Lessons from the partnership with the Bibliothèque nationale de France

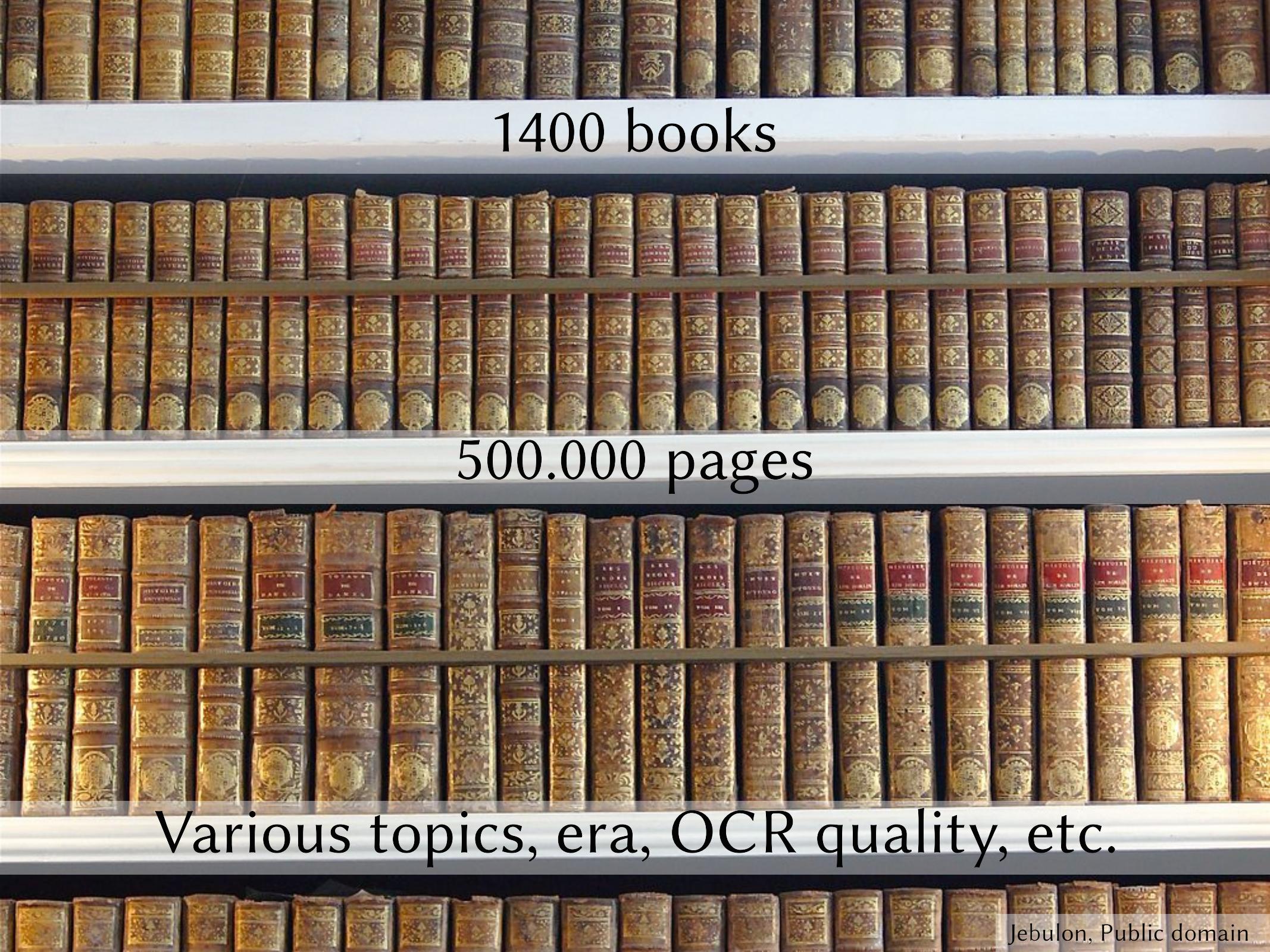# Context

Bibliothèque nationale de France

Negociations

2009

Partnership agreement

2010

Technical work

Show must go on…

1400 books

500.000 pages

Various topics, era, OCR quality, etc.

Technique

Inside...

TIFF multipage

XML "RefNum"
Book metadata
Pages metadata

XML "ALTO"
OCR

HTML ToC

PNG

BnF

TIFF

AltoProd ... AltoProd

RefNum

Tiffsplit

TIFF ... TIFF

Crop Info

TXT ... TXT

TXT Filedesc Commons

TXT Index WS

Wikisource

Annotate

Crop

cjb2

DjVu ... DjVu

DjVu

TAR

Server-side import
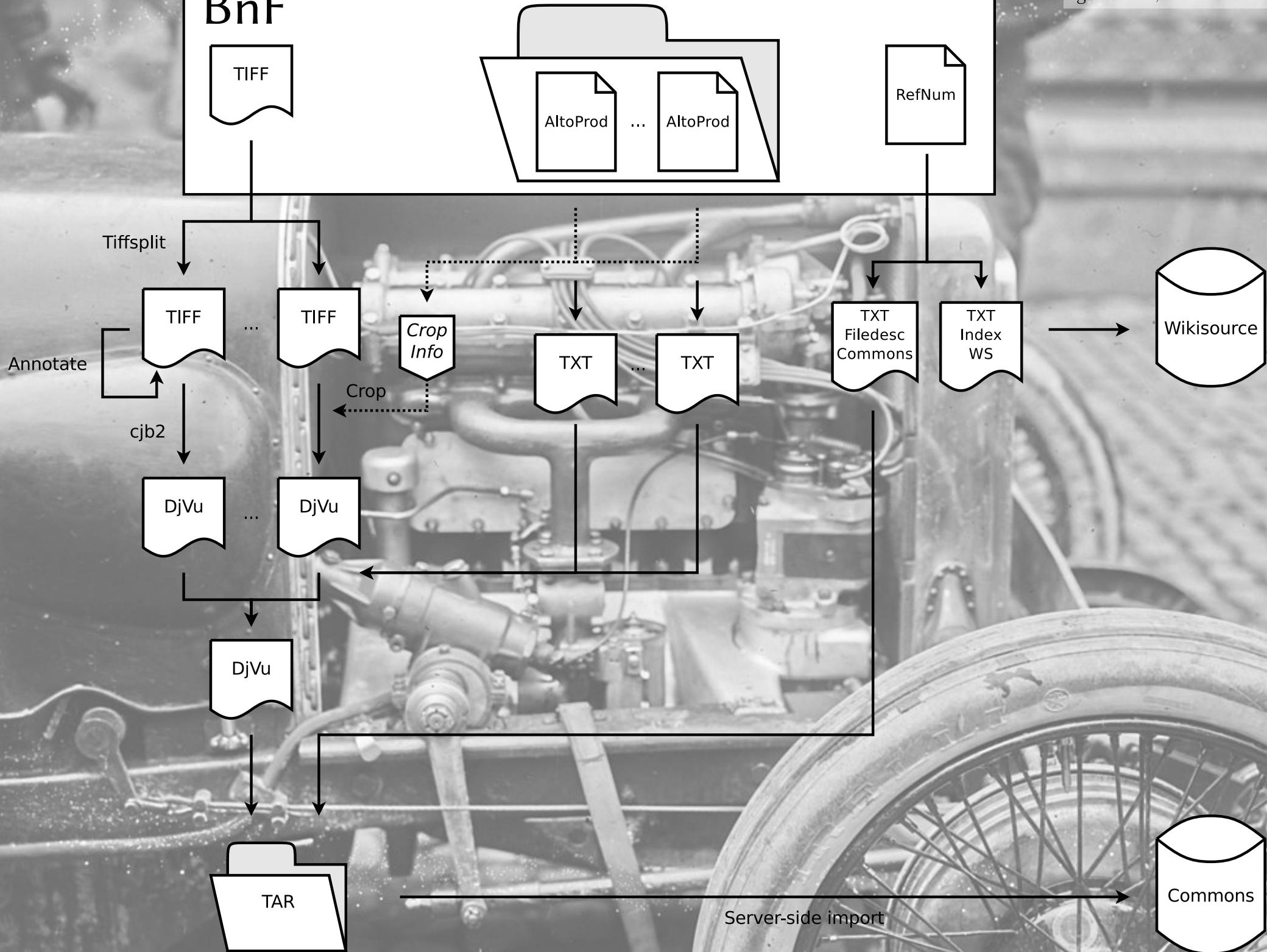
Commons

Text layer addition

Cropping/trimming pages

Coordinate conversion

$$Xmin = HPOS - XminPage$$
$$Xmax = Xmin + WIDTH$$
$$Ymax = tailleY + YminPage - VPOS$$
$$Ymine = Ymax - HEIGHT$$

Unicode #$@!

Image extraction (!)

Pagination extraction (!!)

Cutwords (!!!)

XminPage

HPOS

WIDTH

O1

YminPage

VPOS

HEIGHT

tailleY

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Ymax

Ymin

O2

Xmin

Xmax

Masque de découpage

Page ALTO

Lessons learnt

Wikisorcerers...

...better than machines?

A selection...

...made by
the BnF

The quick brown fox jumps over the lazy dog. Typography is the art and technique of arranging type to make written language legible, readable and appealing when displayed.

So many books...

...for so few hands

Conviviality
productivity

François Louis Jaques, public domain

"Help will always be given at Wikimedia…

…to those who ask for it."

Very focused…

…on ourselves.

WIKIMEDIA

FRANCE

[[meta:BnF]]