

UNIVERSITÉ CAEN NORMANDIE





# The Digital Humanities Turn

#### Morgane PICA & Pierre LARRIVÉE

Unicaen, CRISCO (EA4255)



# The Digital Humanities Turn

#### I. Introduction

The ConDÉ project
"Digital Humanities"?
Digital Humanities "turn"?

II.Digital Humanities,in practical terms1. Digital Humanities: what for?2. What kind of work exactly?

III.What's in store?

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

# 1. The ConDÉ project

## Thanks to the Normandy region

I. Introduction

#### 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)



#### I. Introduction

# 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.



#### I. Introduction

# 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 developped *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.

#### I. Introduction

# 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.

## I. Introduction 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 as about Humanists learning *about* digital tools, if not learning at least some of them.

Everyone cannot spend their life developping 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.

## I. Introduction 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'hyperprésent 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

#### ACCESS DOCUMENTS



#### PROCESS DOCUMENTS



#### AUTOMATIC SEARCH

#### -HELP IN ANALYSIS-



#### AUTOMATIC SEARCH

#### -REPRESENTATION OF RESULTS-



### INSTITUTIONAL DATA



### CONVERT DATA



- → Convert a physical image to machine-understandable pixels.
- $\rightarrow$  Convert image into text (OCR, HTR).
- $\rightarrow$  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.

### 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.



- $\rightarrow$  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).

### **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.

#### LINK DATA TOGETHER



Link a facsimile and its transcription/edition.





#### ANALYSE DATA

Order your data according to your priorities.



Confront theory to massive data

Do fine search in enriched data

Use statistical methods on your results

#### DISPLAY THE DATA

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

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



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

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

#### WHAT WE USE THESE FOR



#### 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 digitsation, their initial quality being insufficient for a semi-automated transcription.

#### DIGITISE THE CORPUS

• • • • • • • • •	🛚 🙀 🤯 🗃 🖱 Search current document 🔍 🚺 🖣 300 /770 🕨 🎽 🦉 🔚 💌 Done	<u>■10 ■  } ● 4 4 4 6 II + II - 10 4 9    </u>
		$a = a + 1 \implies B I X_1 X_2  $
F The second	D'a guerel ou clameurs Linre VIII.	1-1 260
	Defisitat Des rentes doit le recognoiffant eftre fait du derrain terme à quoy les rêtes	2-1 D'act-querel. •ou ·clameurs. ·Liure ·VIII.
	deterra. font deues, deuant celluy, en quoy la defiaitine elt latte quand la rente elt Defiaiane contestenuis fi comme le terme de payer la rente, fut à Nocl, l'enquelle doit	3-1 Des rentes doit-le recognoiffant eftre fait du derrain terme à quoy-les rêtes 🖉
	eftre faite du derrain Noel, deuant ceftuy : & ainfi doit l'en entendre des	3-2 font deues deuant cellus: en quov la della fine eft faite quand la renteeft
	Deflisser -L'en doit feauoir que les vnes faifines font renounclees tous les ans, & 4	
	annuelles. font appelees annuelles dicomme des terres du font cultures fine plus tardiues fine and des rentes qui chacun an font renducs. Les autres font plus tardiues ficom-	
	me les aides sone payces de tiers <sup>6</sup> an en tiers an, dequoy enquelle doit eitre g	3-4 eftre-faite-du-derrain-Noel, deuant-cefftuy-:-8c-ainfi-doit-l'en-entendre-des
	terme deuant celluy qu'elle fuit eucillie.	3-5 autres termes. ♥
	d'aussure d'aussure, ou de grace:ficomme le palnage & les autres fruitages qui vien-	3-6 L'en doit-fçauoir que les vnes faifines font renouuelees tous les ans, & 4
	nent par la grace de Dieu, & non pas d'an en an : nara auconatora toutes nir le peuple. D'auenture cfcheent les feruices di comme quand les cfclufes	3-7 44
C . Charles	d'aucune cau defaiecent: ou il conuient renouueler les foirez, ou rapparei- ler les maifons: ou telles chofes qui ne viennent pas d'anen an, mais d'aucu-	3-8 font appelces annuelles : ficomme des terres qui font cultiuces d'an en an &
	ture:parquoy l'en demande les feruices accouftumez à les rappareiller. Et	3-9 decremente conjucto sono non font rendujic. Je econote control na territivica (font
	fray, & de la derraine fois que les efclufes furent faites, ou les foffez rappa-	5 7 des rentes qui enacun an ione renduces. Les autres ione plus taldides incom
	l'en entendre de toutes les chofes qui ne font pas faites chacun an : mais	3-10 g⇔
	quand la nature en aduient, & la nature des choles le requiert. Ound ce Deflaifine de pafnage ell faite quand il aduient, & il eft contretenu : & 4	3-11 me·les·aides·font·payces·de·tiers·éan·en·tiers·an,dequoy·enquefte·doit·eftre
	herfdoite- fite prime- aller Corr controller de cil qui les doit auoir fe taift, qu'il ne monfreà	3-12 faite, fçauoir-fe-cil-qui-fe-plaint-de-nouuelle-deffaifine, en-fut-faifi-au-derrain 🖑
	la luffice dedans l'an & iour, qu'elles luy font contretenues, recognoillant	3-13 terme deuant cefftuy qu'elle fult cueillie.
	4. miliuse a Brildenmanikeidfuijne. Cebref eh appelé en droit, sundiflum maprende ps [[]a- te miliuse a Brildenmanikeidfuijne.	3-14 Ainfi eft il des faifances & feruices qui ne viennent pas d'an en an, mais
	Rentegil. on sendela poffesion dela terre h celuy qui a torte n à effe deflaif. Rentegil. b. se foregrafie U fun de la terre h celuy qui a torte n à effe deflaif.	3.15 5.4
	lorna, de Rez diceluy. ante e Atari es fonisemon. Pource qu'il s'appartient à aucun quelque droidt qu'ilsit en	
	Commodi- un herinage, d'en deflaifir par voye de faid celuy qui en eft en poffession. Et n'eft faint téde pollet le poffeficur plaider fur la proprieté, infques à ce qu'il foit cognu se duffai dela pol-	5-16 d auenture, ou de grace li comme le painage oc les autres fruitages qui vien-
	Legimente, effision, quis netnealu & chala estic facit, ur aline paffidare, dine à paffidare peres. Et efficies Legimente, efficience il fauorable qu'il n'eff fuier monfirer le tirre de la polifeition: an les	3-17 nent-par-la-grace-de-Dieu, -&-non-pas-d'an-en-an-t-mais-aucunestois-a-loulte-
	s.espi.c.d. fuffictive alleguer, & polletion product. Law magnum granment production. penkens.et d. Daidederson Assif danar cafter. Par cecy eft alle kentendre qu'on obient cohert	3-18 nir·le·peuple. D'auenture escheent·les·feruices·ssi comme·quand·les·escluses¢
	faut menue qu'ail Aout prochain precedent on auoit patiellement & faut contra- faut profession d'Aout prochain precedent on auoit patiellement & faut contra- ent defonuell.	3-19 d'aucune-eau-defpiecent, ou rI-conuient-renouueler-les-foffez, ou rappareil- 🖑
	e stoutout la reconsojfaut, Combié que la forme de ce bref cy deflus efente, loit encost en viage. La accoultume e d'effre inferer en fes propres termes, aux mandement que	3-20 ler-les-maifons, ou-telles-chofes-qui-ne-viennent-pas-d'an-en-an, mais-d'auen-4
	Aulieu de Iuflice baile aux impersans dudit bref: touresfois elle n'eft pratiquer, ne executés verse fice, en coque par icelle si eft mandé au Sergent femondre le recognoifiant du voinné, &	3-21 turerparquoy l'en demande les feruices accouftumez à les rappareiller. Et 🖑
	erdé decla- faire termer la veué, pource que par ordon, d'Efchiquelles veués qui fe fouloret la ration de re au cômencemér des matieres heredicales, font abolies de au lieu de ce fufficientilet	3-22 failetfontuier letrecornoiffont doiteffre frit du dernin pafnare deuant ca-
	checuage. declaration de liberitage & ne fe terme plus de vené finon qu'à les parties la voint cédét, pour terminer la queffi 6 & le defort d'étre eux, dequoy fera partie en fier ins	
		3-23 Ituy,& de la derraine fois que les elclules furent faites, ou les follez rappa-
		3-24 reillez:à·lçauoir·le·la·laifine·que·l'en·demande·ores·fut·lors·euë. Et·ainfi·doit
		3-25 l'en entendre de toutes les chofes qui ne font pas faites chacun an : mais 🖉

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

### BUILD THE DATABASE

- → Building a homogeneously structured TEI-XML database for heterogeneously conceived works is a challenge.
- $\rightarrow$  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.
- $\rightarrow$  Knowing which information to keep as text and which to record within the code.

#### 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.

### **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.
- $\rightarrow$  Make it as simple to use as possible.

## III. What's in store? 2. New developments & expectations

#### **BEYOND BINARY FUNCTIONING**



 $\rightarrow$  Aim to imitate neuronal functioning to teach the computer to think like a human.

Camer Onfinet admorat one and Efter Garthe Don offor Gotomoro asponsy por and pater page hospetrato an the faulter it bank than son puffer it and struggers may ever grown fe not on be guffit paragners are off pubrishing & goin & Ra Arm pros sangest faul to assid bout he now point on de puffer Nem to Bit's Stur lety Super je an officit ge Eg an Son puffit four office me 5) Trigo FE non Control offermo by geganne por southe algorison And my por as a raw course to pland be some on an an an an and shawn out faulps and the put as an the pupt sty and and and and a fifthe mit Mp defro on p. pour on proto and and to co balt most on organic for apportes fit actomet it adment by Tupp it & puds quar egoed of met by hugo partez papers pin rean fut mind pue to teps of bontron spor and ale pur point reason linais for how At It we pre surgent to the on The huger off ( as sois and any as hit of the fund the conference of months for programming the

Charles Morisse. 1386. Notes. Grand Coutumier de Normandie

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

## III. What's in store? **3. How to get started, if you're willing**

#### REMEMBER:

- $\rightarrow$  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.
- $\rightarrow$  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.