

ELTeC-nor

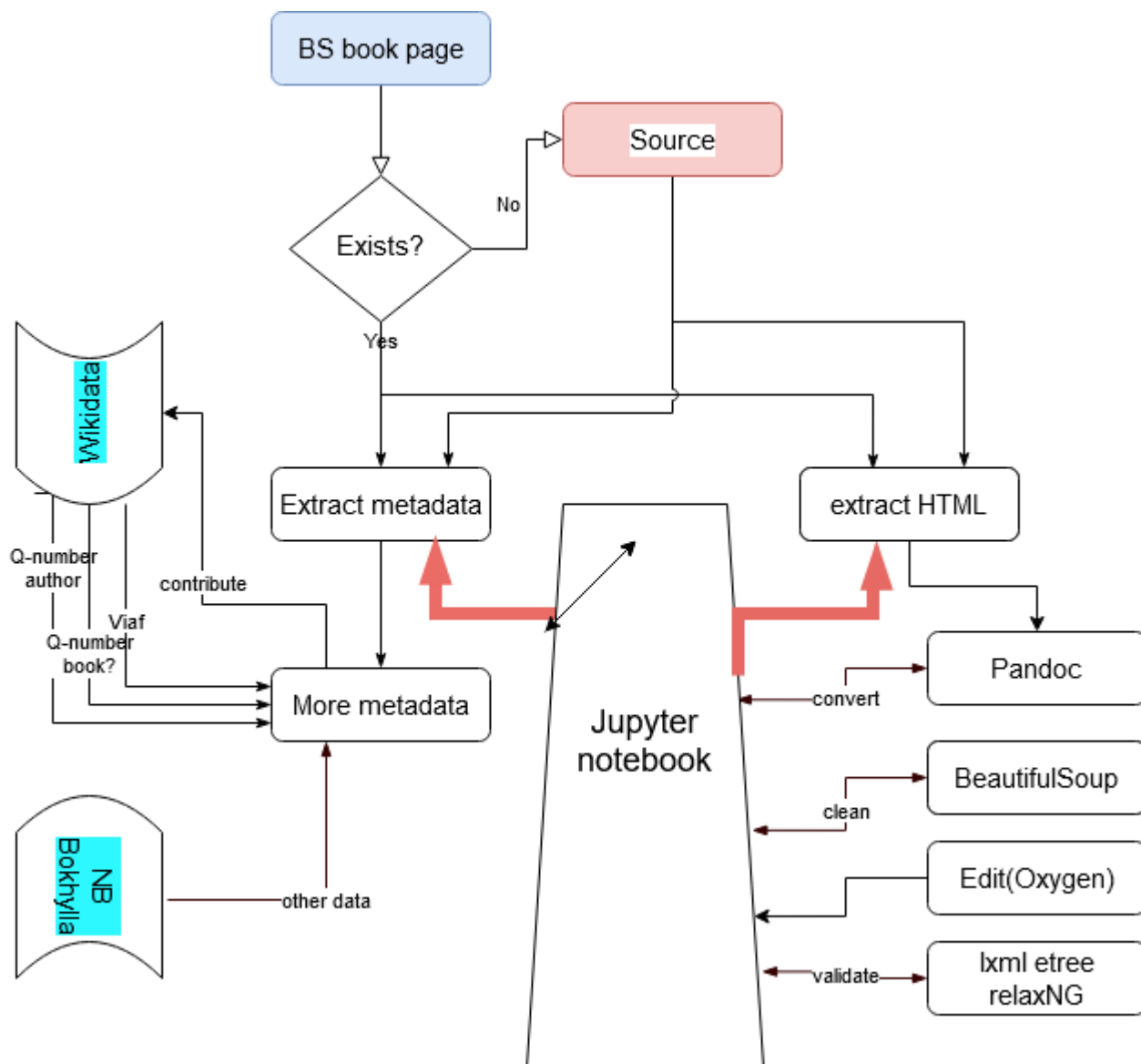
- Michael Preminger
- Christian-Emil Ore

Where did we get the novels from?

- Bokselskapet - a portal developed by NSL that publishes old Norwegian literature as e-books. (TEI-coded)
- The national library (*)

Software used

- Oxygen
- Jupyter Notebook (Python)
 - Pandoc (where necessary)
 - Beautiful soup (for simple transformations)



What we are having right now.

- 50 selected books (still some issues).
 - Had to try and balance the parameters on texts that have decent OCR quality
 - showed not to have a vey good OCR after all.
- 45 Bokmål, 5 nynorsk
- more on the production line. (but for now, not as many as 100).
 - Too many books with challenging OCR quality. (See below)

		AUTHORSHIP					LENGTH		
Last update	Texts	Words	Male	Female	1-title	3-title	Short	Medium	Long
2020-11-17	50	3195845	36	14	20	8	25	17	8

TIME SLOT					REPRINT COUNT		
1840-59	1860-79	1880-99	1900-20	range	Frequent	Rare	E5C
5	2	28	15	26	30	20	71.54

What should we have done differently from the outset?

- Tighter project organization
 - Tighter and more binding cooperation with the national library.
 - Integration in other research and teaching.
 - Participated in Budapest ...

Plans ahead

- (*)The national library is endeavoring on a re-scanning project.
 - ELTeC candidate books will get priority.
- Have some vacant time this spring.
- Hope for some financing for student hours for assistance
 - particularly in the quality assurance step.