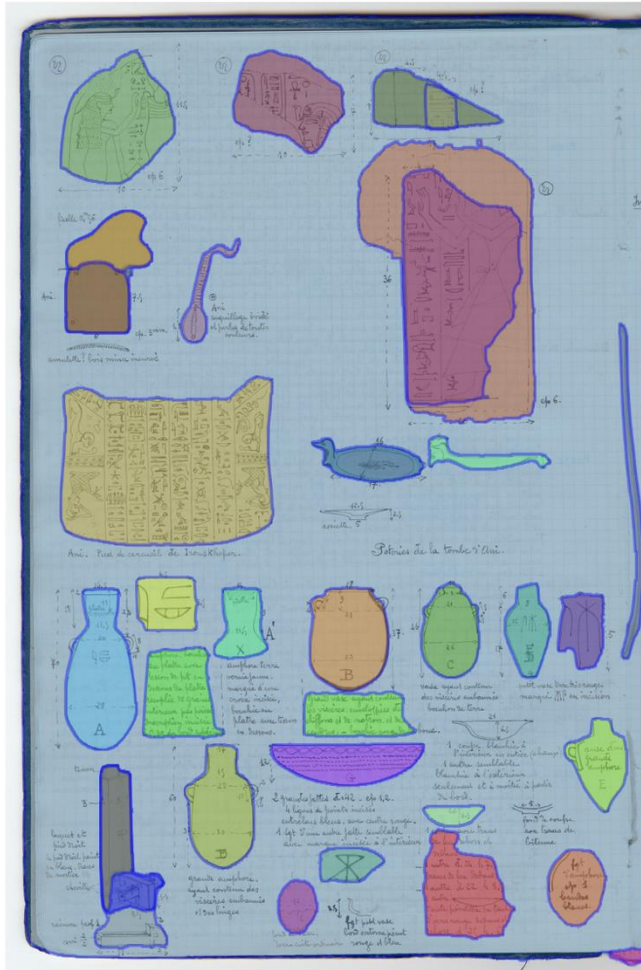
Page ms_2019_03786 / Text line 65

Trouvailles - P.1422 - 2 fgts calc. stèle , 2
registres défilé de famille ramesside . noms .
Ne[...]

100%

VIEW ELEMENT

What about automatic illustration segmentation?



- No training, but not specific to documents
- Good results for a first approach when the segmentation does not need to understand the content
- It needs a very ergonomic interface to correct the predictions (correction, merging, splitting, etc)

Segment Anything (Meta, 2023)

Arkindex: a generic document processing platform

Customization



Process all types of documents

Scaling up



Process 1000 or 10 million pages

Open-source



Dissemination and community participation

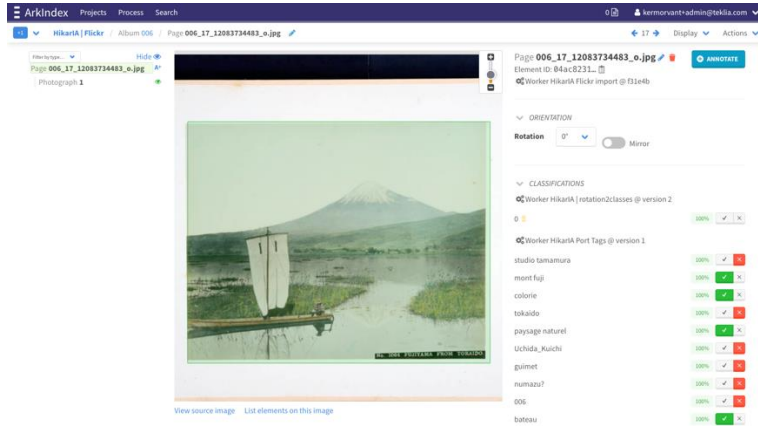
<https://gitlab.teklia.com/arkindex>



TEKLIA's open-source software suite for document processing

Arkindex

Document processing



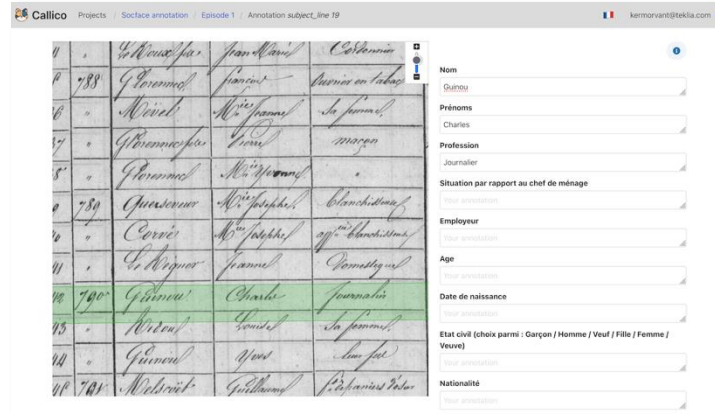
Open-source and Enterprise
licences

<https://gitlab.teklia.com>

<https://support.teklia.com>

Callico

Collaborative annotation campaigns

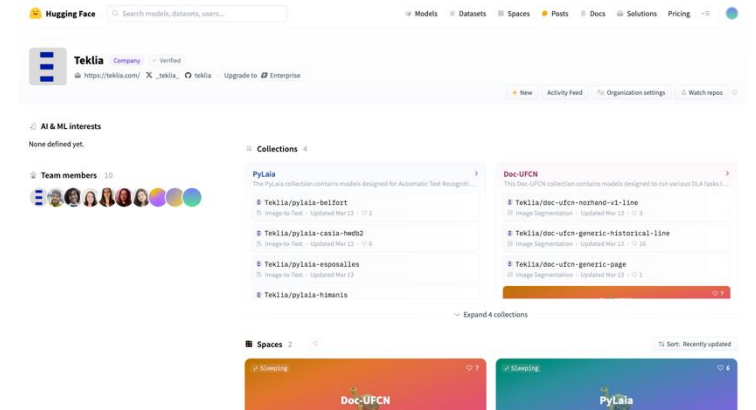


Open-source

kermorvant@teklia.com

HuggingFace

Models and datasets



Open-source, open-weights

<https://huggingface.co/Teklia>