

# Making one's work reproducible

Christophe Pouzat

MAP5, Paris-Descartes University and CNRS

`christophe.pouzat@parisdescartes.fr`

LASCON, January 19 2018, Lecture 1

# Outline

Introduction

What is needed for Reproducible Research?

Note-taking concerns everyone

Note-taking: a quick History

Form text files to lightweight markup languages

Notes (and codes) that are archived but can evolve with version control systems

# Where are we ?

## Introduction

What is needed for Reproducible Research?

Note-taking concerns everyone

Note-taking: a quick History

From text files to lightweight markup languages

Notes (and codes) that are archived but can evolve with version control systems

## The usual presentation of research work

*An article about computational science in a scientific publication is **not** the scholarship itself, it is merely **advertising** of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.*

Thoughts of Jon Claerbout "distilled" by Buckheit & Donoho (1995).

# What do we do as manuscript "producers"?

The preparation of manuscripts and reports in experimental or computational neuroscience often involves:

- ▶ Doing the experiments or coding and doing the simulations.
- ▶ A lot of data analysis.
- ▶ Careful design and realization of figures and tables.

Simulations involve:

- ▶ Algorithmic choices.
- ▶ Software choices or implementation.
- ▶ Parameters setting (e.g., the time step used for conductance based neuronal models).

Data analysis involves:

- ▶ Some data "preprocessing" (often).
- ▶ The development of dedicated scripts, routines and codes.
- ▶ Setting some critical parameters.

## What would we like as manuscript "readers"?

As readers or referees of articles / manuscripts we are therefore often led to ask questions like:

- ▶ What would happen to the analysis (or simulation) results if a given parameter had another value?
- ▶ What would be the effect of applying my preprocessing to the data instead of the one used by the authors?
- ▶ What would a given figure look like with a log scale ordinate instead of the linear scale use by the authors?
- ▶ What would be the result of applying that same analysis to my own data set ?

We can of course all think of a dozen of similar questions.

## The dilemma

Clearly the classical journal article format cannot do the job.

- ▶ Editors cannot publish two versions of each figure to satisfy different readers.
- ▶ Many intricate analysis and modeling methods would require too long a description to fit the usual bounds of the printed paper.
- ▶ We all have, moreover, a lot of different things to do and we cannot afford to systematically look at every piece of work as thoroughly as suggested above.

We need are more systematic and more explicit ways to describe how the analysis (or modeling) was done.

# The solution: Reproducible Research

Reproducible Research (RR) or reproducible data analysis is an approach aiming at complementing classical printed scientific articles with everything required to independently reproduce the results they present.

"Everything" covers here:

- ▶ the data,
- ▶ the computer codes,
- ▶ a precise description of how the code was applied to the data.

The "movement" started with what Economists have been calling replication since the early eighties to reach what is now called reproducible research in computational data analysis oriented fields like statistics and signal processing.

# Where are we ?

Introduction

What is needed for Reproducible Research?

Note-taking concerns everyone

Note-taking: a quick History

From text files to lightweight markup languages

Notes (and codes) that are archived but can evolve with version control systems

# The key "ingredients"

In my view:

- ▶ Good note-taking.
- ▶ Tools and tricks making good note-taking "easy".
- ▶ Thorough code documentation.
- ▶ "General" data formats.
- ▶ Ending with a "fully algorithmic" description of the data analysis (and simulation).
- ▶ Use software that are likely to still be around 15 to 20 years from now.
- ▶ **Use free software.**

## Learn from Unix "philosophy"

Before looking for the last "fancy stuff", try solutions with a serious record like what Arnold Robins exposes in "What's GNU":

- ▶ Break down "complex" analysis into simple steps (*avoid monolithic codes*).
- ▶ Implement each step as a separate program that takes its input from the *standard input* (stdin) and sends its output to the *standard output* (stdout); a **filter** in Unix jargon.
- ▶ Try to stick to the **KISS** (Keep It Simple Stupid) approach for each program.

# Where are we ?

Introduction

What is needed for Reproducible Research?

Note-taking concerns everyone

Note-taking: a quick History

From text files to lightweight markup languages

Notes (and codes) that are archived but can evolve with version control systems

# The scholar annotating his book / manuscript

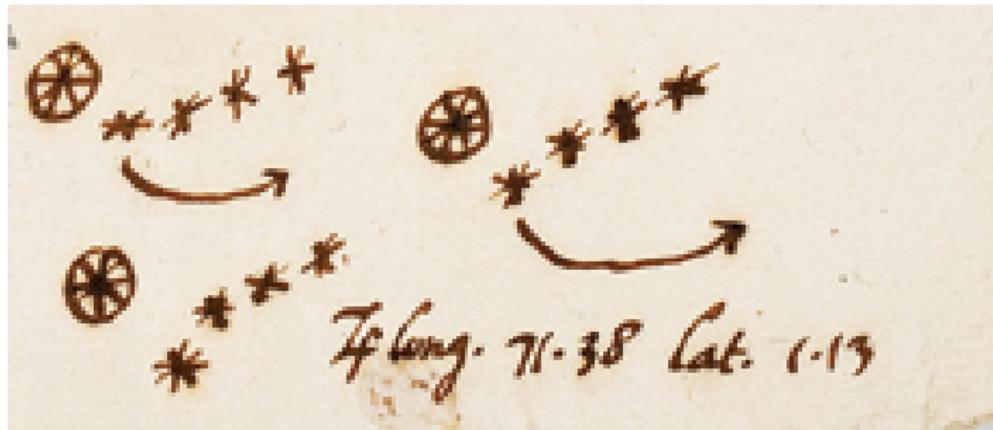


A XIVth century manuscript with the works of Aristotle owned by Nicasius de Planca ([gallica.bnf.fr](http://gallica.bnf.fr) / Bibliothèque nationale de France).

# Galileo observing Jupiter's moons

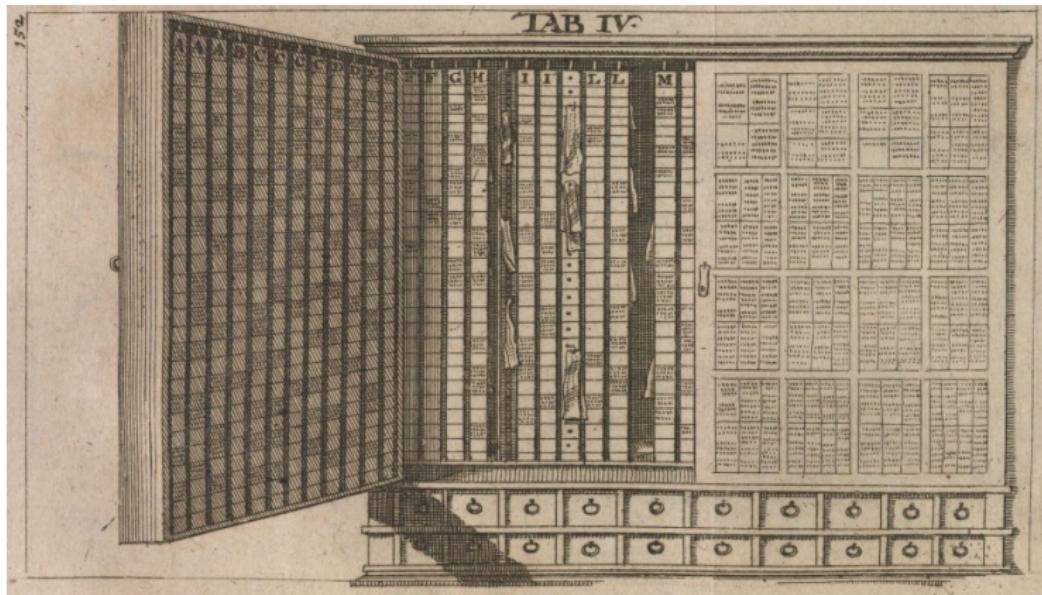
Grazie Gallo Hamilt. Servo della Ser. V. inaugilaro  
so apprezziamo, et lo ogni giorno di potere ho voluto satisfare  
alario che tiene della Altezza Dr. Mathematico nello Stu-  
dio Di Palermo,  
Suo Dureo determinante di presentare al Suo Principe  
l'inchiale et l'opere di giornamento meccanico e ogni  
meccanico et in breve marittimo o terrestre siano di buona qual-  
ita e ragionabile artificio nel maggior segno et along a diffusione  
di l'uso. L'inchiale curato dalla più natale speciezione. Di-  
prospettiva la vantaggiose distribuzione Legion et Vela dell'invia-  
re l'acqua et purificare prima et agli austriaco et difendendo  
il numero et la qualità dei vassalli giudicare le sue forte-  
zze pallottine et la caccia al camuffamento o alla fuga, e pure and  
nella caccia sportiva rivede et particolarmente distinguere ogni suo  
moito di progettamento.

Galileo Galilei's notes while observing Jupiter in January 1610 with his telescope (Wikimedia Commons).

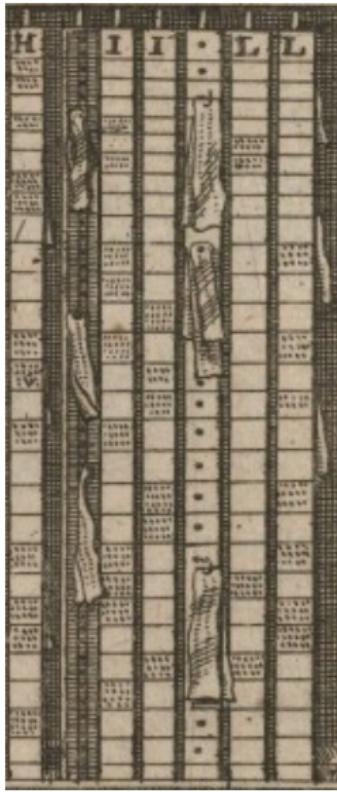


The small "stars" are in fact orbiting around Jupiter, **they are doing what the Moon does around the Earth** (Wikimedia Commons).

# Placcius and Leibniz closet



Organizing notes Placcius' way (Placcius, Vincent, 1642-1699.  
*De arte excerptendi vom gelahrten Buchhalten*, 1689.  
Houghton Library, Harvard University.)



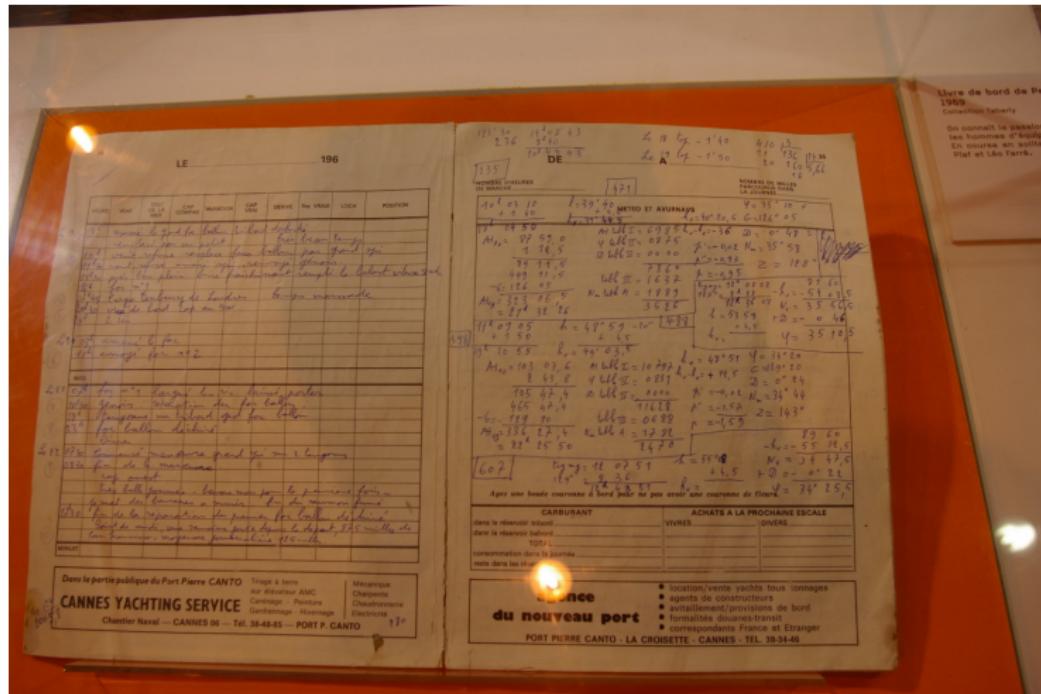
Zoom on the columns of Paccius' cabinet. You can see the "front" (left column), the "side" (second from left) and the "back" (fourth from left).

# Beware of overabundance: Fulgence Tapir's disappearance



In 1908, Anatole France (1844-1924) published "Penguin Island" a parody of French history.  
By Photographer : Wilhelm Benque. Tucker Collection - New York Public Library Archives, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=16240632>.

# A sailor's logbook



The logbook of Eric Tabarly during the San-Francisco / Tokyo transpacific ocean race in 1969.

L 81	02h	foe n°1 largue les ris laissé porter
(2)	10h30	génovis réparation du foe ballon.
	18h	Tangonne sur tribord qui fait ballon
(3)	23h	foe ballon déchiré
		Dinner
L 82	07h30	commence manœuvre grand spi sur 2 tangons
(4)	0830	fin de la manœuvre.
		cap ouest
		tres belles journées - bonne mer pour la pêche forte -
		je met des bananes à cuire. fin du saumon fumé
(5)	17h30	fin de la réparation du poulie foe ballon déchiré
		Point de midi, une seconde grise depuis le départ, 875 milles de San Francisco, moyenne journalière 125 milles.
	MINUIT	

On the left side, Tabarly reports salient events like a ripped jib on March 21 at 11 pm.

		LA JOURNÉE	
		$\varphi = 35^{\circ} 10' N$	
$10^h 03' 10$	$l = 39^{\circ} 40'$	$h_e = 40^{\circ} 20,5$	$G = 126^{\circ} 05'$
+ 1 40	+ 4,5		
	$h_v = 39^{\circ} 44,5$		
$17^h 04' 50$			
$Ah_{v0} = 87 59,0$	$Ah_{tbl\ I} = 69 85$	$h_e = -36$	$D = 0^{\circ} 48' S$
1 12,5	$h_v = 08 75$		
	$tbl\ II = 00 00$	$\mu' = -0,02$	$N_m = 35^{\circ} 58'$
		$\mu'' = -0,93$	<del>126° 05'</del>
$89 11,5$	$78 60$	$Z = 128^{\circ}$	
449 11,5	$tbl\ III = 16 37$		
$-6: 126 05$	$N_m \text{ tabl\ A} = 18 89$		
$Ah_{vg} = 323 06,5$	$35 26$		
= 21 32 26			
$11^h 09' 05$	$h = 48^{\circ} 59' -20''$	$488$	
+ 1 50	+ 4,5		
$19^h 10' 55$	$h_v = 49^{\circ} 03,5$		
$Ah_{v0} = 103 03,6$	$Ah_{tbl\ I} = 10 297$	$h_e = 48^{\circ} 51'$	$\varphi = 34^{\circ} 20'$
2 43,8	$h_v - h_e = +12,5$		$G = 29^{\circ} 20'$
	$tbl\ II = 0831$		$D = 0^{\circ} 24'$
$705 47,4$	$tbl\ III = 0000$	$\mu' = -0,02$	$N_m = 34^{\circ} 44'$
465 47,4	$11628$	$\mu'' = -1,57$	$Z = 143^{\circ}$
$-6: -129 20$	$tbl\ III = 0688$	$\mu = -1,59$	
$Ah_{vg} = 336 27,4$	$N_m \text{ tabl\ A} = 17 92$		
= 29 25 50	$2470$		
<b>[607]</b>	$tg \gamma = 12 07 51$	$h = 55^{\circ} 08'$	$N_v = 34 47,5$
	$129^{\circ} = 8 36$	$+ 4,5$	$-h_v = -55 12,5$
	$12^h 43 51$	$h_v =$	$D = 0^{\circ} 22'$
<i>Ayez une bouée couronne à bord pour ne pas avoir une couronne de fleurs.</i>			$\varphi = 34^{\circ} 25,5$
CARBURANT	ACHATS A LA PROCHAINE ESCALE		

On the right side, he computes his position (that was before GPS time!).

## So, what should we use to take notes?

- ▶ The object of study (like the annotated book)?
- ▶ One or several notebooks?
- ▶ Paper slips or cards?
- ▶ Computer files?
- ▶ Drawings, Pictures?
- ▶ Films?
- ▶ ... ?

# Avoid getting lost

Notes generate an organizational problem:

- ▶ How can we structure our notes?
- ▶ Can we index them, if yes, how?
- ▶ How can we archive them while keeping the capability to make them evolve?

# Where are we ?

Introduction

What is needed for Reproducible Research?

Note-taking concerns everyone

Note-taking: a quick History

From text files to lightweight markup languages

Notes (and codes) that are archived but can evolve with version control systems

## Since note-taking concerns everyone. . .

- ▶ Since we are all "note-takers", our predecessors were also note-takers.
- ▶ This elementary observation will lead us to "study" how our brilliant ancestors took notes.
- ▶ Hopefully, we can learn some useful techniques on the way and put them on daily use.
- ▶ Hopefully, we can avoid thinking that we are the first to face the kind of problem we are now facing: "information overload".

# What are we going to talk about?

- ▶ The practical aspect of note-taking—what Historians dub "materiality"—.
- ▶ The books and notes organization.
- ▶ The link between the concrete and organizational aspects.

We are going to discuss a lot books organization since the "navigation devices" designed for the latter:

- ▶ table of content,
- ▶ index,
- ▶ etc,

also applies to notes.

# The concrete aspects summarized on a single slide

The collage illustrates the transition from ancient artifacts to modern technology, representing the evolution of concrete aspects across different eras and media.

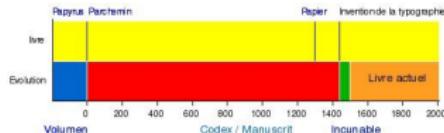
- Top Left:** A terracotta tile featuring a grid of symbols, possibly a game board or a calendar, from ancient Mesopotamia.
- Top Middle:** A Roman fresco depicting a man and a woman, likely a couple, from the Pompeii archaeological site.
- Top Right:** A handwritten page from a notebook with sketches and notes, possibly a technical or artistic drawing.
- Bottom Left:** A handwritten ledger page for the "State Hospital, Buffalo," detailing financial transactions and land transfers.
- Bottom Right:** A modern tablet device displaying the Windows 8 start screen, representing contemporary digital technology.

# Wax tablet and stylus



Musée romain-germanique  
Cologne (Allemagne)  
Photos de Jacques Poitou

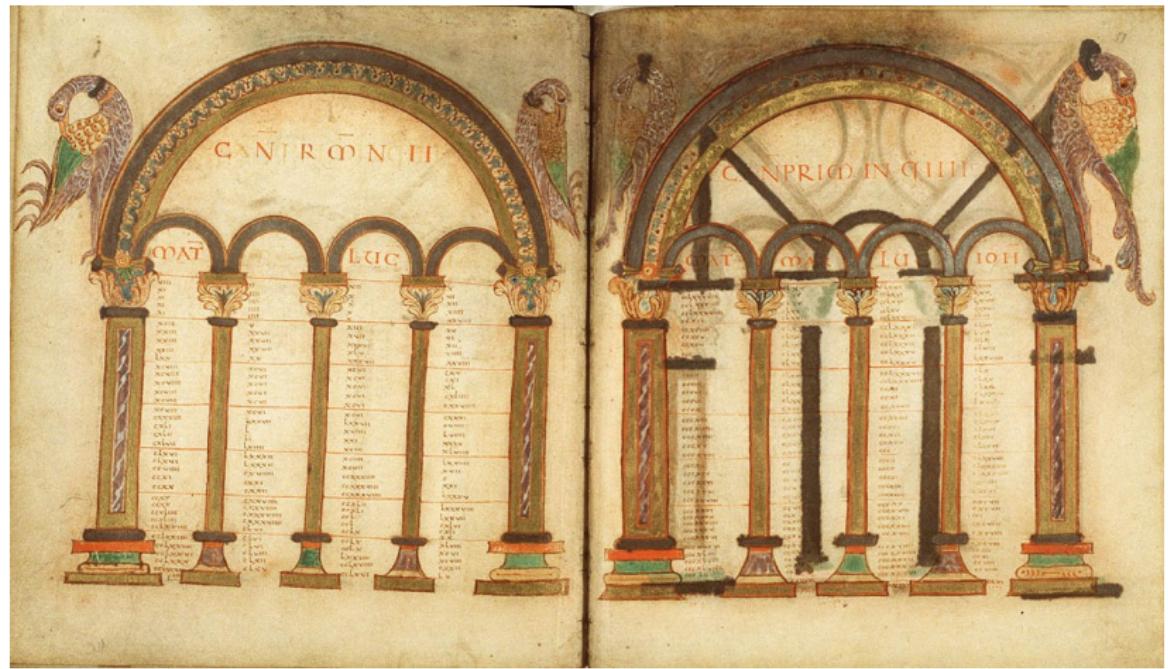
# From the scroll to the codex



# Eusebius and the invention of cross-references



# Eusebian canons



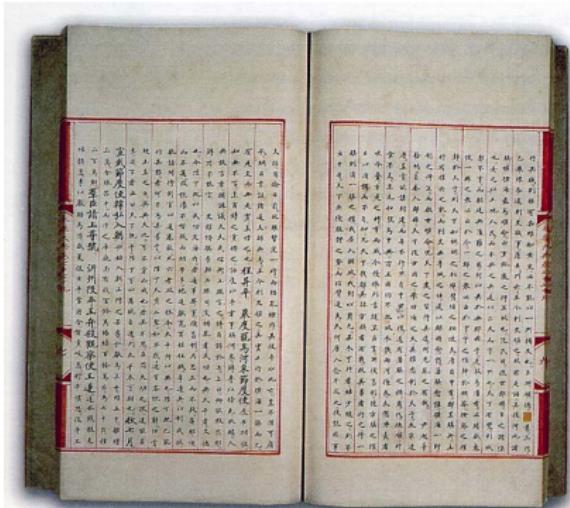
Fol. 10v and 11r of the Egmond Gospels. Canon tables (900 CE).

## *Codex* significance

Following Frédéric Barbier (*HISTOIRE DU LIVRE*, Armand Colin, 2009):

- ▶ *Codex* invention is crucial for the development of written civilization.
- ▶ The *codex* lends itself to **consultation reading**.
- ▶ We can add to the *codex* a "navigation system" making consultation easier.
- ▶ We can take notes while consulting a *codex*.
- ▶ The combination of the *codex* with the *Carolingian minuscule* constitutes an extremely powerful intellectual tools, never seen before.

# Let us not forget China

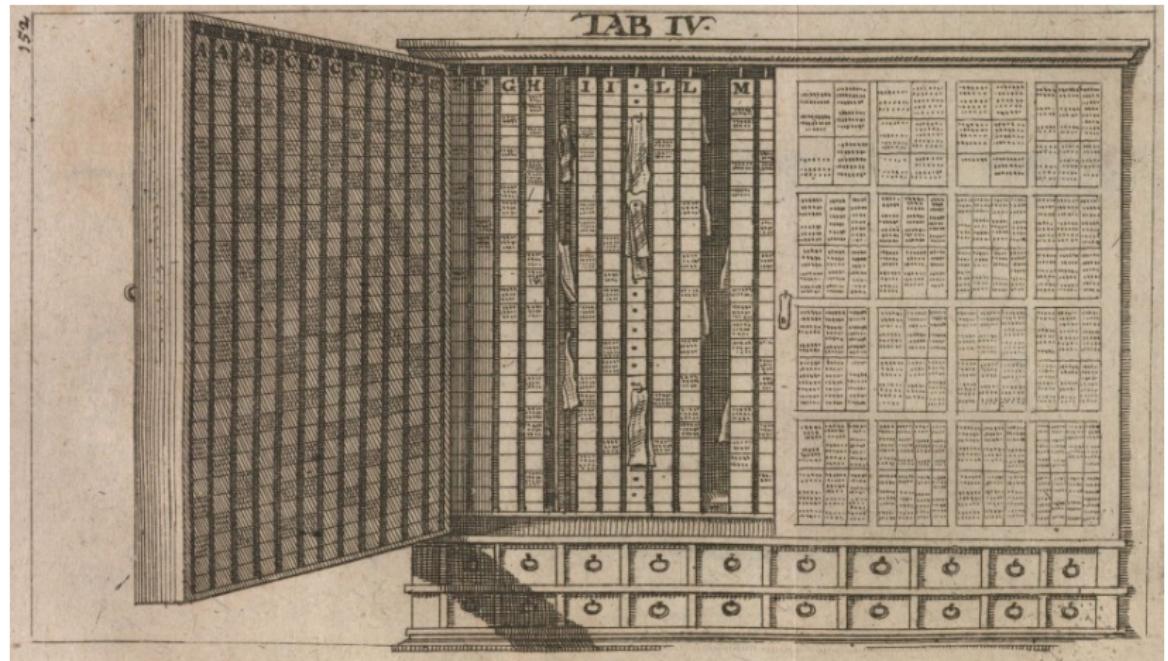


Top: Yongle Dadian (1403) contains 370 millions Chinese characters.



Right: Banknote printing plate from Northern Song Dynasty (960 – 1279).

# Getting organized by using the right slot



Placcius' closet again (Placcius, Vincent, 1642-1699. *De arte excerptendi vom gelahrten Buchhalten*, 1689. Houghton Library, Harvard University.)

# Constructing a notebook index the John Locke way

**Left Page Content:**

- Delta - POMC halfs
  - stack  $60 \times 8 \times 168$  38bit integer
  - time 168 38-bit floating-point
- stack
- 600 chip dimensions =  $60 \times 30$  pixels
- Day = Fura-2
- Excitation wavelength = 360 nm
- Resampling performed by = Andreas Pippal
- Preference = --
- Units = nm
- time
  - Simplification time = 502
  - Units = seconds
- FURA-PI
  - T\_310 Br-Pipette
  - T\_310 Lambdab ("Lambdab bound")
  - T\_360 Lambdab ("Lambdab iso")
  - T\_380 Lambdab ("Lambdab free")
  - CINN C
  - SRO SRO
  - P roIP
  - PA roIP
  - EXTRINSIC opName

**Right Page Content:**

Structure dos bloques da Calibratice

- #.objName → "Fura2" (color)
- #.BrPipette → 400 (nm)
- #.Lambdab → c("Lambdab\_bound": 310, "Lambdab\_free": 360, "Lambdab\_iso": 380)
- #T →  $\text{list}(c(360^{\circ}, 0.01, 380^{\circ}, 0.02, 360^{\circ}, 0.03))$
- #opName → "D09393 EXP"
- #E → 0.146
- #S\_P → 16.4
- #B\_P → 12
- #Transients → c("blast", "stim1", "stim2", "stim3")
- #TransientType → c("leading curve", "stimulus")
- #obj →
- #global →
- #roi → ROI<sub>global</sub>, ROI<sub>310,360</sub>
- #rob → ROI<sub>310,360,1</sub>, ROI<sub>310,360,2</sub>
- #roi3d\_mean → average(ROI) ?
- #roi3d\_norm → -
- #SRO → color(255, 255, 255)
- #PA → color(255, 255, 255)
- #cell, maskcell → c(96, 111, 192)
- #live → numeric(1000)
- #NCell → numeric(1000)
- #RD → c(0.14, 0.01, 0.03) *post stim scaling*

Code / Neuro / Calcium

Code / Neuro / Calcium

My own notebook is used here for illustration.

# Locke's method continued

A	a 6, 17, 52, 83, 80, 115, 151	J	a e o u
B	a e i o u	K	a e i o u
C	a 5, 7-9, 31, 33, 40, 81, 86, 89, 102-103 e 11-16, 46, 55, 65-63 i 10, 24, 25, 29, 56, 76, 79, 80, 88-93, 102, 114, 117-123, 125, 127-128, 152 o u	L	a e i o u
D	a e i o u 11-16, 44, 55, 62-64	M	a e i o u
E	a e i o u	N	a e i o u 36, 7-9, 11-16, 18, 25-27, 42, 55, 58, 64, 69, 86, 107, 115, 117-119
F	a e i o u	O	a e i o u
G	a e i o u	P	a e i o u 23, 35-38, 44-54, 64-67, 70, 77, 91, 91-92, 103-105, 116-117, 119, 126
H	a e i o u	Q	a e i o u
I	a e i o u 10, 72-75, 151 o 53, 93	R	a e i o u

The last pages of my notebook with the index.

## Conclusions of the historical overview

Since it is hard (for me at least) to use paper as a medium for note-taking, learning from "Newton's giants" should save us from reinventing the wheel (and getting it square).

We should nevertheless use numerical medium as much as possible (while keeping in mind what we just learned) since it provides:

- ▶ more organizational and structural flexibility,
- ▶ reliable archiving tools,
- ▶ powerful indexing tools.

# Where are we ?

Introduction

What is needed for Reproducible Research?

Note-taking concerns everyone

Note-taking: a quick History

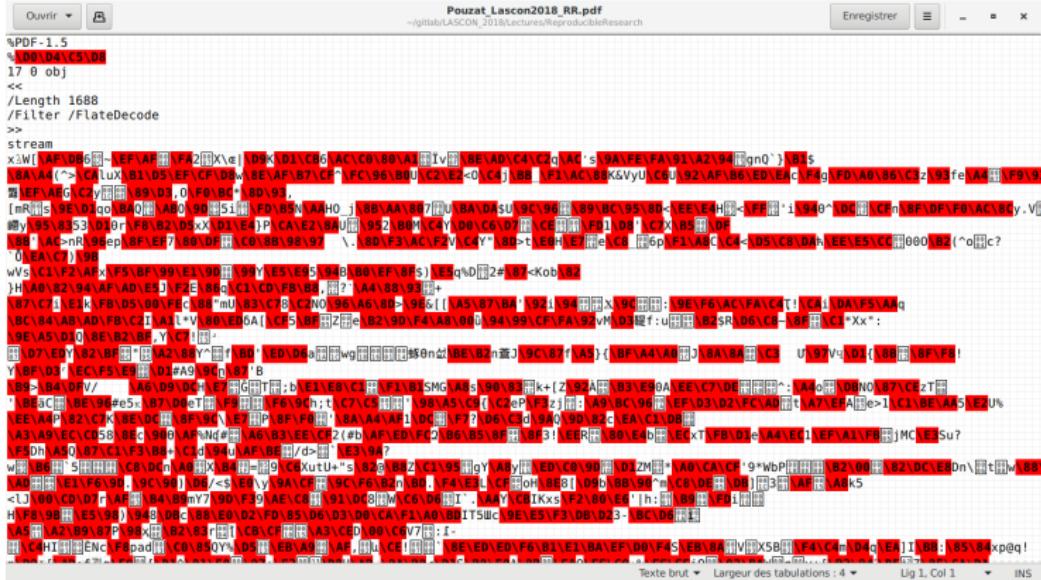
From text files to lightweight markup languages

Notes (and codes) that are archived but can evolve with version control systems

## What is a *text file* or *text format*?

- ▶ From a practical viewpoint, a text files *gives something readable* when opened with a text editor.
- ▶ A text editor enables us to create and modify text files (nice circular definition!). It's a software like:
  - ▶ Notepad++ for Windows,
  - ▶ gedit for Unix/Linux systems (but it also runs on the other two),
  - ▶TextEdit for MacOS.
- ▶ I'm mentioning only open source software since it is hard to do genuinely reproducible research with anything else.
- ▶ A word processor is more sophisticated than a *text editor*.
- ▶ **Warning** the native format used by word processor is rarely a *text format*. Word's doc and docx files and Libreoffice odt files *are not text files*.

Example of a file that cannot be read with a text editor



A pdf file (the file shown right now with a pdf reader) opened with gedit.

# A text file opened with a text editor

The screenshot shows a text editor window with the following details:

- Title Bar:** Pouzat\_Lascon2018\_RR.md  
-gitlab/LASCON\_2018/lectures/reproducibleResearch
- Buttons:** Ouvrir, Enregistrer, Minimize, Maximize, Close.
- Content Area:**

```
Introduction
=====
The usual presentation of research work
-----
> An article about computational science in a scientific publication is
> **not** the scholarship itself, it is merely **advertising** of the
> scholarship. The actual scholarship is the complete software
> development environment and the complete set of instructions which
> generated the figures.

\vspace{0.2cm}

Thoughts of Jon Claerbout \"distilled\" by Buckheit & Donoho (1995).

What do we do as manuscript \"producers\"?

The preparation of manuscripts and reports in experimental or
computational neuroscience often involves:

- Doing the experiments or coding and doing the simulations.
- A lot of data analysis.
- Careful design and realization of figures and tables.

Simulations involve:

- Algorithmic choices.
- Software choices or implementation.
- Parameters setting (*e.g.*., the time step used for conductance based
neuronal models).

Data analysis involves:

- Some data \"preprocessing\" (often).
```
- Bottom Bar:** Markdown ▾ Largeur des tabulations : 4 ▾ Lig 1, Col 1 ▾ INS

A markdown file (a source file for this lecture) opened with gedit.

# Why should we use text files?

Characters contained in text files are now typically encoded in UTF-8.

This implies that:

- ▶ It is "always" possible to read these files with a text editor *even years after their creation*.
- ▶ Desktop search and Concurrent Versions Systems (CVS) software work *very efficiently* with them.

Unless you run into serious memory problems, use text files, always.

## Problems with simple text files

- ▶ The "simple" text file precludes the use of nice navigation tools like hyperlinks.
- ▶ It is not possible to emphasize a word with a **bold** or an *italic* font.
- ▶ If several persons work on the same text, they can't correct each other by ~~striking through~~ text.

These limitations, combined to the benefits of text files led computer scientists to develop markup languages.

# A trivial example is the HTML language.

The screenshot shows a Qutebrowser window with the following details:

- Title Bar:** HTML - Wikipedia - qutebrowser
- Tab Bar:** 1: MargusPanes, 2: Christophe Pouzet, 3: LaTeX/Text Formatting, 4: HTML - Wikipedia, 5: striketh.., 6: cheatsheet-big.png (35), Not logged in, Talk, Contributions, Create account, Log in
- Header:** Article, Read, View source, View history, Search Wikipedia, Lock icon
- Page Content:**
  - Section:** HTML
  - From Wikipedia, the free encyclopedia
  - Text: ".htm" and ".html" redirect here. For other uses, see [HTML](#). For the use of HTML on Wikipedia, see [Help:HTML in wikitext](#).
  - Section:** Hypertext Markup Language (HTML) is the standard markup language for creating web pages and web applications. With Cascading Style Sheets (CSS) and JavaScript it forms a triad of cornerstone technologies for the World Wide Web.<sup>[3]</sup> Web browsers receive HTML documents from a web server or from local storage and render them into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for the appearance of the document.
  - Text:** HTML elements are the building blocks of HTML pages. With HTML constructs, images and other objects, such as interactive forms, may be embedded into the rendered page. It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items. HTML elements are delineated by tags, written using angle brackets. Tags such as <img /> and <input /> introduce content into the page directly. Others such as <p>...</p> surround and provide information about document text and may include other tags as sub-elements. Browsers do not display the HTML tags, but use them to interpret the content of the page.
  - Text:** HTML can embed programs written in a scripting language such as JavaScript which affect the behavior and content of web pages. Inclusion of CSS defines the look and layout of content. The World Wide Web Consortium (W3C), maintainer of both the HTML and the CSS standards, has encouraged the use of CSS over explicit presentational HTML since 1997.<sup>[4]</sup>
  - Table:** A table showing the file format details of HTML.
- Table:** HTML (Hypertext Markup Language)

<!DOCTYPE html>	.html
<html>	.htm
<!-- created 2010-01-01 -->	
<head>	
<title>sample</title>	
</head>	
<body>	
<p>Voluptatem accusantium totam rem aperiam.</p>	
</body>	
</html>	
- Table Headers:** Filename extension, Internet media type, Type code, Developed by, Initial release, Latest release
- Table Data:** .html, text/html, TEXT, W3C & WHATWG, 1993; 25 years ago, 5.1 2nd Edition<sup>[1]</sup> / 5.2 (working draft)<sup>[2]</sup> (1 November 2016; 14)
- Page Address:** https://en.wikipedia.org/wiki/HTML

Wikipedia HTML page viewed with qutebrowser web browser.

# An HTML file opened with a text editor



The screenshot shows a text editor window titled "HTML - Wikipedia.html". The content is the source code of a Wikipedia page about HTML. The code is heavily annotated with color-coded syntax highlighting, primarily in purple and pink, which highlights various HTML tags and attributes. The text describes the history and evolution of HTML, mentioning its standardization, the creation of web pages, and its role as a cornerstone technology. It also discusses the relationship between HTML, CSS, and JavaScript, and the introduction of semantic web technologies. The code is well-formatted with proper indentation and line breaks.

```
</ul>
</div>
</td>
</tr>
</table>
<p><b>Hypertext Markup Language</b> (<b>HTML</b>) is the standard <a href="/wiki/Markup_language" title="Markup language">markup language</a> for creating <a href="/wiki/Web_page" title="Web page">web pages</a> and <a href="/wiki/Web_application" title="Web application">web applications</a>. With <a href="/wiki/Cascading_Style_Sheets" title="Cascading Style Sheets">Cascading Style Sheets</a> (CSS) and <a href="/wiki/JavaScript" title="JavaScript">JavaScript</a> it forms a triad of cornerstone technologies for the <a href="/wiki/World_Wide_Web" title="World Wide Web">World Wide Web</a>. <sup id="cite_ref-3" class="reference"><a href="#cite_note-3">[3]</a></sup> <a href="/wiki/Web_browser" title="Web browser">Web browsers</a> receive HTML documents from a <a href="/wiki/Web_server" title="Web server">web server</a> or from local storage and render them into multimedia web pages. HTML describes the structure of a web page <a href="/wiki/Semantic_Web" title="Semantic Web">semantically</a> and originally included cues for the appearance of the document.</p>
<p><a href="/wiki/HTML_element" title="HTML element">HTML elements</a> are the building blocks of HTML pages. With HTML constructs, <a href="/wiki/HTML_element#Images_and_objects" title="HTML element">images</a> and other objects, such as <a href="/wiki/Fieldset">form fields</a>, may be embedded into the rendered page. It provides a means to create <a href="/wiki/Structured_document" title="Structured document">structured documents</a> by denoting structural <a href="/wiki/Semantics" title="Semantics">semantics</a> for text such as headings, paragraphs, lists, <a href="/wiki/Hyperlink" title="Hyperlink">links</a>, quotes and other items. HTML elements are delineated by <i>tags</i>, written using <a href="/wiki/BracketAngle_brackets" title="Bracket angle brackets">angle brackets</a>. Tags such as <code class="nowrap" style="">&ampltimg /></code> and <code class="nowrap" style="">&ampltinput /></code> introduce content into the page directly. Others such as <code class="nowrap" style="">&ampltp>&lt;/p&gt;</code> surround and provide information about document text and may include other tags as sub-elements. Browsers do not display the HTML tags, but use them to interpret the content of the page.</p>
<p>HTML can embed programs written in a <a href="/wiki/Scripting_language" title="Scripting language">scripting language</a> such as <a href="/wiki/JavaScript" title="JavaScript">JavaScript</a> which affect the behavior and content of web pages. Inclusion of CSS defines the look and layout of content. The <a href="/wiki/World_Wide_Web_Consortium" title="World Wide Web Consortium">World Wide Web Consortium</a>, maintainer of both the HTML and the CSS standards, has encouraged the use of CSS over explicit presentational HTML since 1997. <sup id="cite_ref-deprecated-4-0" class="reference"><a href="#cite_note-deprecated-4">[4]</a></sup></p>
<p></p>
<div id="toc" class="toc">
<div class="toctitle" lang="en" dir="ltr">
<h2>Contents</h2>
</div>
<ul>
<li class="toclevel-1 tocsection-1"><a href="#History"><span class="tocnumber">1</span> <span class="toctext">History</span></a>


```

The Wikipedia HTML page opened with gedit. Markup languages were not designed to be read by humans.

We can summarize our problem as follows:

- ▶ Text files are attractive for note-taking.
- ▶ Markup languages provide a much better "reading experience" when viewed with the proper "browser".
- ▶ Markup language files are text files, **but** require usually a dedicated editing software if we want to modify them.

Is it possible to combine the benefits of "simple" text files with the reading comfort of markup languages?

# Lightweight markup languages: the idea

A lightweight markup language is:

- ▶ A *markup language* with a **simple syntax**.
- ▶ A language designed to be **easily edited** with a *text editor*.
- ▶ A language **easily read** without a browser.

# Markdown's example

Text using Markdown syntax	Corresponding HTML produced by a Markdown processor	Text viewed in a browser
<pre>Heading ----- ## Sub-heading  Paragraphs are separated by a blank line.  Two spaces at the end of a line leave a line break.  Text attributes <i>italic</i>, **bold**, <code>monospace</code>.  Horizontal rule: ---  Bullet list: <ul style="list-style-type: none"> <li>apples</li> <li>oranges</li> <li>pears</li> </ul>  Numbered list: <ol style="list-style-type: none"> <li>wash</li> <li>rinse</li> <li>repeat</li> </ol>  A [link](http://example.com).  ![Image](image_icon.png)  &gt; Markdown uses email-style &gt; characters for blockquoting.</pre>	<pre>&lt;h1&gt;Heading&lt;/h1&gt; &lt;h2&gt;Sub-heading&lt;/h2&gt; &lt;p&gt;Paragraphs are separated by a blank line.&lt;/p&gt; &lt;p&gt;Two spaces at the end of a line&lt;br /&gt; leave a line break.&lt;/p&gt; &lt;p&gt;Text attributes &lt;i&gt;italic&lt;/i&gt;, &lt;strong&gt;bold&lt;/strong&gt;, &lt;code&gt;monospace&lt;/code&gt;.&lt;/p&gt; &lt;p&gt;Horizontal rule:&lt;/p&gt; &lt;hr /&gt; &lt;p&gt;Bullet list:&lt;/p&gt; &lt;ul&gt; &lt;li&gt;apples&lt;/li&gt; &lt;li&gt;oranges&lt;/li&gt; &lt;li&gt;pears&lt;/li&gt; &lt;/ul&gt; &lt;p&gt;Numbered list:&lt;/p&gt; &lt;ol&gt; &lt;li&gt;wash&lt;/li&gt; &lt;li&gt;rinse&lt;/li&gt; &lt;li&gt;repeat&lt;/li&gt; &lt;/ol&gt; &lt;p&gt;A &lt;a href="http://example.com"&gt;link&lt;/a&gt;.&lt;/p&gt; &lt;p&gt;&lt;img alt="Image" src="Image_icon.png" /&gt;&lt;/p&gt; &lt;blockquote&gt; Markdown uses email-style &gt; characters for blockquoting. &lt;/blockquote&gt;</pre>	<p>Heading</p> <p>Sub-heading</p> <p>Paragraphs are separated by a blank line.</p> <p>Two spaces at the end of a line leave a line break.</p> <p>Text attributes <i>italic</i>, <b>bold</b>, <code>monospace</code>.</p> <p>Horizontal rule:</p> <p>Bullet list:</p> <ul style="list-style-type: none"><li>apples</li><li>oranges</li><li>pears</li></ul> <p>Numbered list:</p> <ol style="list-style-type: none"><li>wash</li><li>rinse</li><li>repeat</li></ol> <p>A link<a href="http://example.com">link</a>.</p> <p></p> <p>Markdown uses email-style &gt; characters for blockquoting.</p>

The syntax basics from Wikipedia, see also "Mastering Markdown" (a 3 min read) from GitHub.

# Markdown is not the only lightweight markup language

Among the "most popular":

- ▶ MediaWiki used by Wikipedia (but files are not stored in text format!).
- ▶ DokuWiki like MediaWiki but stored in text format.
- ▶ reStructuredText used for the python documentation.
- ▶ AsciiDoc.
- ▶ Org mode, my favorite, but it requires learning emacs (a good thing to do, if you have time for it).

The good news is that you don't need to be too nervous about choosing the "right" language, thanks to pandoc you can convert one into any other!

## Section's summary

Thanks to lightweight markup languages we will be able to:

- ▶ Work mostly with text files.
- ▶ Write our notes quickly with any editor.
- ▶ Organize our notes.

# Where are we ?

Introduction

What is needed for Reproducible Research?

Note-taking concerns everyone

Note-taking: a quick History

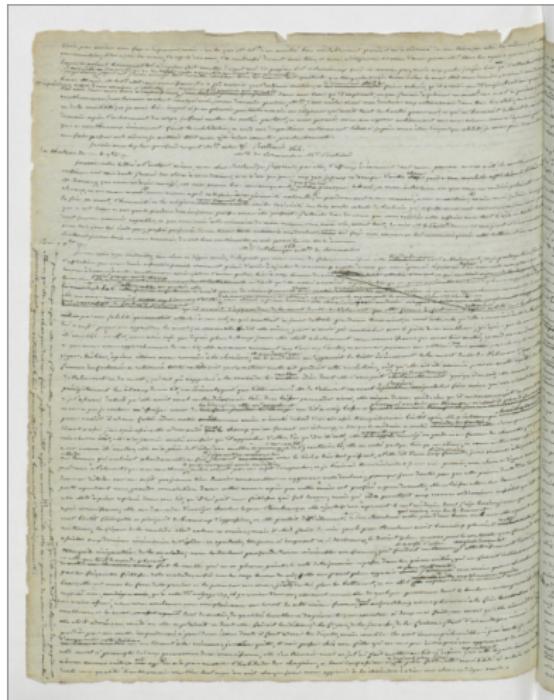
From text files to lightweight markup languages

Notes (and codes) that are archived but can evolve with version control systems

## Introduction of this section

- ▶ The tools we are going to discuss should appeal to a much wider audience than the reproducible research community.
- ▶ Anyone working on a text is concerned, even more so when this work is done in collaboration.
- ▶ **The longevity issue of notes and texts is in noway new.**
- ▶ The humanists and scholars of the early modern period who specialized in text compilations were literally obsessed by this problem and used it to justify their work.
- ▶ Their solution was to use multiple copies, as we now do with a different medium.
- ▶ We should nevertheless remain humble, the paper (and parchment) medium used by humanists as demonstrated its capability to last.
- ▶ When it comes to making notes evolve, I think we can say that some real progress was recently made.

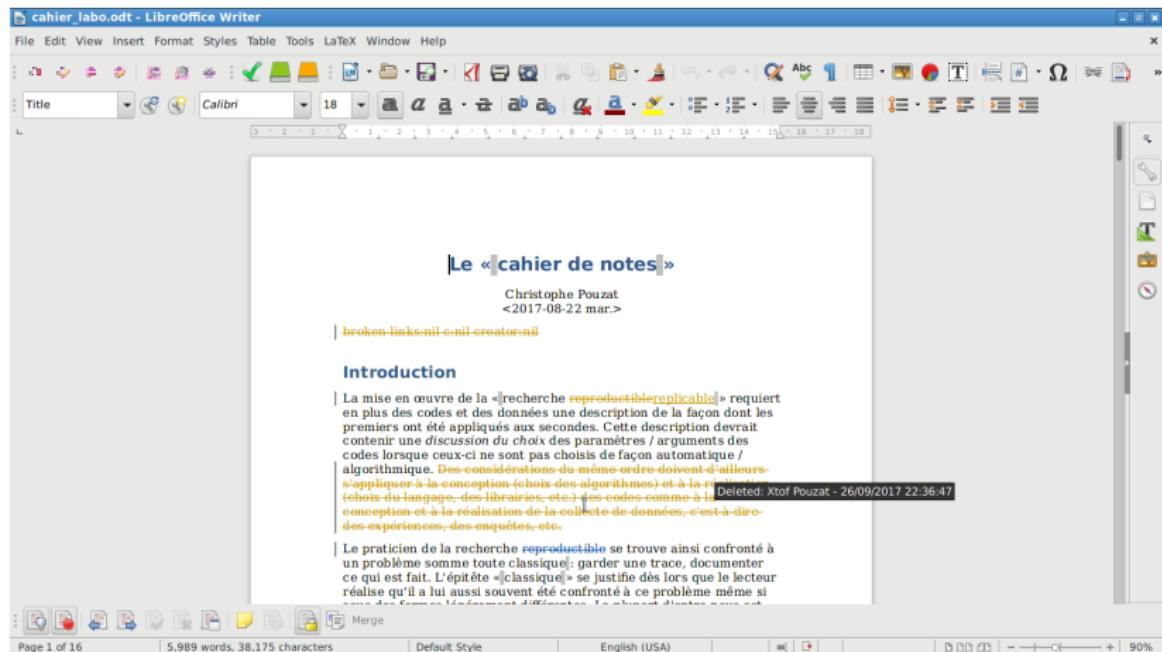
# The nightmare: changing a text on paper medium



Manuscript of *Dangerous Liaisons (Les liaisons dangereuses)* by Pierre Choderlos de Laclos (p. 258, BNF Gallica).

There is clearly a very limited number of changes one can bring in that way!

# Changing a text with a word processor



An early version of this lecture (in French) edited with LibreOffice.

# Making change with a "wiki engine"

The screenshot shows a DokuWiki page titled "les\_notes\_sur\_le\_cahier\_de\_notes [Notes] - Conkeror". The page content is a list of historical notes, many of which are links to external sources like Commons. The page has a "Modifier" button at the top right and a vertical toolbar on the right side with icons for edit, history, and more. The footer includes a link to the page's XML file, a copyright notice for CC Attribution-Noncommercial-Share Alike 4.0 International, and a footer bar with various icons.

jete, car il peut être nécessaire pour expliquer des données antérieures, considérées ultérieurement comme étranges. [...]

Le but de toute cette pratique de prise de notes est de préserver la valeur [le temps et les moyens humains et matériels investis dans la recherche]. Elle devrait être soigneusement conçue pour s'adapter aux conditions de chaque laboratoire et devraient être adéquate mais pas trop élaborés. Si l'on exige trop de la nature humaine, le système ne fonctionnera pas.

**Histoire**

- \* ⓘ Une carte des provinces romaines, la référence est « CC BY-SA 3.0, ⓘ <https://commons.wikimedia.org/w/index.php?curid=2249662> » ;
- \* ⓘ Césarée (Wikipédia) ;
- \* ⓘ Moulin à maitlets de l'Encyclopédie de Diderot et d'Alembert ;
- \* ⓘ Moulin à maitlets de « L'ART DE FAIRE LE PAPIER » de sur le site du Moulin du Verger.
- \* ⓘ Canon eusebien (Wikipédia), un ⓘ exemple avec 4 colonnes (Domaine public, ⓘ <https://commons.wikimedia.org/w/index.php?curid=108009>), Evangéliaire de Lorsch, 778–820, canon I. et ⓘ un autre avec 3 (Domaine public, ⓘ <https://commons.wikimedia.org/w/index.php?curid=108011>) Evangéliaire de Lorsch canon II. Un portrait d'Eusèbe : Domaine public, ⓘ <https://commons.wikimedia.org/w/index.php?curid=643569> ;
- \* ⓘ Un numéro hors série de « Extrême-Orient, Extrême-Occident » sur « Qu'était-ce qu'écrire une encyclopédie en Chine ? » ;
- \* ⓘ Book Culture and Textual Transmission in Sung China, un article de Susan Chernick cité par Ann Blair (p 31) à propos du lien entre matérialité du livre (passage du rouleau au codex) et « l'explosion » des ⓘ leishu dans la Chine du 9e siècle ;
- \* ⓘ Un article (en anglais) sur les common place book dans Wikipedia ;
- \* ⓘ Indexing commonplace books: John Locke's method de Alan Walker ( ⓘ PDF annoté) ;
- \* ⓘ Exemple de John Locke ; sur Locke, « papa du libéralisme » et actionnaire de la Royal African Company principale compagnie négrière britannique, voir l' ⓘ article de Wikipedia en anglais, Philip D. Curtin *The Atlantic Slave Trade: A Census* (University of Wisconsin Press, 1969, p. 121-123), l' ⓘ article de Wikipedia en français sur le bonhomme et « Contre-histoire du libéralisme » de Domenico Losurdo (La Découverte / Poche, 2014, p. 34-36) ;
- \* ⓘ les fiches de Linnaeus ;

les\_notes\_sur\_le\_cahier\_de\_notes.txt - Dernière modification: 2017/10/01 16:47 par xf0f

Sauf mention contraire, le contenu de ce wiki est placé sous les termes de la licence suivante : ⓘ CC Attribution-Noncommercial-Share Alike 4.0 International

[http://localhost/dokupwki/doku.php?id=les\\_notes\\_sur\\_le\\_cahier\\_de\\_notes#histoire](http://localhost/dokupwki/doku.php?id=les_notes_sur_le_cahier_de_notes#histoire)  
Link: <https://commons.wikimedia.org/w/index.php?curid=2249662>

The personal wiki (using the dokupwki engine) I experienced while preparing the French version of this lecture.

les\_notes\_sur\_le\_cahier\_de\_notes - Anciennes révisions [Notes] - Conkeror

## Anciennes révisions

Voici les anciennes révisions de la page en cours. Pour revenir à une ancienne révision, sélectionnez-la ci-dessous, cliquez sur le bouton « Modifier cette page » et enregistrez-la.

<input type="checkbox"/> 2017/10/07 08:20 les_notes_sur_le_cahier_de_notes - [Prise de notes] xtof +128 B (Version actuelle)
<input type="checkbox"/> 2017/10/06 17:59 les_notes_sur_le_cahier_de_notes - [Prise de notes] xtof (127.0.0.1) +948 B
<input checked="" type="checkbox"/> 2017/10/01 16:47 les_notes_sur_le_cahier_de_notes - [Histoire] xtof (127.0.0.1) +98 B
<input checked="" type="checkbox"/> 2017/10/01 16:32 les_notes_sur_le_cahier_de_notes - [Histoire] xtof (127.0.0.1) +95 B
<input type="checkbox"/> 2017/10/01 15:47 les_notes_sur_le_cahier_de_notes - [Histoire] xtof (127.0.0.1) +69 B
<input type="checkbox"/> 2017/09/30 14:09 les_notes_sur_le_cahier_de_notes - [Histoire] xtof (127.0.0.1) +584 B
<input type="checkbox"/> 2017/09/30 13:20 les_notes_sur_le_cahier_de_notes - [Histoire] xtof (127.0.0.1) +101 B
<input type="checkbox"/> 2017/09/30 11:08 les_notes_sur_le_cahier_de_notes - [Histoire] xtof (127.0.0.1) +2 B
<input type="checkbox"/> 2017/09/29 20:56 les_notes_sur_le_cahier_de_notes - [Histoire] xtof (127.0.0.1) +0 B
<input type="checkbox"/> 2017/09/29 20:25 les_notes_sur_le_cahier_de_notes - [Histoire] xtof (127.0.0.1) +49 B
<input type="checkbox"/> 2017/09/29 20:08 les_notes_sur_le_cahier_de_notes - [Histoire] xtof (127.0.0.1) +216 B
<input type="checkbox"/> 2017/09/29 19:55 les_notes_sur_le_cahier_de_notes - [Histoire] xtof (127.0.0.1) +123 B
<input type="checkbox"/> 2017/09/29 18:03 les_notes_sur_le_cahier_de_notes - [Histoire] xtof (127.0.0.1) +363 B
<input type="checkbox"/> 2017/09/29 17:15 les_notes_sur_le_cahier_de_notes - [Histoire] xtof (127.0.0.1) +238 B
<input type="checkbox"/> 2017/09/29 14:49 les_notes_sur_le_cahier_de_notes - [Histoire] xtof (127.0.0.1) +534 B
<input type="checkbox"/> 2017/09/29 14:27 les_notes_sur_le_cahier_de_notes - [Histoire] xtof (127.0.0.1) +294 B
<input type="checkbox"/> 2017/09/29 13:11 les_notes_sur_le_cahier_de_notes xtof (127.0.0.1) +212 B
<input type="checkbox"/> 2017/09/27 09:01 les_notes_sur_le_cahier_de_notes xtof (127.0.0.1) -521 B
<input type="checkbox"/> 2017/09/25 10:07 les_notes_sur_le_cahier_de_notes xtof (127.0.0.1) -8 B
<input type="checkbox"/> 2017/09/24 21:33 les_notes_sur_le_cahier_de_notes - créée xtof (127.0.0.1) +415 kB

Differences entre les versions sélectionnées

les\_notes\_sur\_le\_cahier\_de\_notes.txt - Dernière modification: 2017/10/07 08:20 par xtof

Sauf mention contraire, le contenu de ce wiki est placé sous les termes de la licence suivante : CC Attribution-Noncommercial-Share Alike 4.0 International

http://localhost/dokuwiki/doku.php?id=les\_notes\_sur\_le\_cahier\_de\_notes&do=revisions

Done 15:47 (100, 78) | form/checkbox

Clicking *previous versions (anciennes révisions)* gives access to the list of changes done when and by whom. If I now select two versions...

les\_notes\_sur\_le\_cahier\_de\_notes [Notes] - Conkeror

Connecté en tant que : Christophe Pouzat (xtof) | Administrer | Mettre à jour le profil | Se déconnecter

Rechercher

Derniers changements Gestionnaire Multimédia Plan du site

Piste: • start • les\_notes\_sur\_le\_cahier\_de\_notes

## Définitions

Ci-dessous, les différences entre deux révisions de la page.

Voir les différences : Côte à côté

Lien vers cette vue comparative

2017/10/01 16:47 xtof [Histoire]

les\_notes\_sur\_le\_cahier\_de\_notes [2017/10/01 16:47]  
xtof (127.0.0.1) [Histoire]

Ligne 159:

```
----- Histoire -----
+ * [[https://fr.wikipedia.org/wiki/Province_romaine#/media/File:Ronia_Imperio.png|Une carte des provinces romaines]], la référence est « CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=2249662 » ;
+ [[https://fr.wikipedia.org/wiki/C%C3%A9saree]] (Wikipédia) ;
+ [[http://www.planches.eu/planche.php?nom=PAPETTERIE&n=5|Moulin à maittels]] de l'Encyclopédie de Diderot et d'Alembert ;
```

2017/10/01 16:32 xtof [Histoire]

les\_notes\_sur\_le\_cahier\_de\_notes [2017/10/01 16:32]  
xtof (127.0.0.1) [Histoire]

Ligne 159:

```
----- Histoire -----
- * [[https://fr.wikipedia.org/wiki/Province_romaine#/media/File:Ronia_Imperio.png|Une carte des provinces romaines]] (Wikipédia) ;
- [[http://www.planches.eu/planche.php?nom=PAPETTERIE&n=5|Moulin à maittels]] de l'Encyclopédie de Diderot et d'Alembert ;
```

[http://localhost/dokuwiki/doku.php?id=les\\_notes\\_sur\\_le\\_cahier\\_de\\_notes&do=diff](http://localhost/dokuwiki/doku.php?id=les_notes_sur_le_cahier_de_notes&do=diff)

Done

16:31 (100, 0)

I see the differences between the two versions. You obtain the same thing on Wikipedia by clicking on *View History*.

## Pros and cons

- ▶ A solution with a strong record for collaborative projects (Wikipedia).
- ▶ A text format is used when working with Dokuwiki.
- ▶ A single page can be modified at a time.

# Version Control Systems

I now come to the most "sophisticated" solution:

- ▶ A dedicated software, git, is used to manage the successive versions of a **set** of files in **different formats** (text, images, etc.). In fact files arborizations can be managed.
- ▶ git like software require a repository, that can be built on the user's computer, but that is usually on a dedicated server like GitHub or GitLab.
- ▶ The repository allows several people to work on the same project and to exchange their modifications. Each project member has a **full copy** of the repository (dating back to his/her last *synchronization*).

LASCON\_2018

Commits · ff927c0b2ce3e5c8ad7cafc696653299894f5d19 · Christophe Pouzat / LASCON\_2018 · GitLab - qutebrowser

1: MarquesPages 1: 2: Commits · ff927c0b2ce3e5c8ad7cafc696653299894f5d19 1: 3: Git 1: 4: Version control - Wikipedia 1: 5: GitHub

GitLab Projects Groups Activity Milestones Snippets

Overview

Repository

Commits

Branches

Tags

Contributors

Graph

Compare

Charts

Locked Files

Registry

Issues

Merge Requests

<< Collapse sidebar

Christophe Pouzat > LASCON\_2018 > Commits

ff927c0b2ce3e5c... LASCON\_2018

18 Jan, 2018 1 commit

 Start git section  
Christophe Pouzat authored 5 minutes ago

ff927c0b   Browse Files

12 Jan, 2018 1 commit

 Evolving notes up to wiki  
Christophe Pouzat authored 6 days ago

a940200a   Browse Files

11 Jan, 2018 1 commit

 Lightweight markup languages added  
Christophe Pouzat authored a week ago

3560fd12   Browse Files

10 Jan, 2018 5 commits

 Historical part done.  
Christophe Pouzat authored a week ago

d8e3f608   Browse Files

 Placcius' closet again.  
Christophe Pouzat authored a week ago

e4bc2562   Browse Files

 Leishus done.  
Christophe Pouzat authored a week ago

f673ef8e   Browse Files

 Codex done.  
Christophe Pouzat authored a week ago

5239e6e3   Browse Files

 Mourning's week.

[https://gitlab.com/c\\_pouzat/LASCON\\_2018/commits/ff927c0b2ce3e5c8ad7cafc696653299894f5d19](https://gitlab.com/c_pouzat/LASCON_2018/commits/ff927c0b2ce3e5c8ad7cafc696653299894f5d19) [top] [x] [2/5]

The GitLab interface containing the files of this presentation.

Start git section (ff927c0b) · Commits · Christophe Pouzat / LASCON 2018 · GitLab - qutebrowser

1: MarquesPages 2: Start git section (ff927c0b) 3: Git 4: Version control - Wikipedia 5: GitHub

**GitLab** Projects Groups Activity Milestones Snippets This project Search + 9 - 3

L LASCON\_2018

Overview Repository Files Commits Branches Tags Contributors Graph Compare Charts Locked Files Registry Issues Merge Requests Collapse sidebar

Showing 1 changed file ▾

```

@@ -870,7 +870,7 @@ Thanks to lightweight markup languages we
will be able to:
- Write our notes quickly with any editor.
- Organize our notes.

@@ -874,7 +874,7 @@ Thanks to lightweight markup languages we
will be able to:
- Write our notes quickly with any editor.
- Organize our notes.

@@ -949,7 +949,7 @@ I started using
[[https://www.dokuwiki.org/dokuwiki#][dokuwiki]] for this lecture

#*END_EXPORT latex
\begin{document}
\begin{center}
\begin{table}[ht]
\begin{tr}
\td{Clicking /previous versions/ (/anciennes révisions/)} gives access
to the list of changes done when and by whom. If I now select two
versions...
\end{tr}
\end{table}
\end{center}
\end{document}
#*END_EXPORT

```

[https://gitlab.com/c\\_pouzat/LASCON\\_2018/commit/ff927c0b2ce3e5c8ad7cafef96653299894f5d19 \[27%\] \[<\] \[>\] \[2/5\]](https://gitlab.com/c_pouzat/LASCON_2018/commit/ff927c0b2ce3e5c8ad7cafef96653299894f5d19 [27%] [<] [>] [2/5])

Modifications are easily visualized...

The screenshot shows a GitLab project interface for 'LASCON\_2018'. The left sidebar lists various project sections: Overview, Repository (selected), Files, Commits, Branches, Tags, Contributors, Graph, Compare, Charts, Locked Files, Registry, Issues (0), Merge Requests (0), and a Collapse sidebar button. The main content area has a heading 'Markdown is not the only lightweight markup language'. Below it, a list of languages is provided:

- MediaWiki used by Wikipedia (but files are not stored in text format).
- DokuWiki like MediaWiki but stored in text format.
- reStructuredText used for the python documentation.
- AsciiDoc.
- Org mode, my favorite, but it requires learning emacs (a good thing to do, if you have time for it).

The text 'The good news is that you don't need to be too nervous about choosing the "right" language, thanks to pandoc you can convert one into any other!' follows. A section titled 'Section's summary' is present, followed by a note about the benefits of using lightweight markup languages.

**Notes that are archived but can evolve with version control systems**

**Introduction of this section**

- The tools we are going to discuss should appeal to a much wider audience than the reproducible research community.
- Anyone working on a text is concerned, even more so when this work is done in collaboration.
- The longevity issue of notes and texts is in noway new.
- The humanists and scholars of the early modern period who specialized in text compilations were literally obsessed by this problem and used it to justify their work.
- Their solution was to use multiple copies, as we now do with a different medium.

[https://gitlab.com/c\\_pouzat/LASCON\\_2018/blob/master/Lectures/ReproducibleResearch/Pouzat\\_Lascon2018\\_PR\\_slides.org](https://gitlab.com/c_pouzat/LASCON_2018/blob/master/Lectures/ReproducibleResearch/Pouzat_Lascon2018_PR_slides.org) [82%] [4] [3/3]

Text files entered with a lightweight markup language get automatically formatted (an example with org).

## Pros and cons

- ▶ A "sophisticated" approach that takes a bit more time to learn and master than the other two.
- ▶ A strong record for collaborative projects (Linux kernel, . . .).
- ▶ Can manage modifications on several files at once.
- ▶ A centralized version **copied** by each member of the project.

# Acknowledgements

This lecture owes a lot to feed-back from:

- ▶ Arnaud Legrand
- ▶ Konrad Hinsen
- ▶ Aurélie Bayle
- ▶ Marie-Hélène Comte