

# Group Projects

---

Benjamin Kiessling

October 5, 2023

# Group Projects

The grade for this class will be based on a group project.

The project allows you to apply what you've learned during the course (and learn new things by yourself).

You can form groups of up to 4 people to work jointly on a single problem.

Workload in a group should be roughly equal. This doesn't mean that everyone has to participate in every single task.

# Group Projects

Pick a topic/problem (find one yourself or choose from the list I brought today) that interests you.

Your topic should consist of a humanities research question that is to be answered with computer vision methods.

It can be something that is already known! Getting to already established results with new methods is valid.

## Group Projects - Structure

Grade based on a report (ca. 15 pages) and code (a few hundred lines).

Report should include analysis of problem, methods, and results.

Reports can be written in English or French.

Code can (and should!) be adapted from other projects.

If you're writing a lot of code, a shorter report is acceptable.

# Group Projects - Structure

There are two intermediate presentations.

They are not part of the grade but to give feedback on the progress and sanity checking.

First one is roughly 30 minutes of problem statement, data sources, research question, methods, ...

There will be extended discussion afterwards.

## Group Projects - Structure

Second presentation is intended to give an update on the current state of your project and present results and/or preliminary analysis.

After the 2nd presentation everything should be clear with only minor questions remaining.

Ideally the remaining time after the 2nd presentation is spend on analysis of obtained results and writing the report.

# Group Projects - Timeline

13/10/2023 - Presentation of rough proposals!

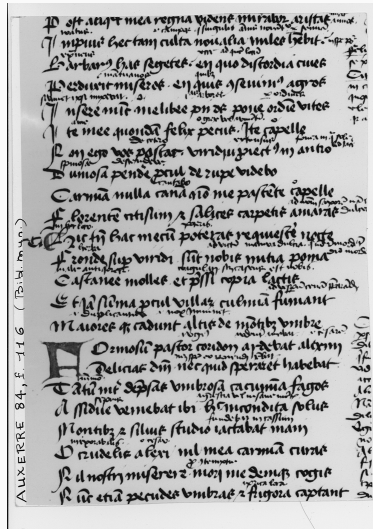
27/10/2023 - Presentation of intermediate results/progress report

early January - Deadline for final report + code

# Ideas - Classification of Medieval Manuscripts

The CLAMM dataset consists of 5000 pages of medieval handwritten books dating from 500 CE to 1600 CE.

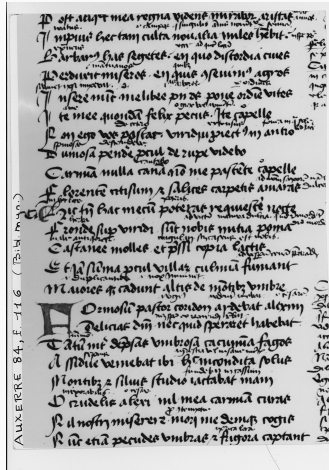
Each page is labelled according to style. There are 12 different styles. In addition, each page is dated.





## Ideas - Classification of Medieval Manuscripts

Either build a classification system that determines the style of page images or an automatized dating method.



There is a dataset is available at  
<https://clamm.irht.cnrs.fr/icdar-2017/icdar2017-clamm/>.

## Ideas - Linear B Line Drawings

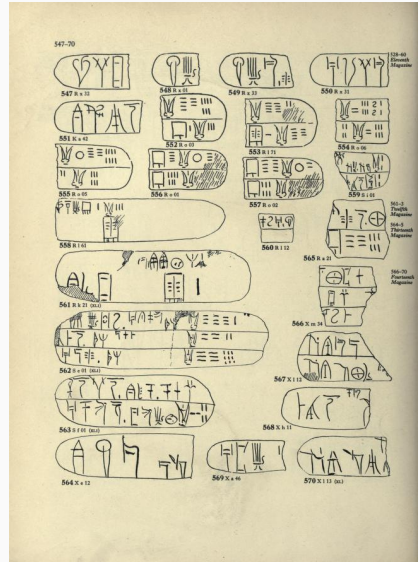
Archeological illustration is often used to provide experts and the public an encoded representation of the original artifact.

Historically largely used because of the limits of photography and reproduction technology it remains a valuable tool.

# Ideas - Linear B Line Drawings

Linear B is the earliest writing system for the Greek language. It is a syllabic script with ca. 200 different characters.

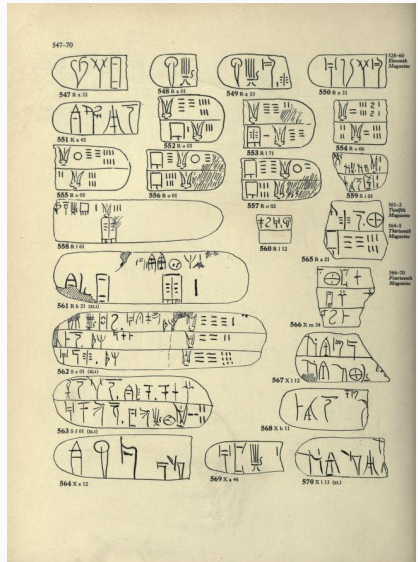
It was largely used for administrative records on clay tablets and the majority of texts are quite short.



# Ideas - Linear B Line Drawings

Build a system converting photographic images into line drawings of Linear B tablets. Both low-level and high-level methods could be used.

As there is no ready dataset you will have to collect the image data first.



# Ideas - Collection Color Analysis

IIIF allow harvesting of image data quickly. The Biblissima project aggregates a large number of IIIF data sources, including many illuminated manuscripts over the centuries.

Sample the corpus and perform color analysis on the illumination across time and origin of the material.

The data is accessible through Biblissima

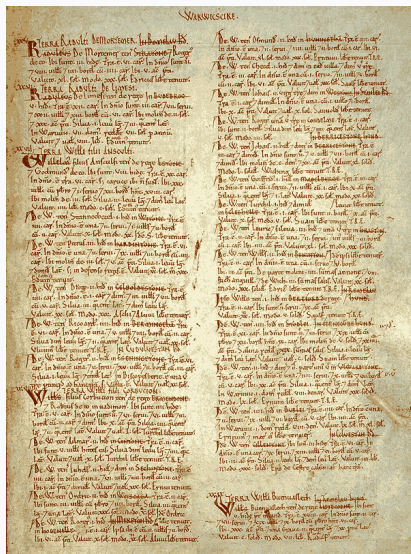
(<https://portail.biblissima.fr>).



Source: gallica.bnf.fr / Bibliothèque nationale de France, Département des manuscrits, f. 117, 192

# Ideas - Codicological Classification of the Domesday Book

The Domesday Book is a survey of much of England and Wales commissioned in 1086 by William the Conqueror. There is extensive codicological information for a part called the Exon Domesday available. This includes human readable information on the parchment characteristics. Build a classification system for parchment types from this information.



# Ideas - Paleographic Analysis of Initials

Drop caps or initials are decorated letters at the beginning of pages or paragraphs.

The Bibliothèques Virtuelles Humanistes has a large collection of renaissance printed works of central France available as facsimiles.

Analyse the initials and try to find patterns with the aid of the associated metadata.

The library is available at <http://www.bvh.univ-tours.fr>

