Lab books and note books

Christophe Pouzat
MAP5, Paris-Descartes University and CNRS
christophe.pouzat@parisdescartes.fr

September 3, 2018

Lab books and note books
Christophe Pouzat
MAP5, Paris-Descartes University and CNRS
christophe.pouzat@parisdescartes.fr
September 3, 2018

Outline

M1-S0: Lab books and note books

M1-S1: Note-taking concerns everyone

M1-S2: Note-taking: a quick history

M1-S3: From text files to lightweight markup languages

M1-S4: Notes (and codes) that are archived but can evolve with version control systems

M1-S5: Finding one's way with tags and desktop search application

Outline

M1-S0: Lab books and note books

M1-S1: Note-taking concerns everyone

M1-S2: Note-taking: a quick history

M1-S3: From text files to lightweight markup languages

M1-S4: Notes (and codes) that are archived but can evolve with version control systems

M1-S5: Finding one's way with tags and desktop search application

Where are we?

M1-S0: Lab books and note books

M1-S1: Note-taking concerns everyone

M1-S2: Note-taking: a quick history

M1-S3: From text files to lightweight markup languages

M1-S4: Notes (and codes) that are archived but can evolve with version control systems

M1-S5: Finding one's way with tags and desktop search application

Lab books and note books

1. Note-taking Concerns Everyone
2. A Quick History of Note Taking
3. From Text Files to Lightweight Markup Languages
   - Demo: markdown
4. Note Archiving and Evolution with Version Control
   - Demo: gitlab
5. Labels and Search Engines
   - Demo: DocFetcher

Lab books and note books

1. Note-taking Concerns Everyone
2. A Quick History of Note Taking
3. From Text Files to Lightweight Markup Languages
   - Demo: markdown
4. Note Archiving and Evolution with Version Control
   - Demo: gitlab
5. Labels and Search Engines
   - Demo: DocFetcher

Where are we?

M1-S0: Lab books and note books

M1-S1: Note-taking concerns everyone

M1-S2: Note-taking: a quick history

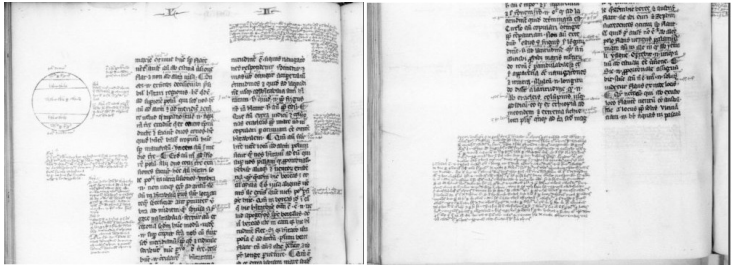
M1-S3: From text files to lightweight markup languages

M1-S4: Notes (and codes) that are archived but can evolve with version control systems

M1-S5: Finding one's way with tags and desktop search application

Notes This section discusses a much wider issue than *reproducible research* (RR). Implementing RR requires thorough note-taking and note-taking concerns everyone. The purpose of this section is therefore to remind the reader / auditor that he/she already knows: note-taking concerns everyone. Few examples are used to that end.

The scholar annotating his book / manuscript



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

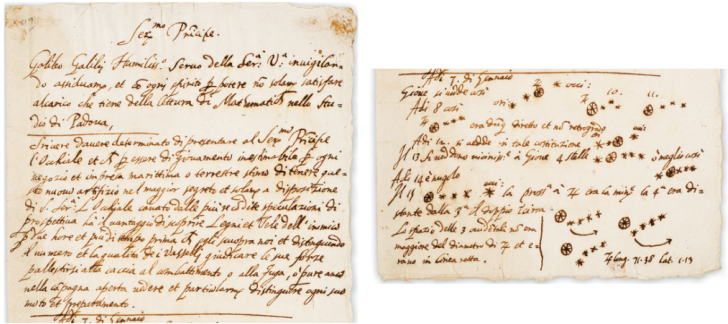
The scholar annotating his book / manuscript

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

Notes We see a manuscript from the XIVth century heavily annotated by its owner Nicasius de Planca. This kind of note-taking was and remains extremely common. You should nevertheless avoid it when reading books from a library or from your friends!

The next two slides show a case of paramount importance for the History of Science.

Galileo observing Jupiter's moons



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

Galileo observing Jupiter's moons

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

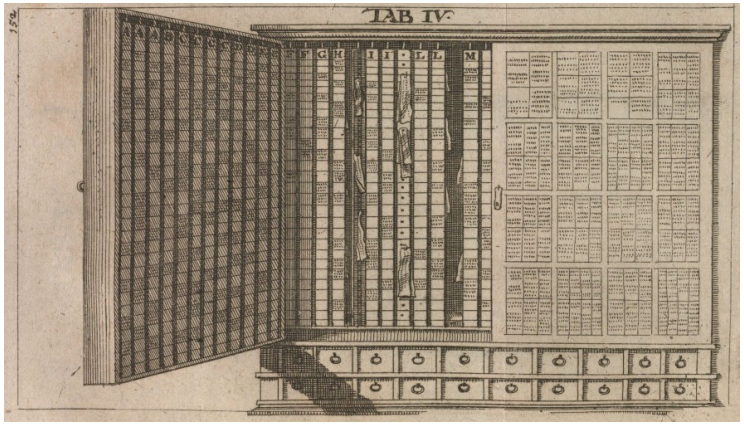
Notes The first observation was done on January 7 1610. Galileo Galilei first thought that he found new stars close to Jupiter (see the Wikipedia page). But after several nights of observation, he realized that these "stars" were in fact circling around the planet, they are satellites! He named the group of four the Medicean stars, in honour of his future patron, Cosimo II de' Medici, Grand Duke of Tuscany, and Cosimo's three brothers (Wikipedia).

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

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

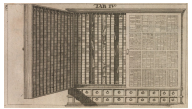
Notes These observations lead Galileo to reject the geocentric hypothesis in favor of the heliocentric one. This brought him much later, and after a somewhat tortuous path that I don't have the space to describe now, in front of the Inquisition that sentences him on June 22 1633 to house arrest, which he remained under for the rest of his life.

Placcius' and Leibniz' closet



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

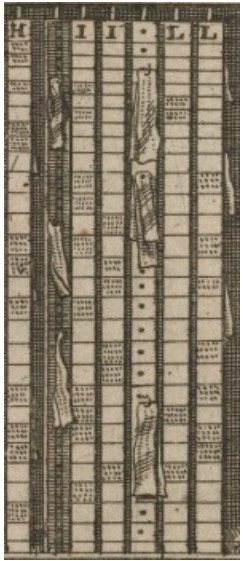
Placcius' and Leibniz' closet



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

Notes With printing appearance, demand for paper increased and paper's price ended up decreasing (after a large production increase). In addition to the use of the *codex* with pages made of paper, many scholars started using paper slips.

But taking abundant notes on paper slips is good only if one can find efficiently retrieve this stored information when needed. Vincent Placcius (1642-1699) and Gottfried Leibniz (1646-1716) had a custom made closet to solve this retrieval problem. This example is discussed in Ann Blair's book *TOO MUCH TO KNOW*, Yale Univ. Press, 2010 (pp. 93-95).



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

Notes This cabinet had many columns that could rotate about their (vertical) axis. The column's front was used to write what we would now call keywords relating to the content of the notes that were hooked on the column's back side.

Notice the advantage of these paper slips over Galileo's codex: with the former, notes can be reorganized.

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.

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.

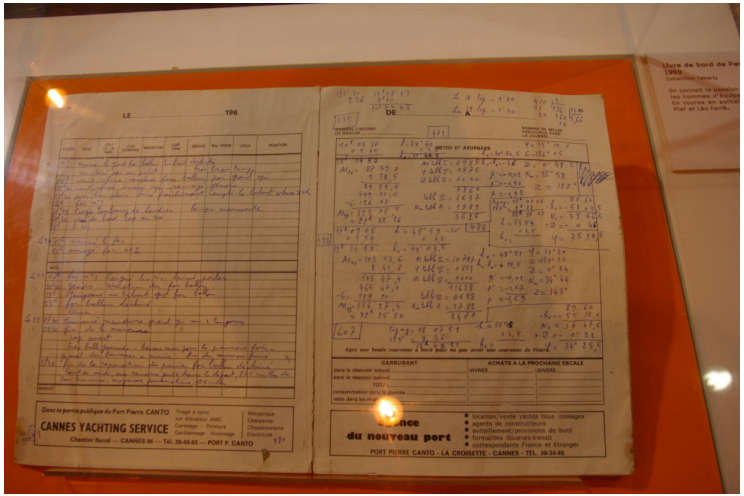
Notes The text can be found *legally* at several places, the Project Gutenberg one is missing the "Preface", so don't use it, go to one of the versions available on Internet Archive: `https://tinyurl.com/MOOC-RR-penguin-island`. The importance of the preface in illustrated by the following two quotations:

> *One word more if you want your book to be well received, lose no opportunity for exalting the virtues on which society is based — attachment to wealth, pious sentiments, and especially resignation on the part of the poor, which latter is the very foundation of order. Proclaim, sir, that the origins of property — nobility and police — are treated in your history with the respect which these institutions deserve. Make it known that you admit the supernatural when it presents itself. On these conditions you will succeed in good society.*

And more importantly for our subject:

> *The idea occurred to me, in the month of June last year, to go and consult on the origins and progress of Penguin art, the lamented M. Fulgence Tapir, the learned author of the 'Universal Annals of Painting, Sculpture and Architecture'*
>
> *Having been shown into his study, I found seated before a roll-top desk, beneath a frightful mass of papers, an amazingly short-sighted little man whose eyelids blinked behind his gold-mounted spectacles.*
>
> *To make up for the defect of his eyes his long and mobile nose, endowed with an exquisite sense of touch, explored the sensible world. By means of this organ Fulgence Tapir put himself in contact with art and beauty. It is observed that in France, as a general rule, musical critics are deaf and art critics are blind. This allows them the collectedness necessary for æsthetic ideas. Do you imagine that with eyes capable of perceiving the*

A sailor's logbook

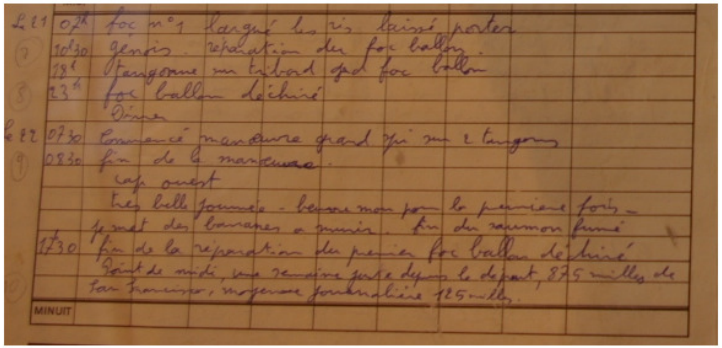


The logbook of Eric Tabarly during the San-Francisco / Tokyo transpacific ocean race in 1969.

A sailor's logbook

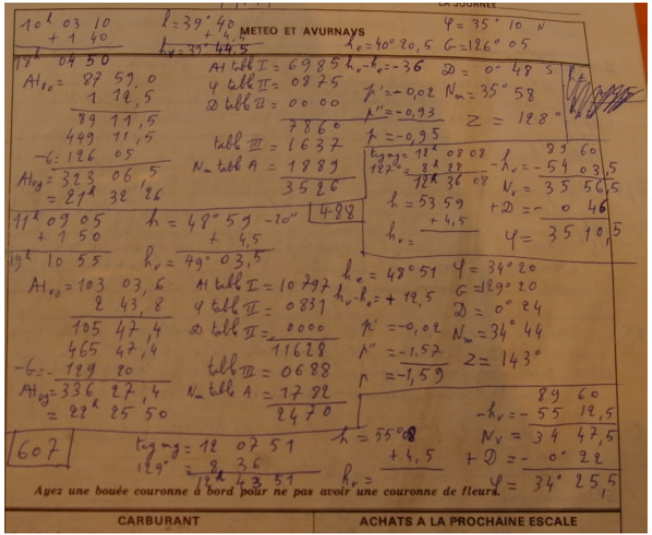
The logbook of Eric Tabarly during the San-Francisco / Tokyo transpacific ocean race in 1969.

Notes This example is only superficially anecdotal. Information about the source can be found at: `https://commons.wikimedia.org/wiki/File:LivredebordpenduickV.jpg`.

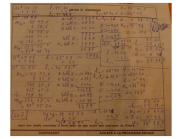


On the left side, Tabarly reports salient events like a ripped jib on March 21 at 11 pm.

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



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



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

Notes This example is only superficially anecdotal. Ten years ago, a European project was aiming at estimating the Atlantic and Indian Oceans climates during the 18th century using logbooks from ships of the West- and East-India companies from the Kingdoms of Portugal, Spain, Holland, Britain and France. See the Climatological Database for the World's Oceans 1750-1850.

In the same vein, logbooks from slave ships give a lot of quantitative information about the slave trade between Africa and the "New World".

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?
- . . . ?

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?

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?

Notes Notes are necessarily heterogeneous—because of their subject matter as well as, often, their material support—and that creates a serious organizational problem.

Without organization, notes usability barely exceeds our capability of memorizing facts and events.

In the sequel we are going to give *tentative* answers to the questions raised in the last two slides.

Where are we?

M1-S0: Lab books and note books

M1-S1: Note-taking concerns everyone

M1-S2: Note-taking: a quick history

M1-S3: From text files to lightweight markup languages

M1-S4: Notes (and codes) that are archived but can evolve with version control systems

M1-S5: Finding one's way with tags and desktop search application

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 to daily use.
- Hopefully, we can avoid thinking that we are the first to face the kind of problem we are now facing: "information overload".

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 to daily use.
- Hopefully, we can avoid thinking that we are the first to face the kind of problem we are now facing: "information overload".

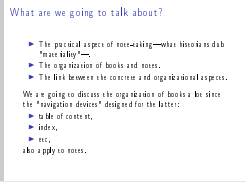
What are we going to talk about?

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

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

- table of content,
- index,
- etc,

also apply to notes.


What are we going to talk about?

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

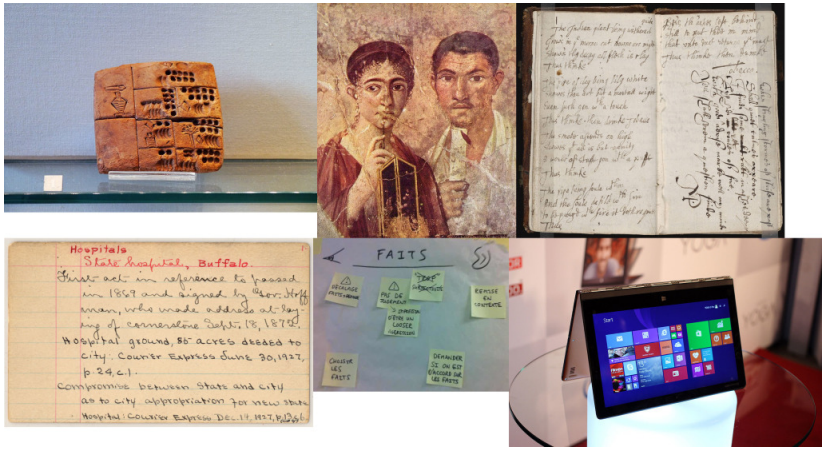
We are going to discuss the organization of books a lot since the "navigation devices" designed for the latter:

- table of content,
- index,
- etc,

also apply to notes.

Clarification We will mostly refer to the "Western" part of this History, with a single slide on Chinese contributions and nothing on Muslim, Indian or pre-Colombian contributions. This bias must be clearly understood as a reflection of my ignorance (I'm actively learning on the subject) and because it's easier, as always, to find illustrative material for "Western" contributions...

The concrete aspects summarized on a single slide



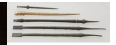
Details All illustrations are taken from Wikimedia Commons

- Top left: A clay tablet (pre-cuneiform period, -3000).
- Top center: A fresco from Pompeii with the portrait of Terentius Neo and his wife. She carries a wax tablet and a *stylus* (the main medium of note-takers up to the 19th century); he carries a *volumen* or scroll, the stuff of books until the beginning of the Common Era.
- Top right: a notebook made of paper from the 17th century with commonplaces. "Commonplace" is a translation of the Latin term locus communis (from Greek tópos koinós, see literary topos) which means "a theme or argument of general application", such as a statement of proverbial wisdom (Wikipedia).
- Bottom left: An index card, a notes medium whose use exploded with bureaucratization and the development of libraries. Still heavily used in the humanities. Apparently first used (if not created) by the father of taxonomy, Carl Linneaus. You can find his cards at: `http://linnean-online.org/61332/#/0`.
- Bottom center: A Post-it note as most of us use every day.

Wax tablet and stylus

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

Wax tablet and stylus



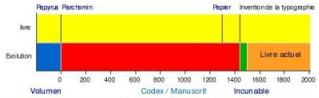
Musée romain-germanique
Cologne (Allemagne)
Photos de Jacques Poitou

Details From the Wikipedia page:

A wax tablet is a tablet made of wood and covered with a layer of wax, often linked loosely to a cover tablet, as a "double-leaved" diptych. It was used as a reusable and portable writing surface in Antiquity and throughout the Middle Ages.

Writing on the wax surface was performed with a pointed instrument, a stylus. Writing by engraving in wax required the application of much more pressure and traction than would be necessary with ink on parchment or papyrus,[1] and the scribe had to lift the stylus in order to change the direction of the stroke. Therefore, the stylus could not be applied with the same degree of dexterity as a pen. A straight-edged, spatula-like implement (often placed on the opposite end of the stylus tip) would be used in a razor-like fashion to serve as an eraser. The entire tablet could be erased for reuse by warming it to about 50 °C and smoothing the softened wax surface. The modern expression of "a clean slate" equates to the Latin expression "tabula rasa".

From the *scroll* to the *codex*



Details

The shift from the *scroll* to the *codex* is fundamental for development of written civilization.

A scroll (from the Old French escroe or escroue), is a roll of papyrus, parchment, or paper containing writing.

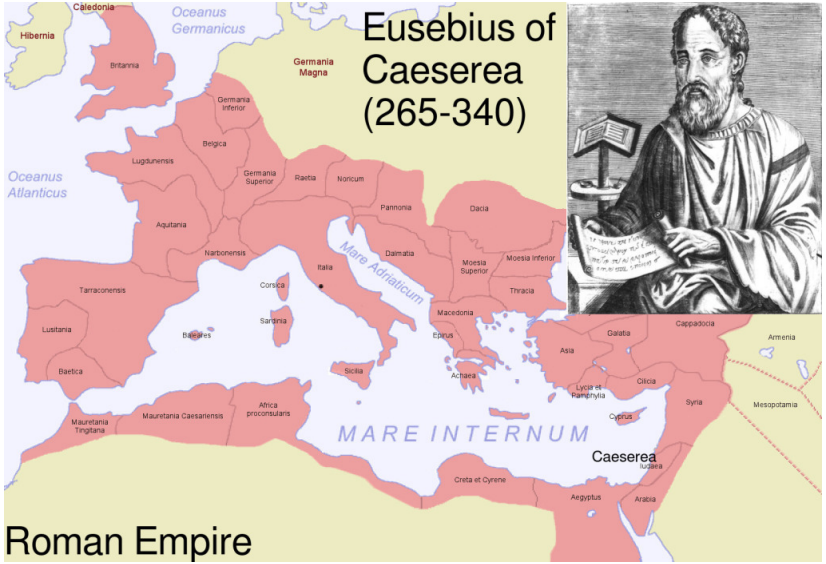
From Wikipedia:

The codex was a new format for reading the written word, consisting of individual pages loosely attached to each other at one side and bound with boards or cloth. It came to replace the scroll thanks to several problems that limited the scroll's function and readability. For one, scrolls were very long, sometimes as long as ten meters. This made them hard to hold open and read, a difficulty not helped by the fact that most scrolls in that era were read horizontally, instead of vertically as scrolling virtual documents are read now. The text on a scroll was continuous, without page breaks, which made indexing and bookmarking impossible. Conversely, the codex was easier to hold open, separate pages made it possible to index sections and mark a page, and the protective covers kept the fragile pages intact better than scrolls generally stayed. This last made it particularly attractive for important religious texts.

The bottom left mosaic shows Virgil seating (70-19 BCE) holding a scroll of the *Aeneid*, with Clio, muse of history, also holding a scroll.

As explained by Frédéric Barbier (*Histoire du Livre*): "The scroll / volumen imposes a complex reading practice: one must unroll (*explicare*) and roll at the same time;

Eusebius and the invention of cross-references



Details Illustrations from Wikimedia Commons.
From the Wikipedia page on Eusebius:

> *Eusebius of Caesarea (ad 260/265 – 339/340), also known as Eusebius Pamphili, was a historian of Christianity, exegete, and Christian polemicist. He became the bishop of Caesarea Maritima about 314 AD. Together with Pamphilus, he was a scholar of the Biblical canon and is regarded as an extremely learned Christian of his time. He wrote Demonstrations of the Gospel, Preparations for the Gospel, and On Discrepancies between the Gospels, studies of the Biblical text.*

According to Anthony Grafton and Megan Williams (2006) *Christianity and the Transformation of the Book*, The Belknap Press of Harvard University Press, his writings are crucial for our knowledge of the first three centuries of Christian history. *He brought several essential innovations to the book's organization like the cross-references.*

Eusebian canons

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

Eusebian canons

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

Details Source: `https://commons.wikimedia.org/wiki/File:Fol._10v-11r_Egmond_Gospels.jpg`. Public Domain.
Quote from Wikipedia:

> *For an easier survey of the material of the four Evangelists, Eusebius divided his edition of the New Testament into paragraphs and provided it with a synoptical table so that it might be easier to find the pericopes that belong together. These canon tables or "Eusebian canons" remained in use throughout the Middle Ages, and illuminated manuscript versions are important for the study of early medieval art, as they are the most elaborately decorated pages of many Gospel books.*

The significance of the *codex*

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

- The invention of the *codex* 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.

The significance of the *codex*

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

- The invention of the *codex* 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.

Details Example of *Carolingian minuscule* can be found on the corresponding Wikipedia page.
Over centuries, *codices*—that we often call *manuscripts*—will slowly evolve and gain modern days book attributes:

- separation between words (VIIth century),
- start of punctuation (VIIIth century),
- table of content,
- running title,
- paragraph marks (rubrication, XIth century),
- pagination,
- index (XIIIth century).

An interesting point: Torah's content got "fixed" before the *codex* generalization and today Torah scrolls are still used.

Let us not forget China

En haut : leishu Yongle Dadian
(1403) 370 millions de caractères.
À droite : matrice d'impression
de billets de banque
Dynastie Song du Nord (960-1127).

Details

The link between the *codex* generalization, on the one hand, and the apparition of "navigation guides" like the table of content, the index, the running title, on the other hand as a counterpart in the Chinese civilization.

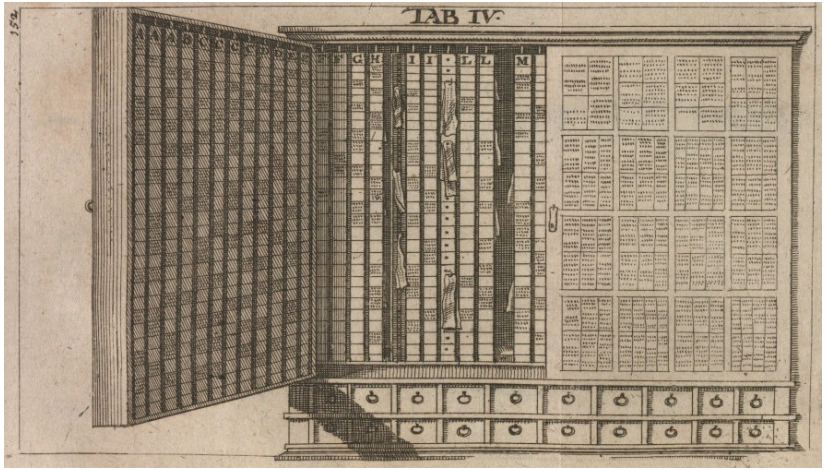
In China, competitive examinations to become a high ranking state employee developed in the IXth century (CE). The main part of these exam was a paper on what we would now call general knowledge of the Classics where the students were asked to demonstrate their knowledge through appropriate quotations.

To fulfill the need of "textbook" appropriate for this kind of examination what is called leishus were produced. They are described as follows on Wikipedia:

> *The leishu are composed of sometimes lengthy citations from other works and often contain copies of entire works, not just excerpts. The works are classified by a systematic set of categories, which are further divided into subcategories. Leishu may be considered anthologies, but are encyclopedic in the sense that they may comprise the entire realm of knowledge at the time of compilation.*

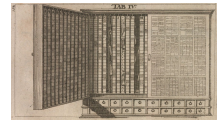
The efficient use of the leishu requires an indexing system, a table of content, etc. Very interestingly, the scroll will be abandoned and the codex will generalize in China around

Getting organized by using the right slot



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

Getting organized by using the right slot



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

Details

Now that we briefly reviewed the timeline of the main navigation elements of the books—navigation elements that can of course be applied to our lab/note-books—we come back to the paper slips and cards as notes media.
We see (again) Placcius' and Leibniz's closet since it displays both the benefits and the shortcomings of media that hold a single note.
Obvious shortcomings are:

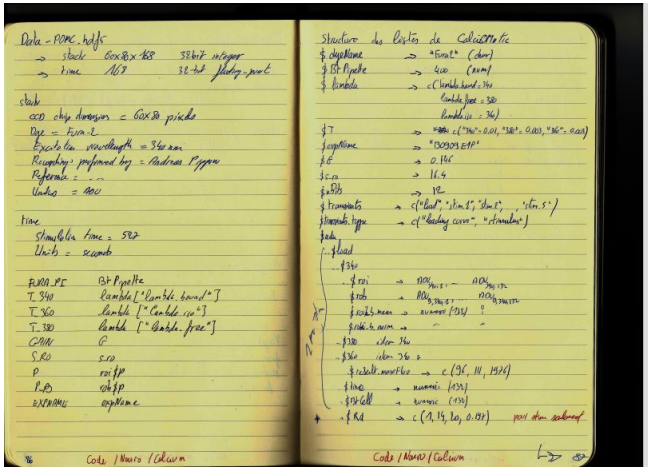
- Paper slips and cards get easily lost.
- They are essentially useless if they are not classified in addition to being filed.

These problems are solved by Placcius' cabinet, the content of which is fundamentally accessed through the index.
Clear benefits are:

- Paper slips can be easily reorganized when they contain information on several subjects.
- Paper slips can be directly pasted in a book when composing an anthology or a compendium.

This last technique (pasting when making an anthology) was systematically used by

Constructing a notebook index the John Locke way



My own notebook is used here for illustration.

Constructing a notebook index the John Locke way

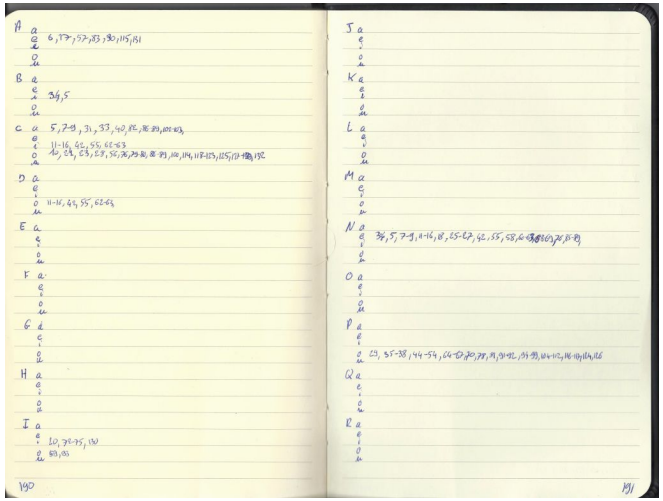
My own notebook is used here for illustration.

Details We will now learn about an index construction technique due to John Locke (1632-1704), the grand-father of liberalism and a major investor in the *Royal African Company*, the largest company in the slave-trade business at that time...
The indexing method is here illustrated using my own notebook. The two pages that are displayed describe the structure of a dataset in the HDF5 format on the left side and the corresponding structure (designed to map the former one) of a `data frame` object of the R language. This dataset contain calcium concentration measurements made in neurons. This notes were taken while writing some computer code to analyze the data. The precise content of the pages does not matter here in order to understand how Locke's method works. The important points are:

- The pages are numbered (we are seeing here pages 86 and 87).
- Keywords are written at the bottom of the page: code; neuro; calcium.

This method can be applied after note-taking, you just need to have few pages left at the end of your notebook. That's in fact what I did since I had started filling my notebook before learning about the method (I learned about while preparing the French version of this lecture last September).

Locke's method continued



The last pages of my notebook with the index.

Locke's method continued

The last pages of my notebook with the index.

Details

We know the index. It is located at the end of the notebook although Locke recommends placing it at the beginning. Since I did not know about the method when I started the notebook, I had to place it at the end...

The idea is to enter the keywords used in the notebook based on their first letter and the first vowel following the first letter.

The index is therefore made of the 26 letters (you see letters "A" to "R" here, the remaining ones are on the next page) subdivided the five most common vowels ("y" goes together with "i" in that case).

Pages 86 and 87 contained the keyword code that goes into the entry "Co" of the index (you see "86-89" because the following pages also concern code for the same project).

The keyword Neuro giving an entry on line "Ne", while the keyword Calcium gives an entry on line "Ca".

The keyword Criquet (not shown above) gives an entry on line "Ci".

It is also a good idea to list the set of keywords used in the notebook on the page preceding or following 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 digital media as much as possible (while keeping in mind what we just learned) since they provide:

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

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 digital media as much as possible (while keeping in mind what we just learned) since they provide:

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

Where are we?

M1-S0: Lab books and note books

M1-S1: Note-taking concerns everyone

M1-S2: Note-taking: a quick history

M1-S3: From text files to lightweight markup languages

M1-S4: Notes (and codes) that are archived but can evolve with version control systems

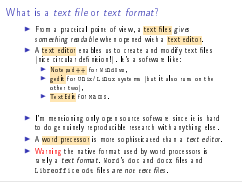
M1-S5: Finding one's way with tags and desktop search application

Section introduction We now start the "technical" part of this lecture with the tools that computers provide for note-taking like text files and lightweight markup languages.

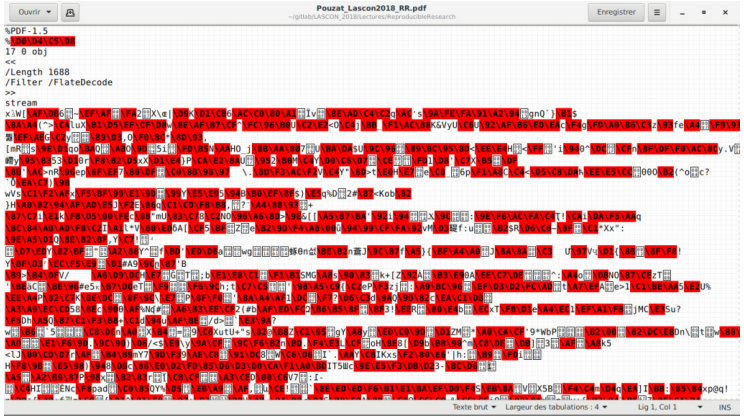
What is a *text file* or *text format*?

- From a practical point of view, 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 precessor is more sophisticated than a *text editor*.
- Warning the native format used by word processors 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

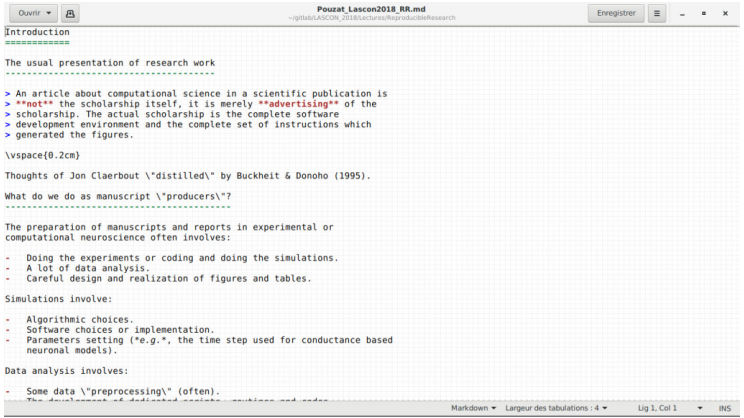


A `pdf` file (the file shown right now with a pdf reader) opened with `gedit`.

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



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

A text file opened with a text editor

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

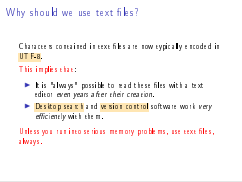
Why should we use text files?

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 version control software work *very efficiently* with them.

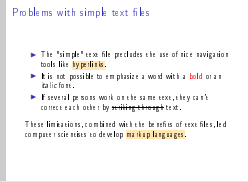
Unless you run into serious memory problems, use text files, always.



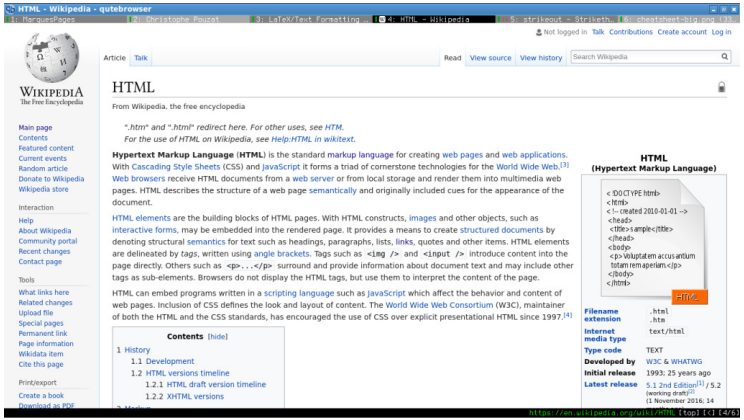
Problems with simple text files

- 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 with the benefits of text files, led computer scientists to develop markup languages.



A trivial example is the HTML language.

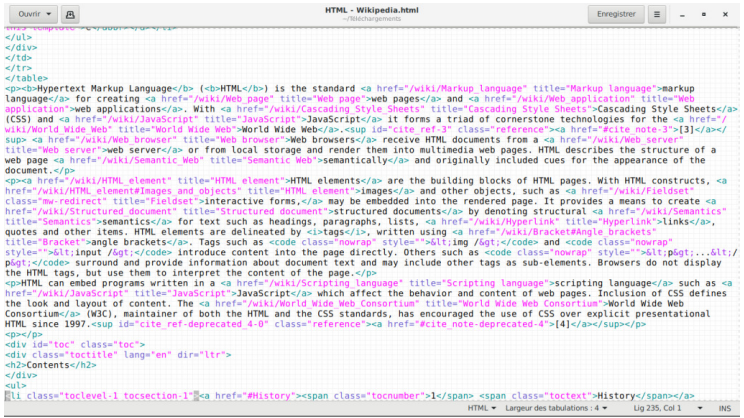


Wikipedia HTML page viewed with qutebrowser web browser.

A trivial example is the HTML language.

Wikipedia HTML page viewed with qutebrowser web browser.

An HTML file opened with a text editor



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

Note The content of files written with a markup language are typically processed by a dedicated software like a web browser or converted into a format for which readers are available like LaTeX files that get "compiled" into PDF files.

If you look carfuly the last figure, you can find the text of the first main paragraph of the previous figure.

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 usually require dedicated editing software if we want to modify them.

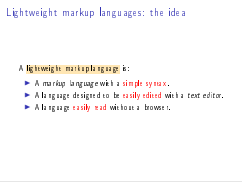
Is it possible to combine the benefits of "simple" text files with the reading comfort of markup languages?

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 usually require 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?

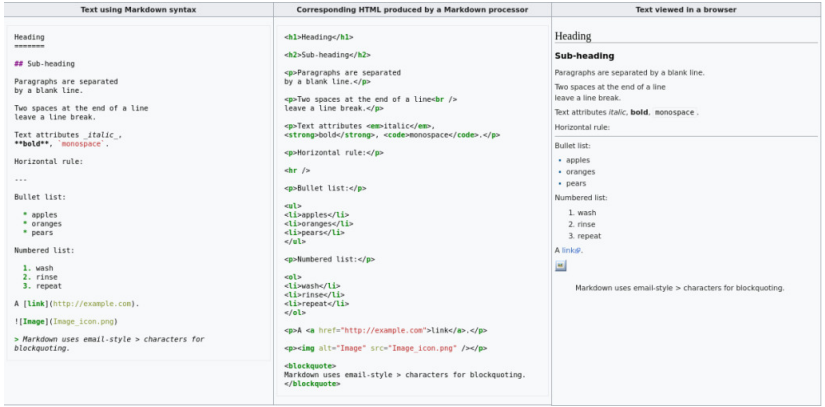
Lightweight markup languages: the idea

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.

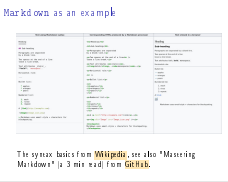


Markdown as an example



Text using Markdown syntax	Corresponding HTML produced by a Markdown processor	Text viewed in a browser
`Heading` `=======` `## Sub-heading` `Paragraphs are separated` `by a blank line.` `Two spaces at the end of a line` `leave a line break.` `Text attributes _italic_,` `**bold**, `monospace`.` `Horizontal rule:` `---` `Bullet list:` `  * apples` `  * oranges` `  * pears` `Numbered list:` `  1. wash` `  2. rinse` `  3. repeat` `A [link](http://example.com).` `![Image](Image_icon.png)` `> Markdown uses email-style > characters for` `blockquoting.`	`<h1>Heading</h1>` `<h2>Sub-heading</h2>` `<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 <em>italic</em>,` `<strong>bold</strong>, <code>monospace</code>.</p>` `<p>Horizontal rule:</p>` `<hr />` `<p>Bullet list:</p>` `<ul>` `<li>apples</li>` `<li>oranges</li>` `<li>pears</li>` `</ul>` `<p>Numbered list:</p>` `<ol>` `<li>wash</li>` `<li>rinse</li>` `<li>repeat</li>` `</ol>` `<p>A <a href="http://example.com">link</a>.</p>` `<p><img alt="Image" src="Image_icon.png" /></p>` `<blockquote>` `Markdown uses email-style > characters for blockquoting.` `</blockquote>`	**Heading** **Sub-heading** Paragraphs are separated by a blank line. Two spaces at the end of a line leave a line break. Text attributes *italic*, **bold**, `monospace`. Horizontal rule: Bullet list: • apples • oranges • pears Numbered list: 1. wash 2. rinse 3. repeat A link#. Markdown uses email-style > characters for blockquoting.

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

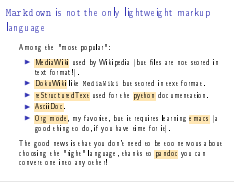


Markdown is not the only lightweight markup language

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).

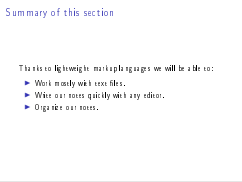
The good news is that you don't need to be too nervous about choosing the "right" language, thanks to pandoc you can convert one into any other!



Summary of this section

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?

M1-S0: Lab books and note books

M1-S1: Note-taking concerns everyone

M1-S2: Note-taking: a quick history

M1-S3: From text files to lightweight markup languages

M1-S4: Notes (and codes) that are archived but can evolve with version control systems

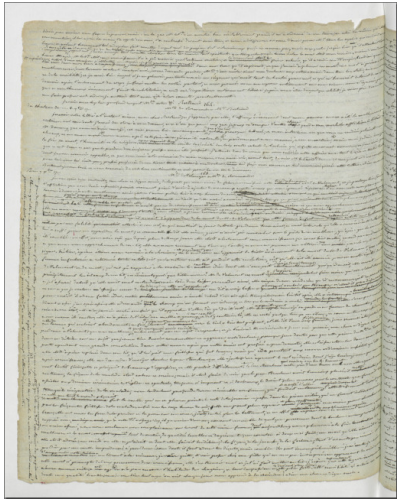
M1-S5: Finding one's way with tags and desktop search application

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 with text is concerned, even more so when this work is done in collaboration.
- The longevity issue of notes and texts is in no way 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 has demonstrated its capability to last.
- When it comes to making notes evolve, I think we can say that some real progress was recently made.

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 with text is concerned, even more so when this work is done in collaboration.
- The longevity issue of notes and texts is in no way 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 has demonstrated its capability to last.
- When it comes to making notes evolve, I think we can say that some real progress was recently made.

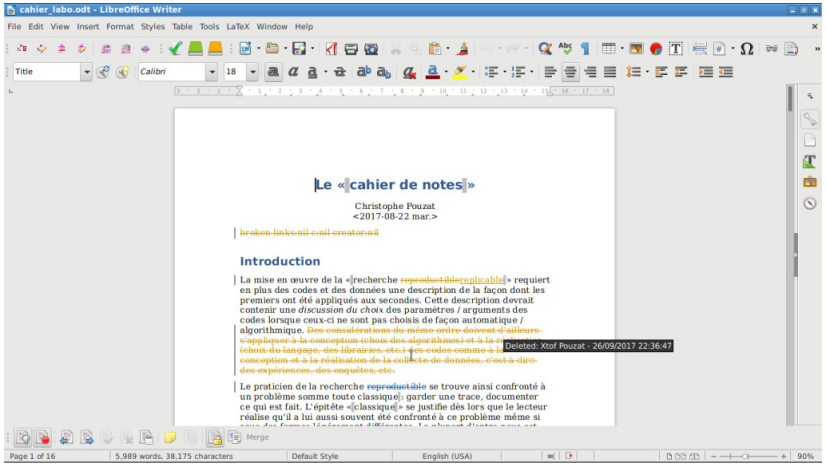
The nightmare: changing a text on paper medium



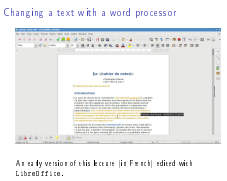
Manuscript of *Dangerous Liaisons* (*Les liaisons dangereuses*) by Pierre Choderlos de Laclos (p. 258, BNF Gallica).
There is clearly a very limited number of changes one can bring in that way!

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`.

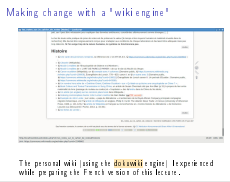


Details We see a way of working in collaboration on a text: most word processing software have a way to follow changes brought to the text.
This is not the solution I recommend but this is probably the most widely known concurrent version facility.
Notice the buttons at the bottom left. They appear when you navigate in *view* -> *Toolbars* -> *track changes*.
This "solution":

- is easy to implement,
- *does not generate text files*
- does not take care of archiving the files.

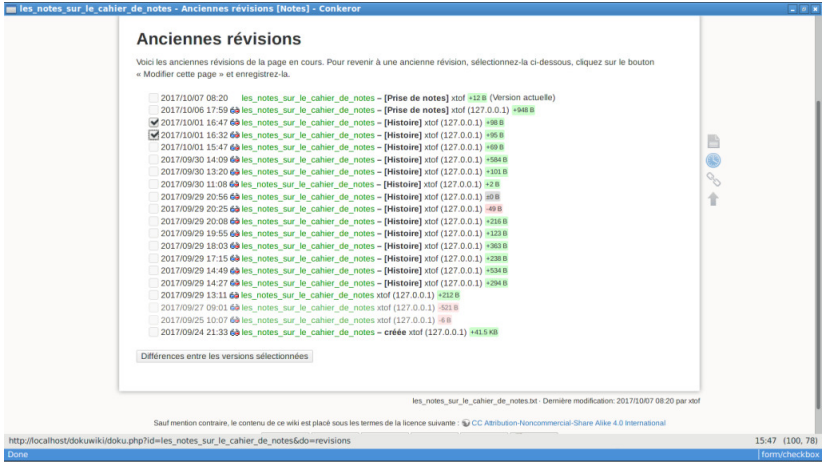
Making change with a "wiki engine"

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

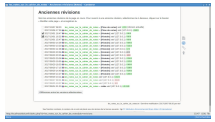


Details I started using dokuwiki for this lecture, it is therefore simple enough to learn.

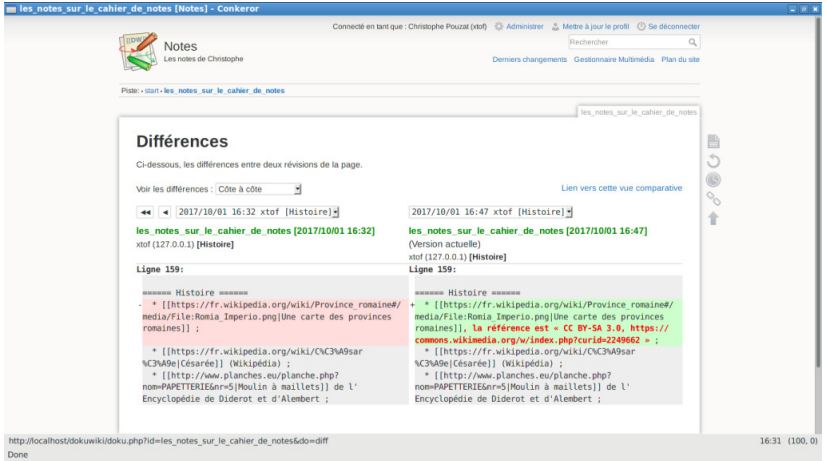
Dokuwiki uses a test format.



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



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



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.

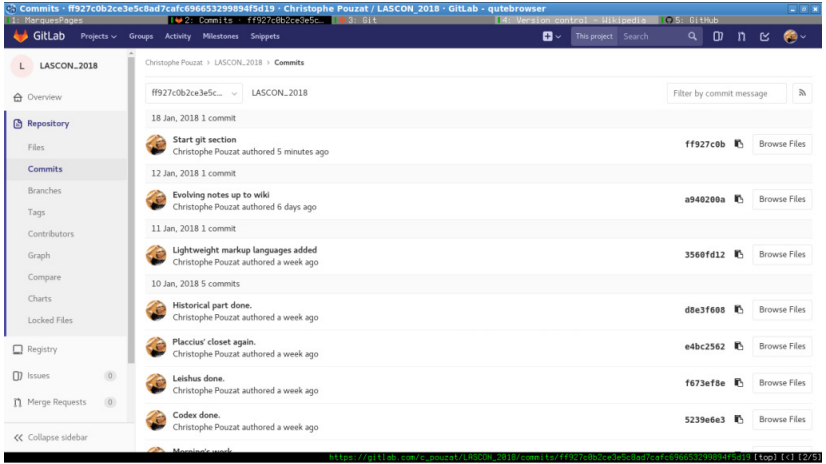
Pros and cons
- A solution with a strong record for collaborative projects (Wikipedia).
- A text format is used when working with Dokuwiki.
- A single page can be modified at a time.

Version Control Systems

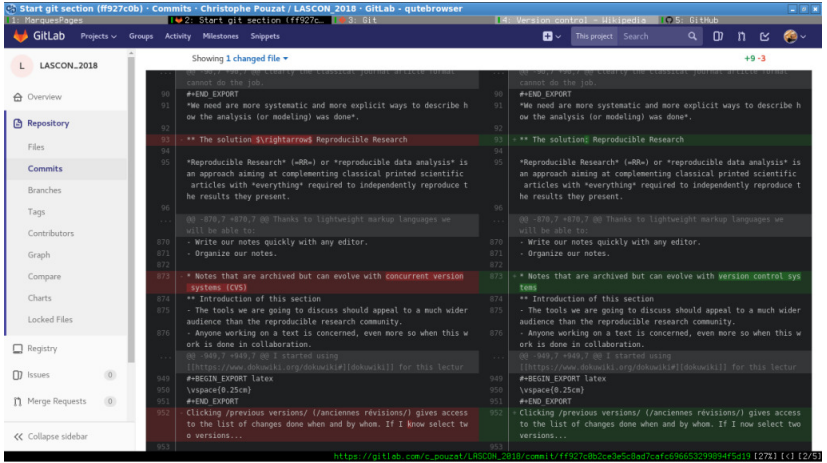
I now come to the most "sophisticated" solution:

- A dedicated software, git, is used to manage the successive versions of a set of files in different formats (text, images, etc.). In fact, file arborizations can be managed.
- `git`-like software requires a repository, that can be built on the user's computer, but 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*).

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, file arborizations can be managed.
git-like software requires a repository, that can be built on the user's computer, but 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).

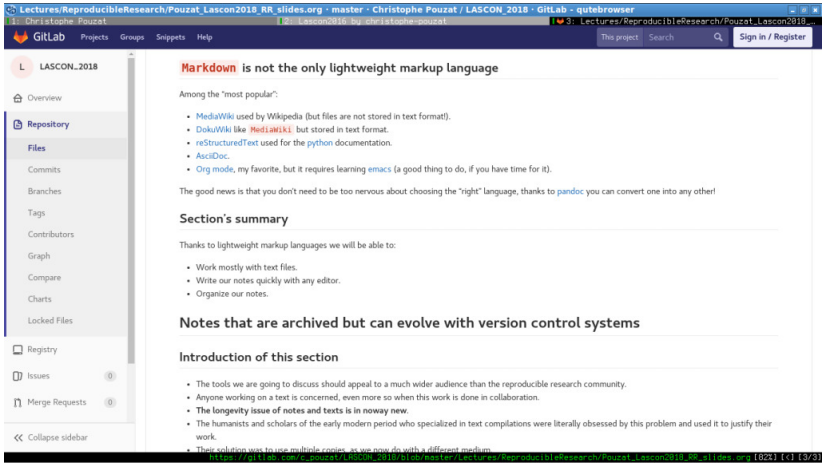


The GitLab interface containing the files of this presentation.



Modifications are easily visualized...

Modifications are easily visualized...



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


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.

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.

Where are we?

M1-S0: Lab books and note books

M1-S1: Note-taking concerns everyone

M1-S2: Note-taking: a quick history

M1-S3: From text files to lightweight markup languages

M1-S4: Notes (and codes) that are archived but can evolve with version control systems

M1-S5: Finding one's way with tags and desktop search application

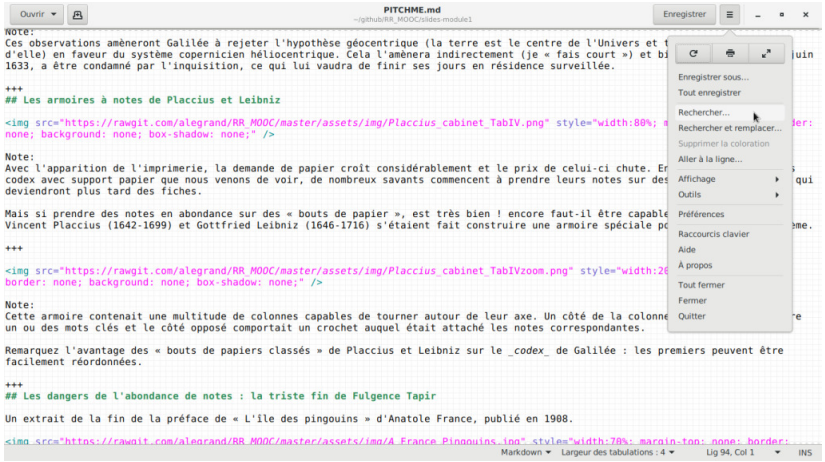
Leibniz again

"It seems to me that the apparatus of contemporary scholarship is comparable to a very large store which, though it keeps a great variety of goods, yet is totally confused and in disorder, because all items are mixed up, because no numbers or letters of an index are displayed, and because inventories or account ledgers which could throw some light on the matter are missing."
"The larger the mass of collected things, the less will be their usefulness. Therefore, one should not only strive to assemble new goods from everywhere, but one must endeavor to put in the right order those that one already possesses."

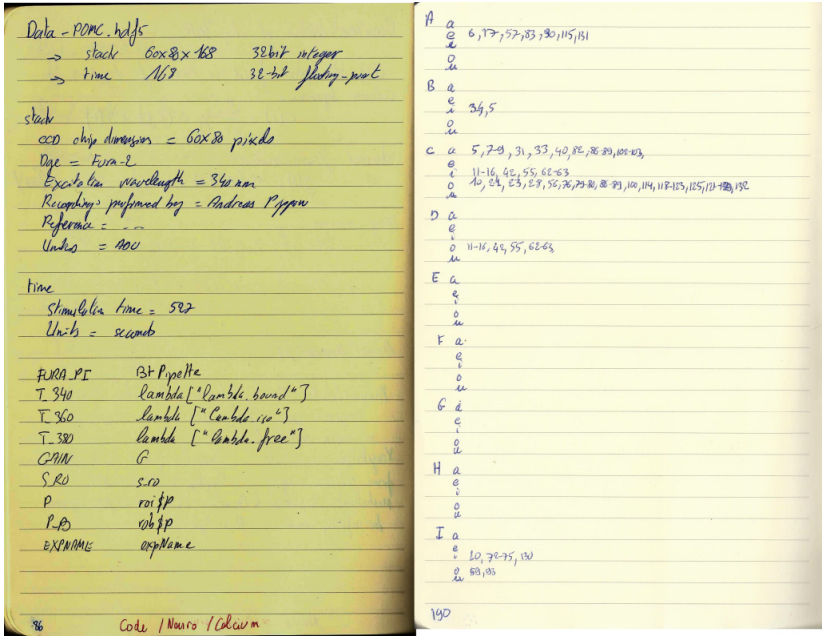
Leibniz again

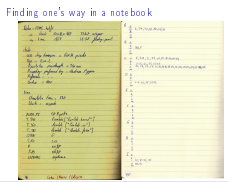
"It seems to me that the apparatus of contemporary scholarship is comparable to a very large store which, though it keeps a great variety of goods, yet is totally confused and in disorder, because all items are mixed up, because no numbers or letters of an index are displayed, and because inventories or account ledgers which could throw some light on the matter are missing."
"The larger the mass of collected things, the less will be their usefulness. Therefore, one should not only strive to assemble new goods from everywhere, but one must endeavor to put in the right order those that one already possesses."

Finding one's way in a text file

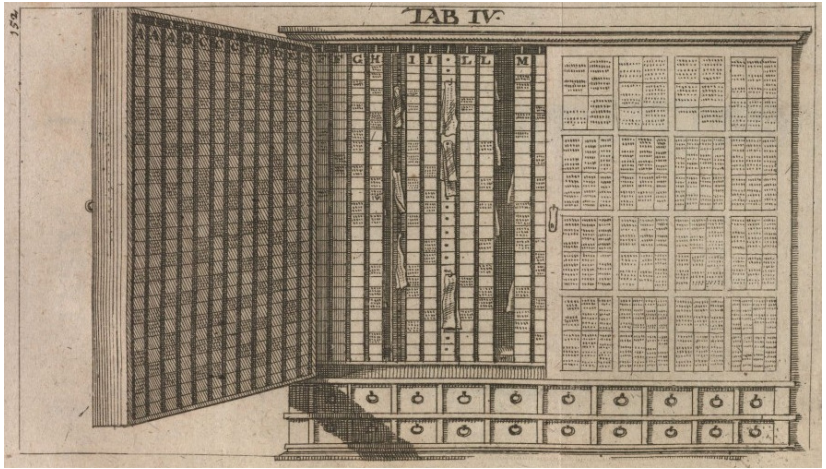


Finding one's way in a notebook

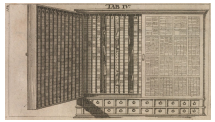




Finding one's way in a cards collection



Finding one's way in a cards collection

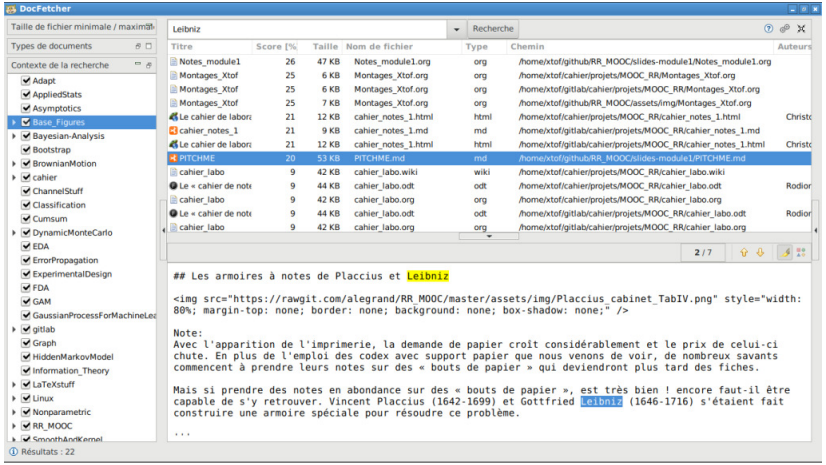


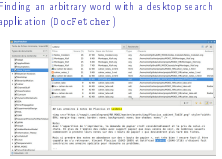
Problems, limitations, solutions?

- A single document at a time
- Numerical files indexation
- Tagging numerical files in general (not only text format files)
- Using a desktop search application for indexation and general search

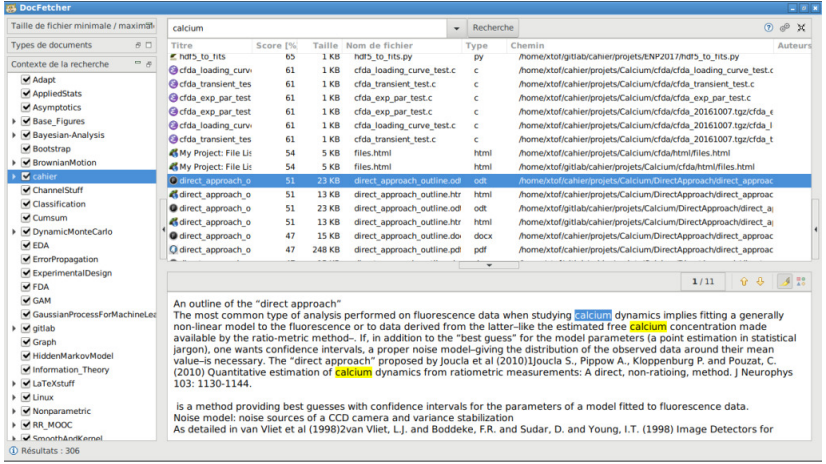
Problems, limitations, solutions?
- A single document at a time
- Numerical files indexation
- Tagging numerical files in general (not only text format files)
- Using a desktop search application for indexation and general search

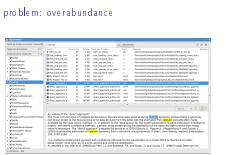
Finding an arbitrary word with a desktop search application (DocFetcher)





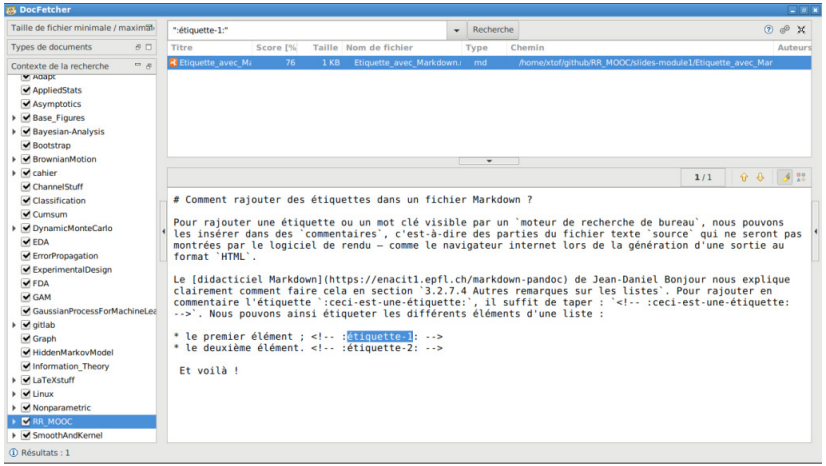
A problem: overabundance





Adding tags / keywords in a text file (Markdown)

Finding a tag with a desktop search application (DocFetcher)



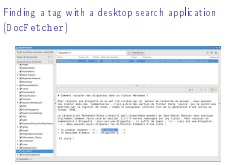


Image files contain metadata

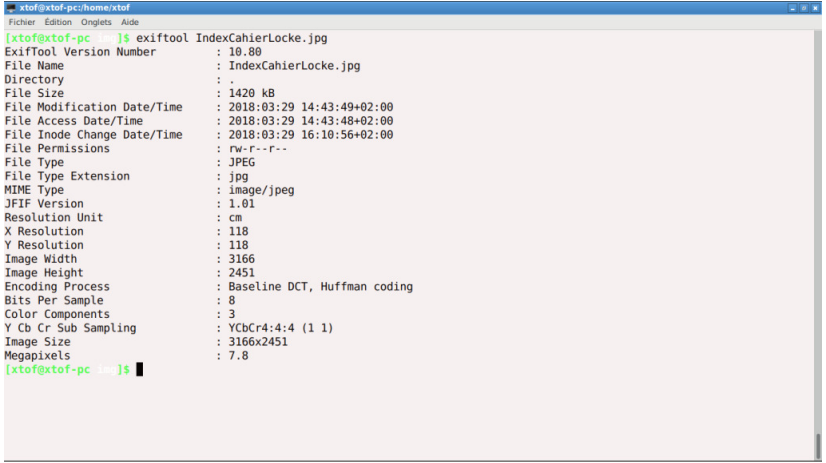
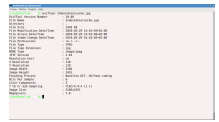


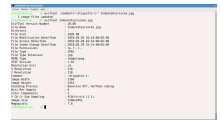
Image files contain metadata



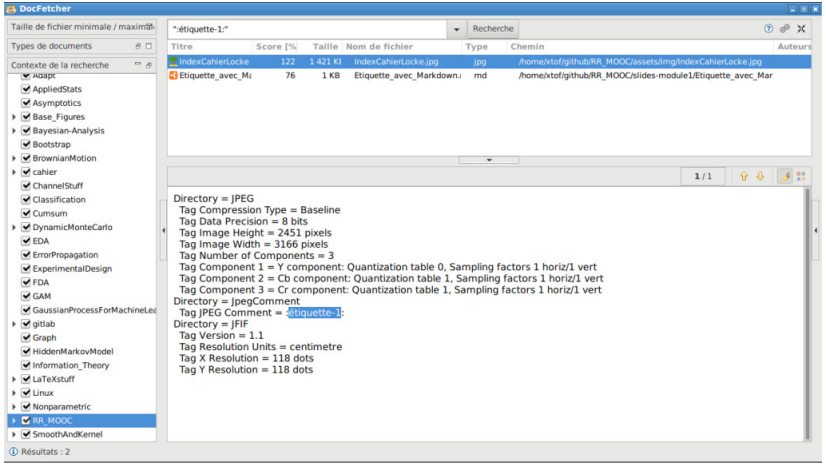
Metadata can be set

```
xtof@xtof-pc:/home/xtof
Fichier  Édition  Onglets  Aide
[xtof@xtof-pc ~]$ exiftool -comment=":étiquette-1:" IndexCahierLocke.jpg
    1 image files updated
[xtof@xtof-pc ~]$ exiftool IndexCahierLocke.jpg
ExifTool Version Number         : 10.80
File Name                       : IndexCahierLocke.jpg
Directory                       : .
File Size                       : 1420 kB
File Modification Date/Time     : 2018:03:29 16:14:06+02:00
File Access Date/Time           : 2018:03:29 16:14:06+02:00
File Inode Change Date/Time     : 2018:03:29 16:14:06+02:00
File Permissions                : rw-r--r--
File Type                       : JPEG
File Type Extension             : jpg
MIME Type                       : image/jpeg
JFIF Version                    : 1.01
Resolution Unit                 : cm
X Resolution                    : 118
Y Resolution                    : 118
Comment                         : :étiquette-1:
Image Width                     : 3166
Image Height                    : 2451
Encoding Process                : Baseline DCT, Huffman coding
Bits Per Sample                 : 8
Color Components                : 3
Y Cb Cr Sub Sampling            : YCbCr4:4:4 (1 1)
Image Size                      : 3166x2451
Megapixels                      : 7.8
[xtof@xtof-pc ~]$
```

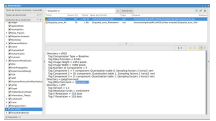

Metadata can be set



Desktop search applications can read metadata



Desktop search applications can read metadata



Conclusions

Using:

- tags / keywords inserted in our numerical files (text, images, PDF, etc.)
- a desktop search application

we can (perhaps) avoid "Leibniz's nightmare".

Conclusions

Using:

- tags / keywords inserted in our numerical files (text, images, PDF, etc.)
- a desktop search application

we can (perhaps) avoid "Leibniz's nightmare".