



UNIVERSITÉ
CAEN
NORMANDIE



The Digital Humanities Turn

Morgane PICA & Pierre LARRIVÉE

Unicaen, CRISCO (EA4255)



The Digital Humanities Turn

I. Introduction

1. The ConDÉ project
 2. "Digital Humanities"?
 3. Digital Humanities "turn"?
-

II. Digital Humanities, in practical terms

1. Digital Humanities: what for?
 2. What kind of work exactly?
-

III. What's in store?

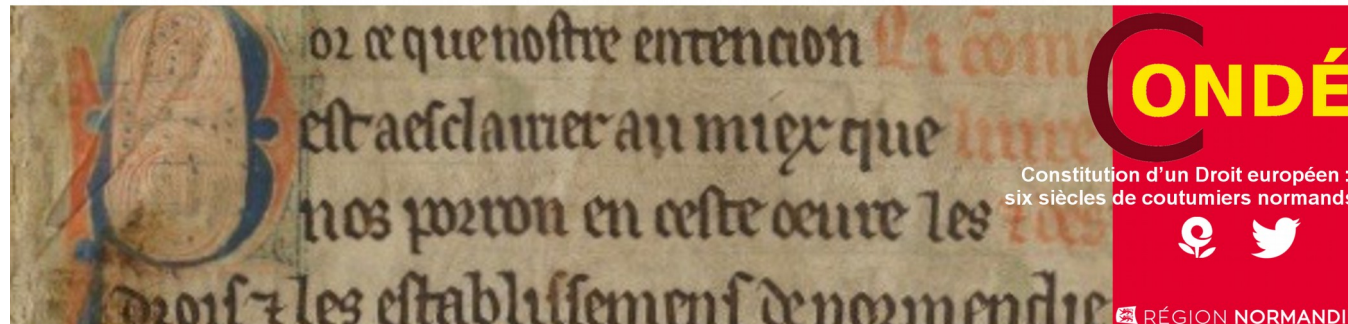
1. The D.H. in the ConDÉ project
2. New developments and expectations
3. How to get started, if you're willing

1. The ConDÉ project

Thanks to the Normandy region

The constitution of a European Law: the ConDÉ project

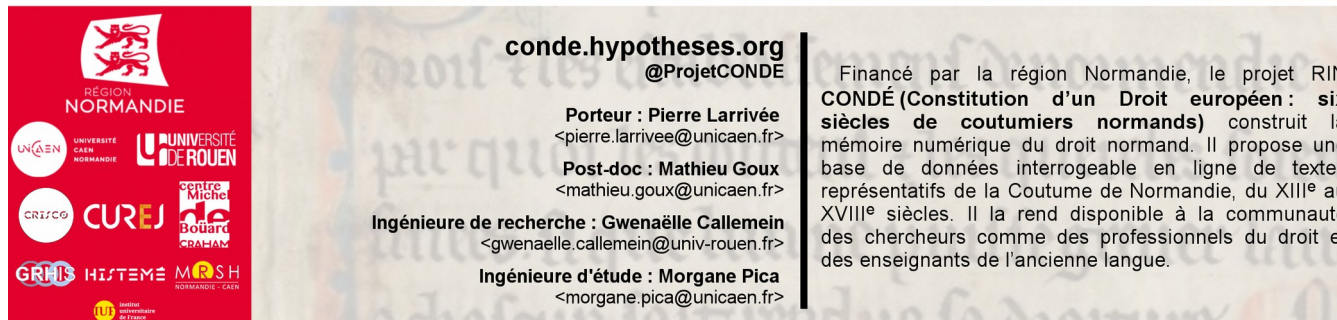
Research project funded by the Normandy region for a period of three years (December 2018 => September 2021)



1. The ConDÉ project

Three objectives:

- (i) *Patrimonial* dimension: build a database compiling texts representative of the Norman Laws, over a long stretch of time (1250 => end of the 18th century), and establish a digital map of the owners/authors to show the vitality of written texts in Normandy.
- (ii) *Legal* dimension: facilitate access for historians and law historians to a highly homogeneous legal corpus, characterized by a rich arrestography and a strong tradition of commentary.
- (iii) *Linguistic* dimension: enriching existing textual corpora with "speciality" speeches.



The image shows a project logo on the left and contact information on the right. The logo is a red square with a white lion rampant in the center. Below the lion, it says 'RÉGION NORMANDIE'. Underneath that are logos for 'UNICAEN UNIVERSITÉ CAEN NORMANDIE' and 'UNIVERSITÉ DE ROUEN'. Further down are logos for 'CRIFCO CUREJ' and 'Centre Michel Bouard GRAHAM'. At the bottom are logos for 'GRIS HISTEME' and 'MRS H NORMANDIE - CAEN'. To the right of the logo, the text reads: 'conde.hypotheses.org @ProjetCONDE'. Below that, it lists the 'Porteur : Pierre Larrivée <pierre.larrivee@unicaen.fr>', 'Post-doc : Mathieu Goux <mathieu.goux@unicaen.fr>', 'Ingénieure de recherche : Gwenaëlle Callemein <gwenaelle.callemein@univ-rouen.fr>', and 'Ingénieure d'étude : Morgane Pica <morgane.pica@unicaen.fr>'. To the right of this text, there is a vertical line, and to the right of the line, a paragraph of text: 'Financé par la région Normandie, le projet RIN CONDÉ (Constitution d'un Droit européen: six siècles de coutumiers normands) construit la mémoire numérique du droit normand. Il propose une base de données interrogeable en ligne de textes représentatifs de la Coutume de Normandie, du XIII^e au XVIII^e siècles. Il la rend disponible à la communauté des chercheurs comme des professionnels du droit et des enseignants de l'ancienne langue.'

2. “Digital Humanities”?

For a long time, Human Sciences have shown little interest in digital tools, which means they are still nowadays seldom taught or encouraged in “regular” academic training.

For this reason, they were made a subject in themselves, as a crossover between Human Sciences and digital tools. Usually, digital humanists have at least a degree in a Human science and learn digital tools in a separate training (often at MSc level).

The fact is, digital tools are becoming more and more central to the Humanities. Tools have been and are being developed *for* your subjects, by scholars *from* your subjects. They help you gain time and precision in your analysis. They make documents available and workable worldwide.

2. “Digital Humanities”?

The impression Humanities students usually have is that you don't need digital tools in their subject. More often than not, they even seem to feel they are incompatible with or at the very least, not directly relevant to their domain.

This is something I feel I should address now and ask you to disseminate among you:

Digital tools are not out of your reach. Feel free to look into what interests you. Look up tutorials. It may present a challenge at some point because the nuanced way of thinking you've been taught all along is different from the computer's binary running.

But I promise, with time and practice, you will gain *a lot* of time and efficiency.

3. Digital Humanities “turn”?

Said “turn” is already taken in “hard” sciences, probably because a large part of their work is calculation, which is basically what the first computers were for. Using computers was therefore quickly obvious.

Humanities are now realizing computers have moved beyond basic calculation and enable them to, first of all, deal with the largest sets of data in a matter of seconds, when tens of scientists’ life’s work would not suffice to analyse the same sets manually.

The “turn” is, in fact, quite as much about some engineers specialising in digital tools for Human sciences as it is about Humanists learning *about* digital tools, if not learning at least some of them.

Everyone cannot spend their life developing the most up-to-date tools (these do evolve *very* fast), but anyone can learn enough about formats and digital languages to use tools and know what can be done and what cannot –as of yet.

3. Digital Humanities “turn”?

Another point of view: did the D.H. in fact acquire a specific academic identity and *raison d'être* in the process?

« Les Humanités Numériques peuvent s'appuyer sur une histoire passée mais n'ont pour objet que l'hyper-présent ou le futur à venir. Elles servent à comprendre le monde de demain, à le faire advenir dans de meilleures conditions et à l'expliquer aux citoyens, aux dirigeants, etc. Le chercheur en Humanités est ici dans une toute autre situation par rapport à ce qu'il a l'habitude de faire dans sa pratique. Il doit inventer, comprendre et analyser en s'appuyant sur une solide culture classique scientifique qui le met en position de **mieux anticiper les changements dans la société**. Tel est, à mon sens, le rôle des HN. Bien loin des outils numériques utilisés pour notre recherche dans des secteurs disciplinaires donnés... »

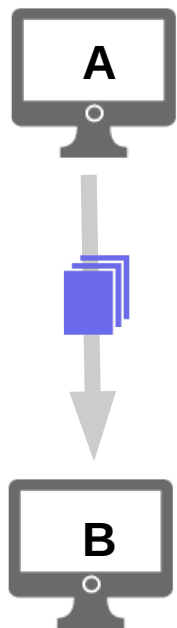
Source : « Les Humanités Numériques : une nouvelle discipline universitaire ? », Suzanne Dumouchel, in Digital Humanities à l'Institut historique allemand, publié le 19/06/2015, consulté le 05/04/2019 - <https://dhiha.hypotheses.org/1539>

« Digital Humanities may rely on passed history, but their only goal is the hyper-present or coming future. **They help understand the world of tomorrow, make it happen in better conditions and explain it to citizens, leaders, etc.** The researcher in Humanities here is in a completely different situation, compared with what they are used to do in their practice. They must invent, comprehend and analyse, relying on a solid scientific classical culture, which enables them to **better anticipate change in society**. Such is, in my opinion, the role of Digital Humanities. Far from digital tools used for our research in given disciplinary area... »

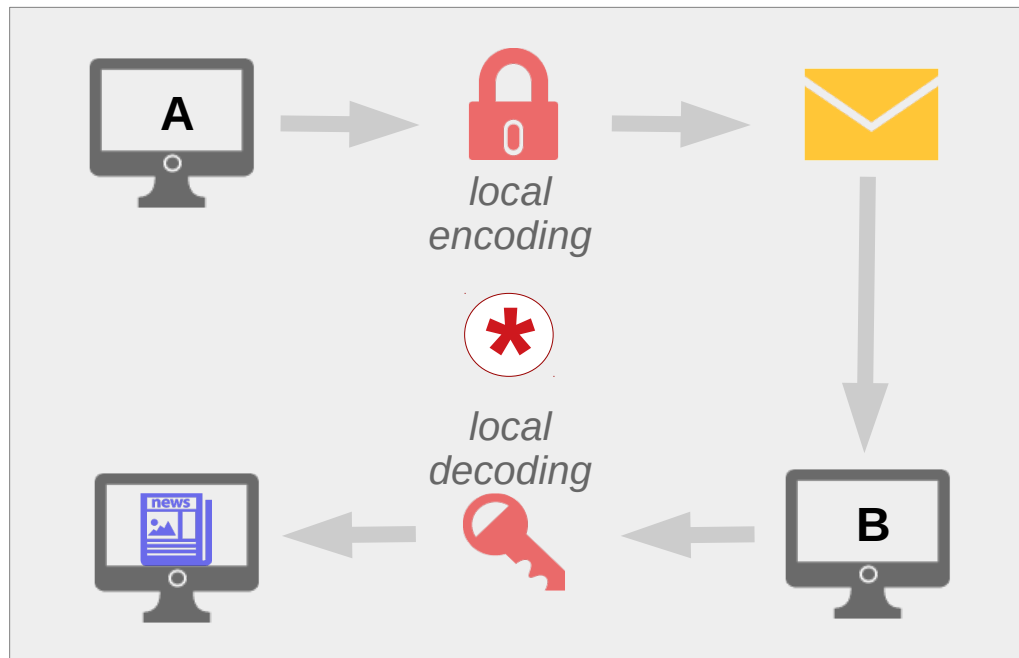
Source : « Digital Humanities : a new academic disciplin ? », Suzanne Dumouchel, in Digital Humanities at the Institut historique allemand, published on 06/19/2015, read on 04/05/2019 - <https://dhiha.hypotheses.org/1539>

1. What are D.H. for?

ACCESS DOCUMENTS



=

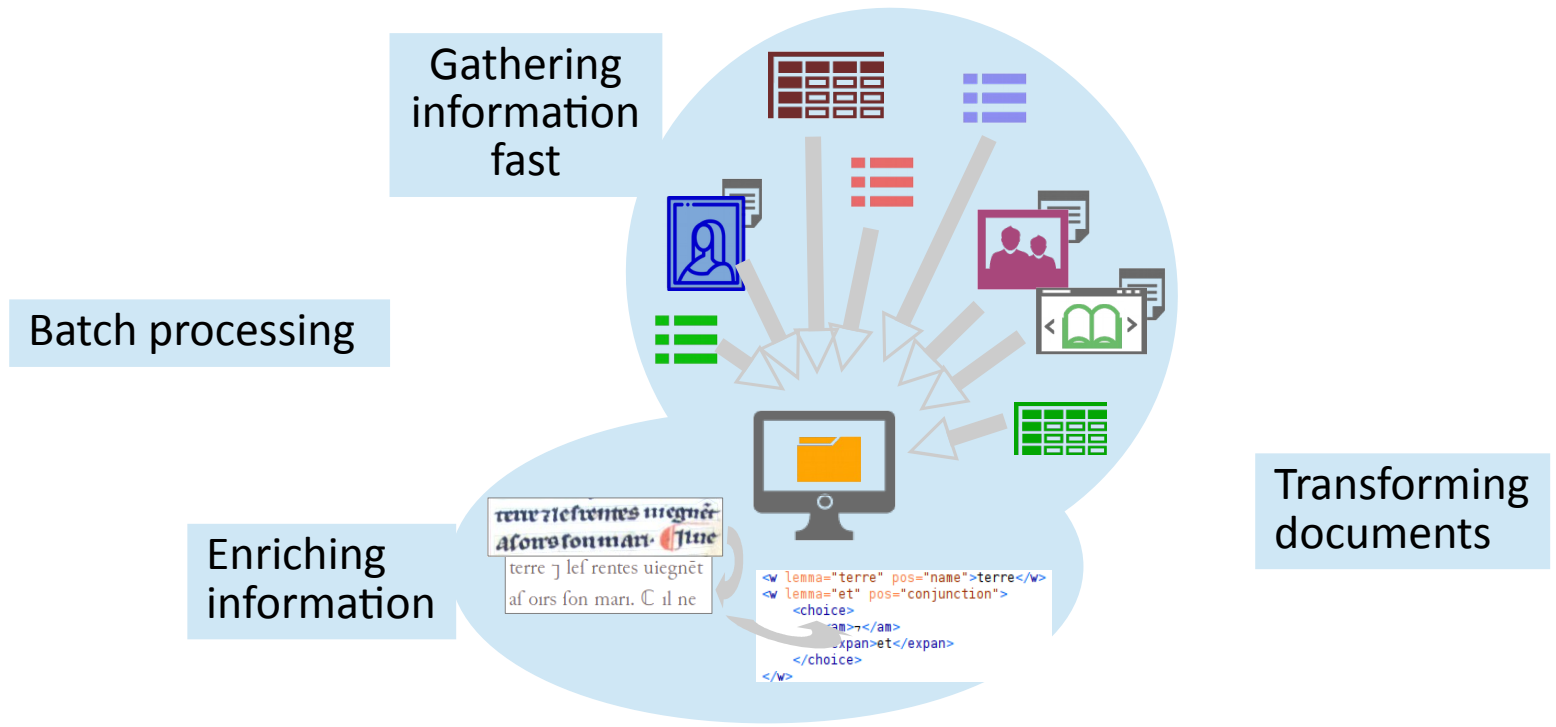


Hence the need for international standard formats.

D.H. deal with specialized standards and establish them when needed.

1. What are D.H. for?

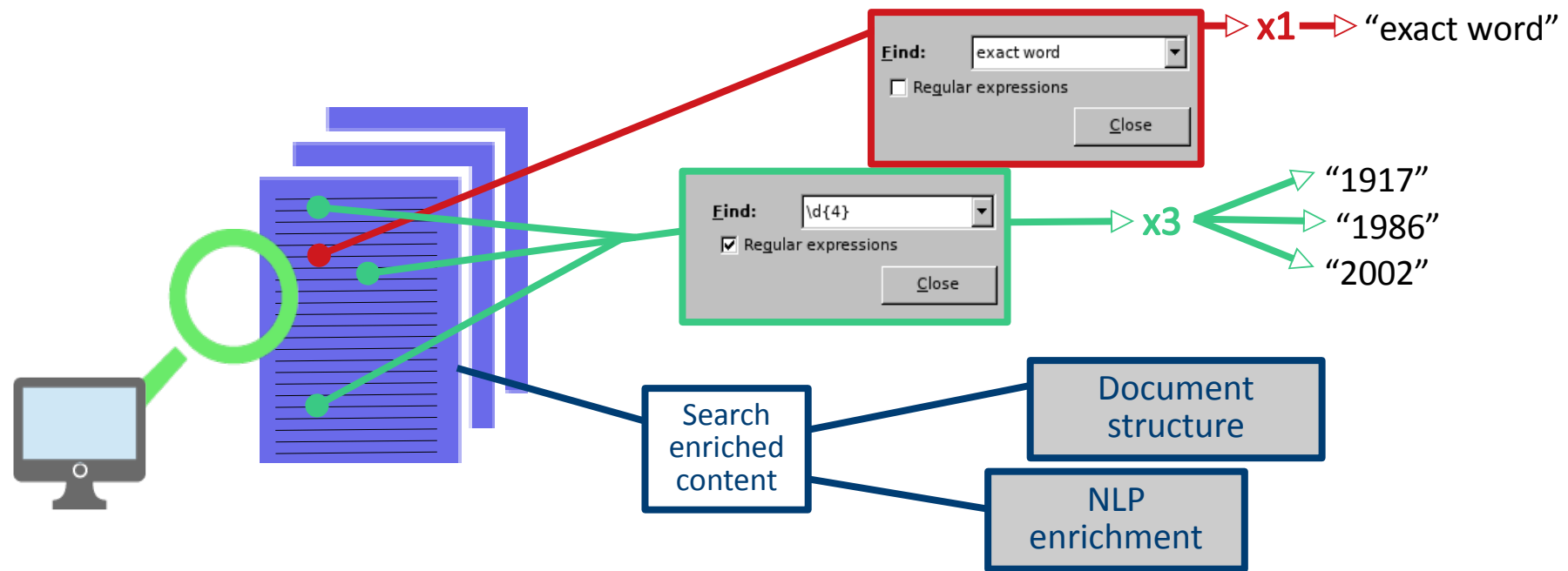
PROCESS DOCUMENTS



1. What are D.H. for?

AUTOMATIC SEARCH

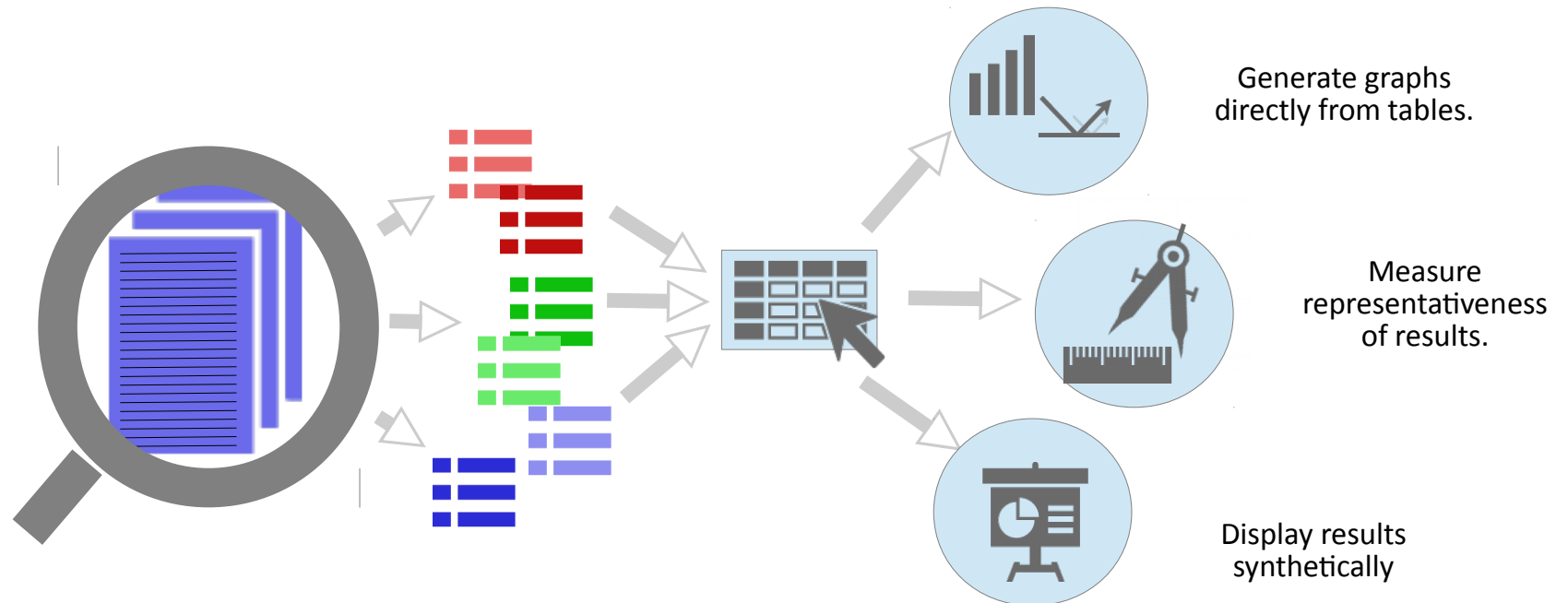
—HELP IN ANALYSIS—



1. What are D.H. for?

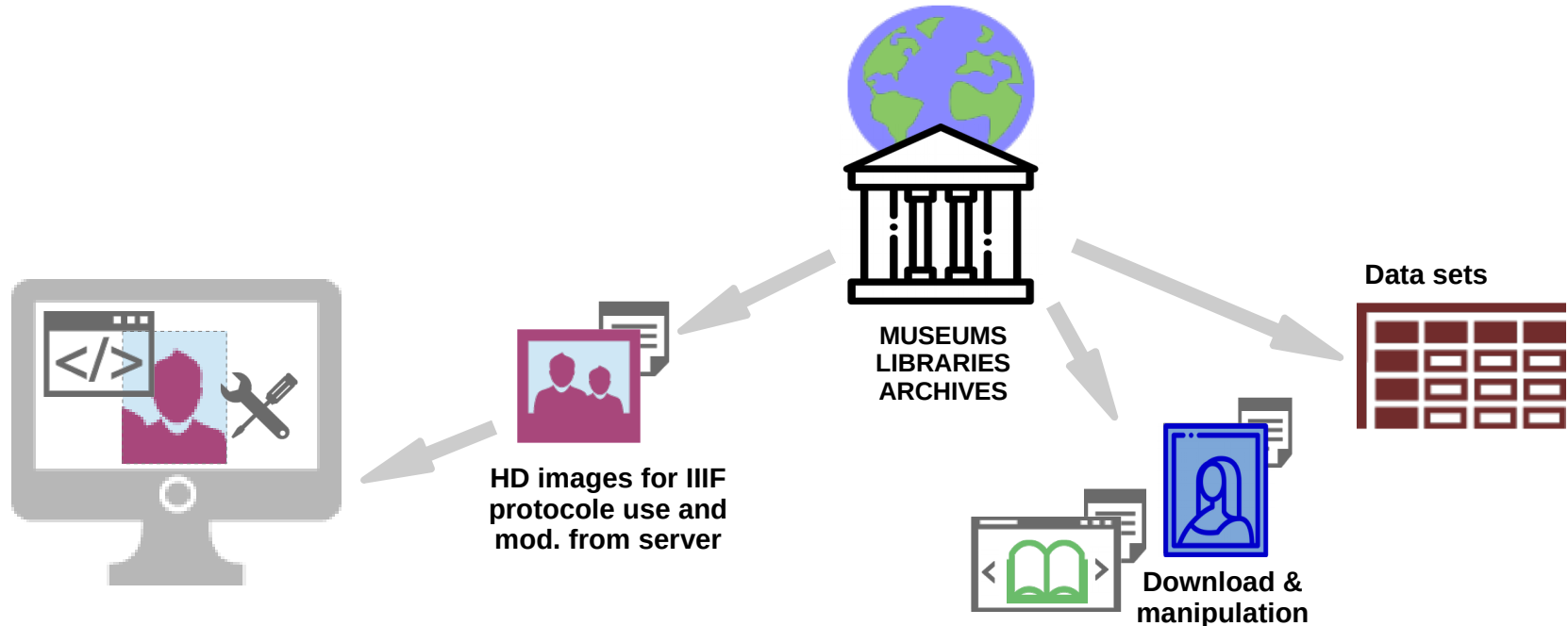
AUTOMATIC SEARCH

—REPRESENTATION OF RESULTS—



2. What kind of work?

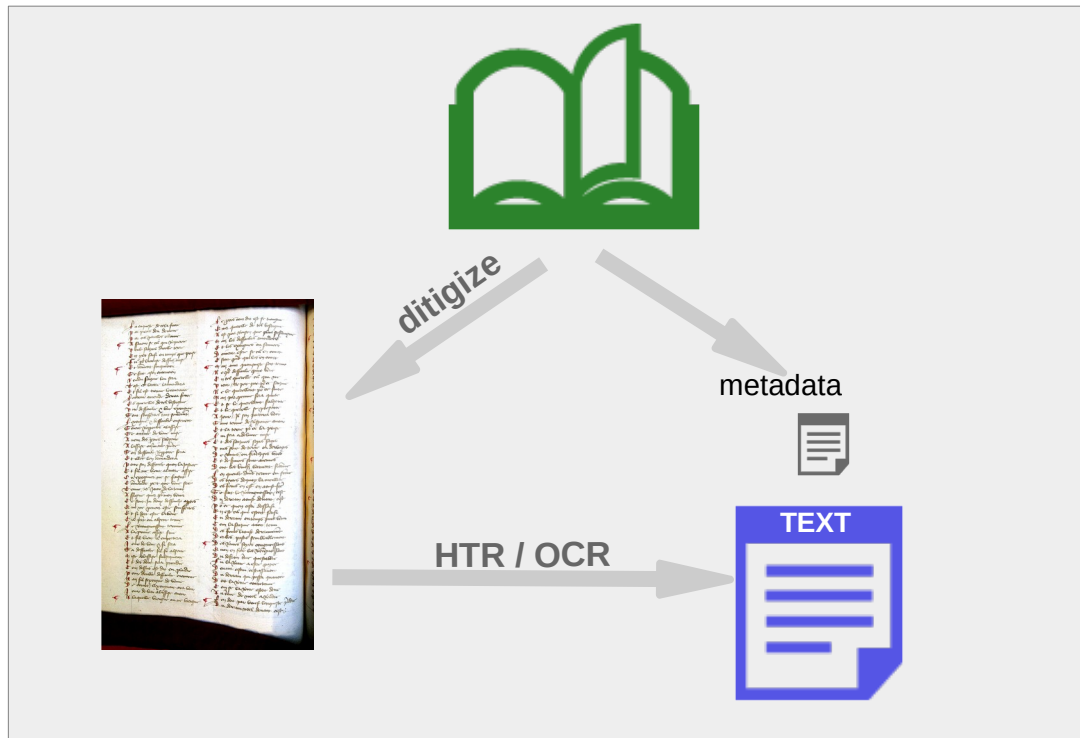
INSTITUTIONAL DATA



II. Digital Humanities, in practical terms

2. What kind of work?

CONVERT DATA

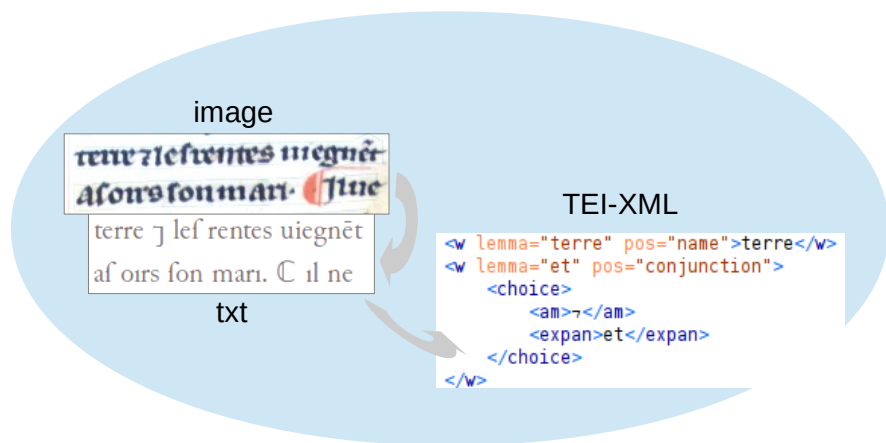


- Convert a physical image to machine-understandable pixels.
- Convert image into text (OCR, HTR).
- Convert full texts into a list of relevant occurrences for analysis.
- Convert an endless table with multiple columns to intelligible information.
- Convert a complex TEI-XML text into a human-readable HTML page or digital book.

2. What kind of work?

ENCODE TEXTS

→ A computer “thinks” in terms of what is and isn’t there. If you give it basic text and ask it to spot a paragraph, it will return an error because it only sees bits (0s or 1s) forming characters. You have to tell it where paragraphs start and end. This is enrichment.

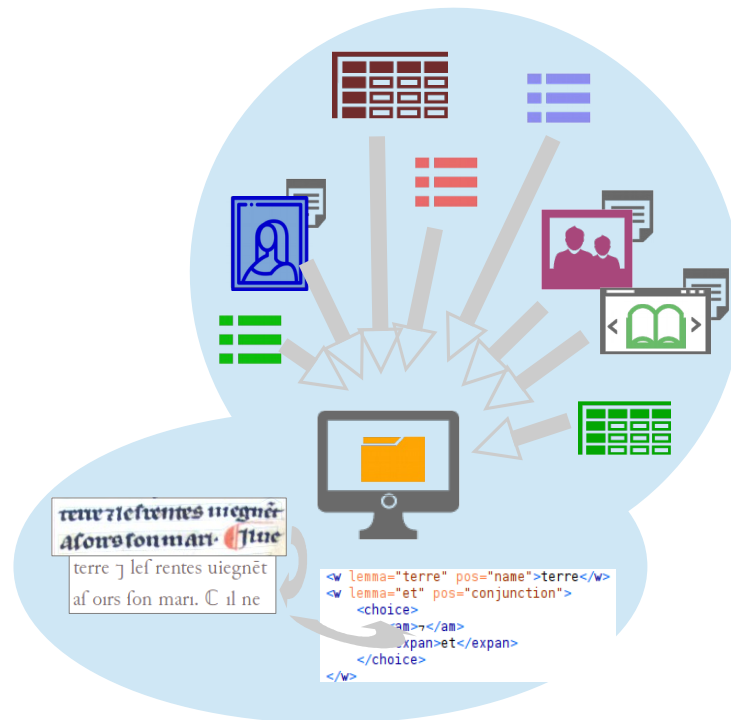


- Encoding = enriching basic text data with the nature of things.
- Structure, grammar, syntax, translation, synchronisation with another document...
- Used languages: mostly XML for texts (usually TEI-XML <<https://tei-c.org>>, but M.Office and L.Office use their own XML).

2. What kind of work?

BUILD DATABASES

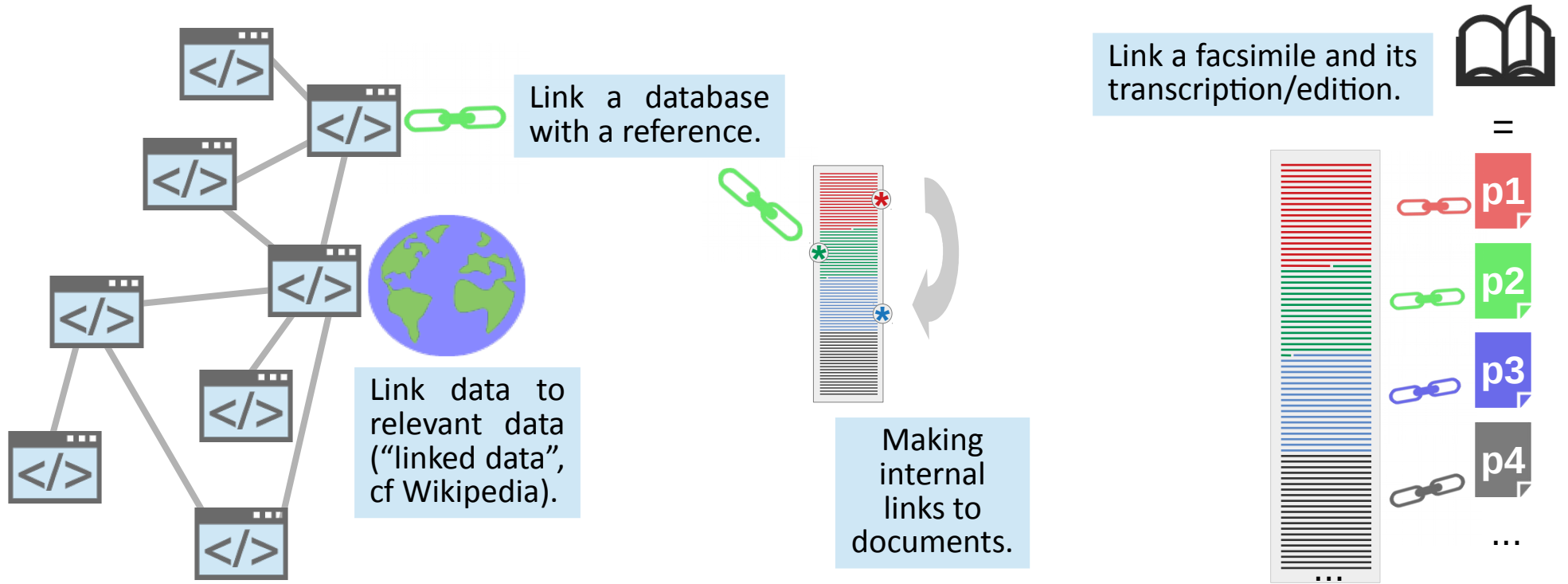
- Digitize information for research.
- Make implicit information explicit via digital formats.
- The aim is to be able to use your data but also to share information: usually, databases are built to be made available to the public in the end.



- Build thematic and homogeneous databases on your subject, including useful information, for more precise and faster research.
- Homogeneous structure does not mean homogeneous data (diachrony, text/image/music, genres...) Building a database is as much about what the data is as it is about how you want to use it.

2. What kind of work?

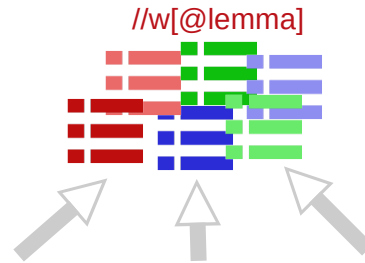
LINK DATA TOGETHER



2. What kind of work?

ANALYSE DATA

Order your data according to your priorities.



Confront theory to massive data

Do fine search in enriched data

```
<w lemma="terre" pos="name">terre</w>
<w lemma="et" pos="conjunction">
  <choice>
    <am>~</am>
    <expan>et</expan>
  </choice>
</w>
```

Use statistical methods on your results

2. What kind of work?

DISPLAY THE DATA

In clear, give the public the tools to work with your data along with said data



Document everything you did, so that later, when every standard has been replaced with another, someone might be able to understand your data

Design tools relevant to them, give them the possibility to process it themselves if they need for something you yourself did not need

Make sure the data and tools are maintained to insure as long-term access as possible

2. What kind of work?

WHAT WE USE THESE FOR

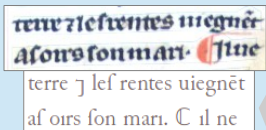
Cultural data management

Programming for cultural institutions

Digital archive management

Engineering for research projects

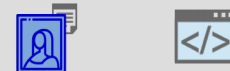
Digital text edition



```
<w lemma="terre" pos="name">terre</w>  
<w lemma="et" pos="conjunction">  
<choice>  
<w lemma="et" pos="conjunction">et</w>  
</choice>  
</w>
```



MUSEUMS
LIBRARIES
ARCHIVES



Build and feed
digital libraries
and exhibits

1. D.H. in the ConDÉ project

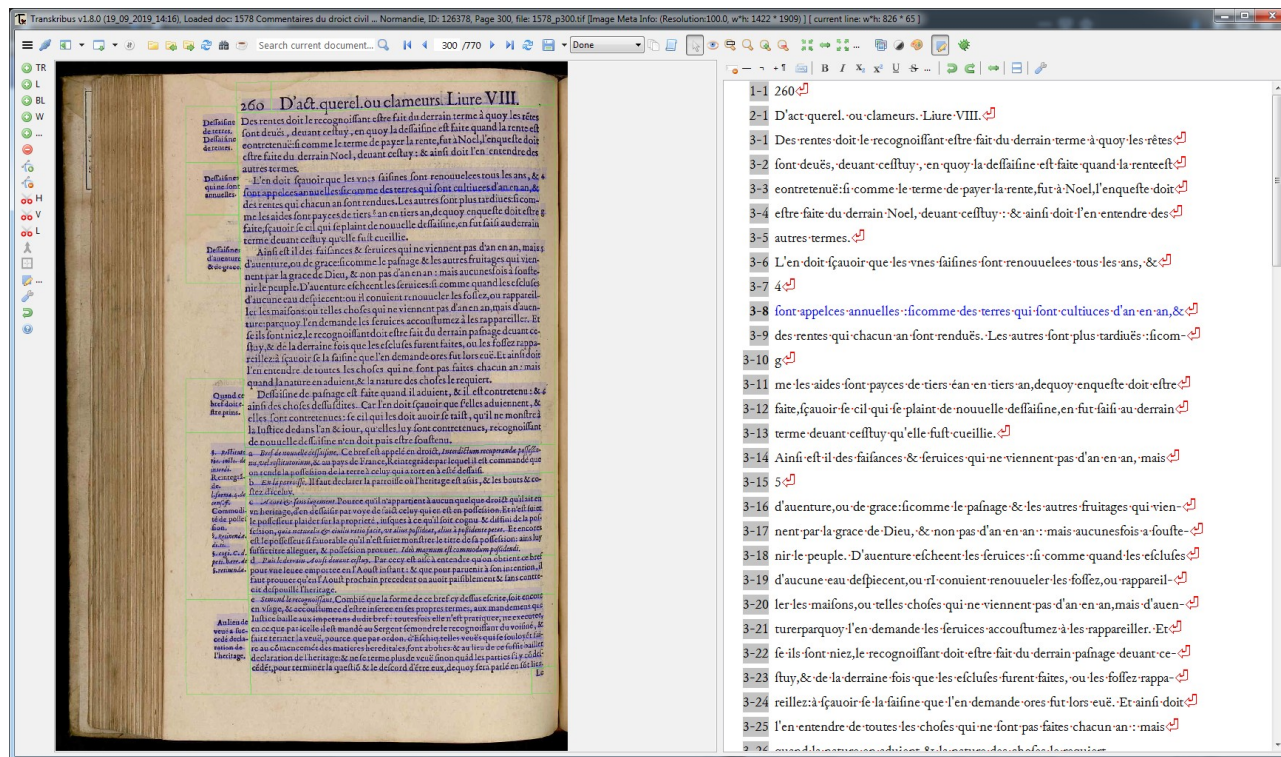
DIGITISE THE CORPUS

- P. Larrivée and G. Cazals identified the representative texts before the project began. They selected about fifteen texts, from the Très Ancien Coutumier (mid/late 13th century), the oldest known text of the Custom, to Pesnelle's Coutume expliquée (1771), the last major work on Norman Law before the establishment of the Napoleonic Civil Code, which marks the end of French customary law.
- An attempt was made to include one text for every 50 years, knowing that before the sixteenth century, sources were scarce.
- Many texts were already digitised / accessible on large databases, including Gallica (<<https://gallica.bnf.fr>>) and Google Books. However, we had to take pictures of some of the manuscripts ourselves, or ask for a new digitisation, their initial quality being insufficient for a semi-automated transcription.

III. What's in store?

1. D.H. in the ConDÉ project

DIGITISE THE CORPUS



→ After digitisation, we had to transcribe the texts using the OCR/HTR *Transkribus* software <<https://transkribus.eu/Transkribus/>>.

→ Success rates: more than 97/99% correct characters on something in the order of 15/20 million characters, in less than a year. During this phase we were also able to prepare the TEI-XML encoding.

1. D.H. in the ConDÉ project

BUILD THE DATABASE

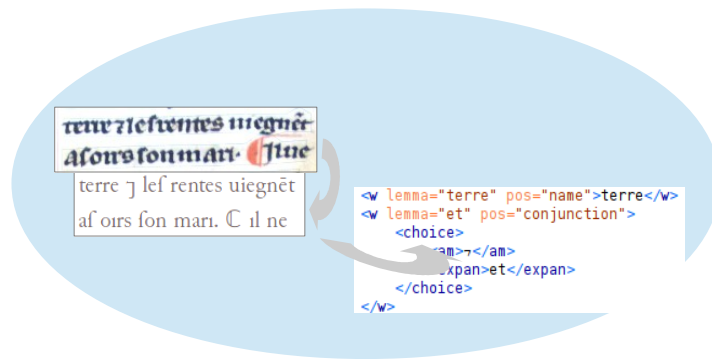
- Building a homogeneously structured TEI-XML database for heterogeneously conceived works is a challenge.
- Identifying the basic units of text, how they are connected to others.
- Knowing very early on who will use the database and what they want it for, so as to know which information to keep and which to ignore.
- Knowing which information to keep as text and which to record within the code.

III. What's in store?

1. D.H. in the ConDÉ project

NLP ENRICHMENT

- NLP: Natural Language Processing.
- Tokenize words and record their lemma and POS information



- This allows fine search and analysis for linguists.
- Recorded in the XML file of the text, in the same TEI standard.

1. D.H. in the ConDÉ project

BUILD AN INTERFACE

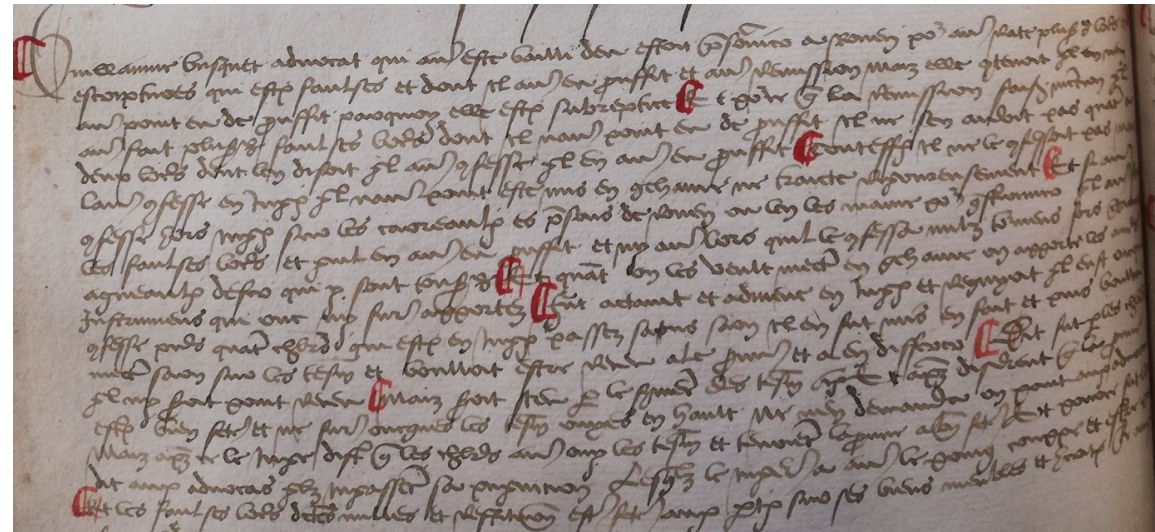
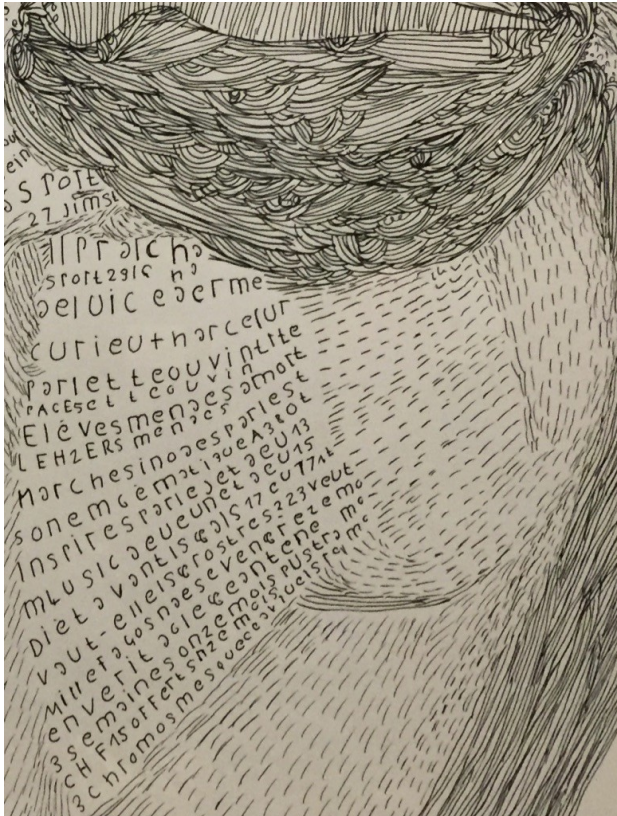
- Build an interface to go through the database and fine-search it. We wish to enable the visitors to choose which translation they will see: diplomatic, semi-diplomatic or modernized. We also wish to enable visualization of facsimile images.
- This has influence on how we enrich the texts, therefore it was one of the earliest questions we asked ourselves.
- Once again, we start by remembering who will need and use the interface and try to meet their needs.
- Make it as simple to use as possible.

III. What's in store?

2. New developments & expectations

BEYOND BINARY FUNCTIONING

→ Aim to imitate neuronal functioning to teach the computer to think like a human.



↑ Charles Morisse. 1386. Notes. Grand Coutumier de Normandie

← Language imitation. Outsider artist Paul Vonlanthen (2019, no title)

3. How to get started, if you're willing

REMEMBER:

- For now, the computer gives you facts, but it still is *your* job to analyse them.
- Digital tools are not “magic”: they will not allow you to do *anything*. They allow you to do new things but they are also based on very real parameters you need to consider. Start by learning about possibilities and limits. Learning the basics of one language (Python or R, for instance) will already give you an idea. Without being able to *program* in a language, learning to *read* it will help you better understand how code works.
- A real “digital” edition is *not* a paper edition recorded in DOCX format. Before you start handling data, think, read, ask people about the new possibilities digital tools give you and the limits they present.

III. What's in store?

3. How to get started, if you're willing

- Get a text editor (Notepad++, Atom...) to write/read code or handle text data, instead of a word-processing package (yuk!!). They are specialized and much lighter.
- Learn to phrase your theories in binary possibilities:
 - ✗ “How dependant is parameter A upon parameter B?”
 - ✓ “Is parameter A significantly higher when parameter B is present or absent?”
 - ✗ “How is this word used in this text?”
 - ✓ 1 - “What are the left and right contexts of the word?”
 - ✓ 2 - “Are some context parts frequent?”
- Learn regular expressions. They are a way to categorise characters and a *very* powerful way to handle text data. Most text editors, word-processors, spreadsheets, etc. accept them. It may be a challenge and take more than one try but I guarantee it is worth every bit of effort and it will be your best ally in dealing with texts.