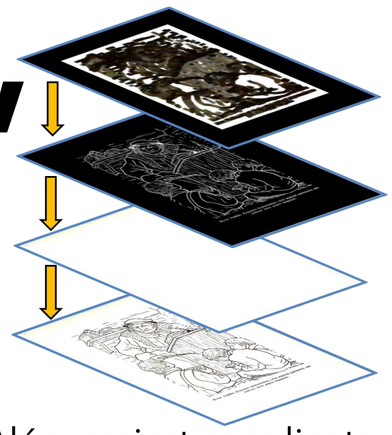
Monograph Digitization Workflow



The SKS-SHS digitization project

https://www.doria.fi/handle/10024/136372

Niklas Alén, project coordinator

Realized book digitization process

1. Charting material → 2. Choosing file format → 3. Choosing a digital repository → 4. Metadata production → 5. Rights management; in conjunction with 3-4

1. Charting the material

- The material was charted and saved in a relational database.
- Information merged from library catalogues, online resources and publication databases.

2. Choosing file format

- 2 PDF versions: Master and web.
- Master PDF: PDF/A-1b; Uncompressed TIFF; OCR-text-layer. Complies with Finnish national long-term preservation standards.
- Web PDF: MRC-PDF; OCR text-layer. Good compromise between image quality and compression ratio (e.g. 3,2 Mb TIFF vs. 64,4 Kb MRC).

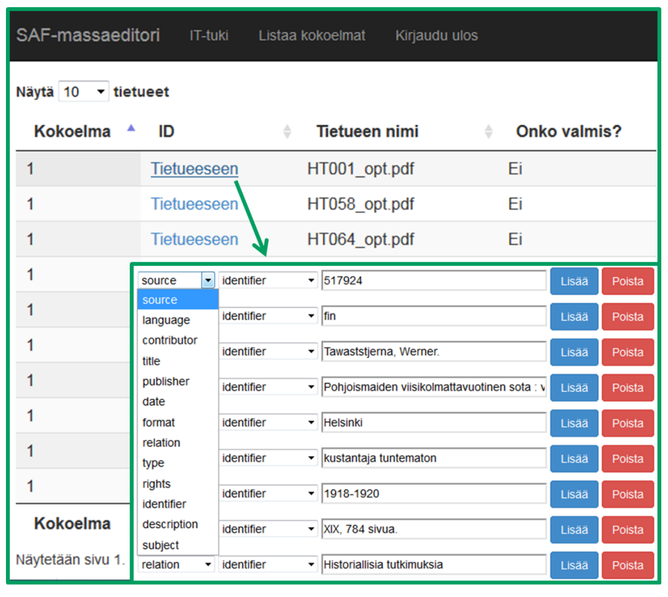
3. Choosing a digital repository

- Dspace instance provided by the Finnish National Library.
- Direct integration with URN-PID's administered and maintained by the Finnish National Library.
- Good SEO.
- Material also indexed in FINNA and BASE.

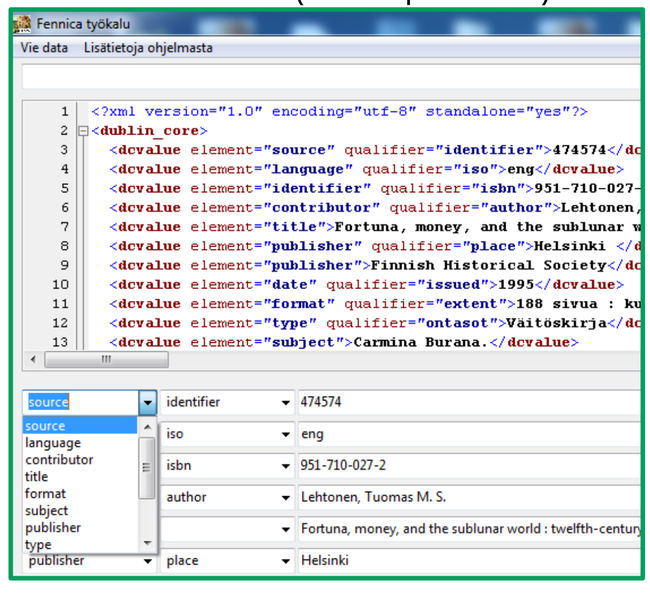
4. Metadata production

- Metadata harvested from the National Bibliography of Finland in MARCXML format.
- Metadata converted from MARCXMLto Simple Archive Format DC.
- Metadata supplemented in purpose built web app.
- Packaged in SAF-format and uploaded to Dspace.

metadata editor (web application):



metadata editor (desktop version):



5. Rights management

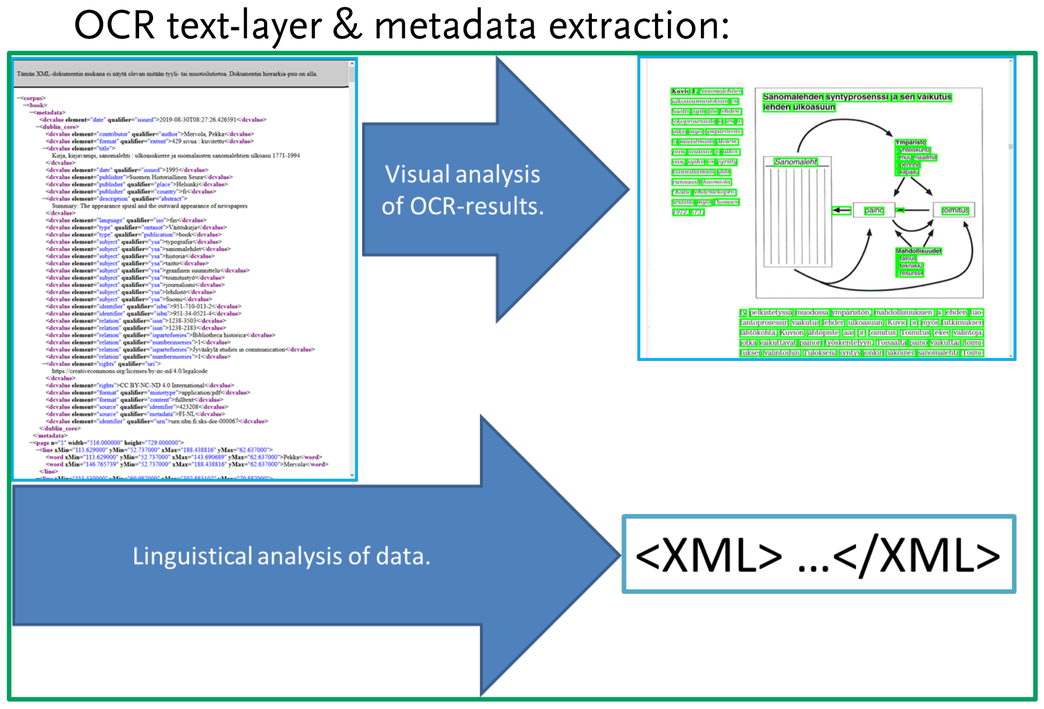
- Performed in conjunction with steps 3-4.
- Backed up by a relational database of authors.
- Proved to be far more time consuming than initially expected.
- Possible solutions: annual compensations to copyright societies, usage of EU orphan work legislation.

Suggested book digitization process

1. Rights management → 2. Charting material → 3. Choosing file format → 4. Choosing a digital repository → Metadata production

Use cases for scholars:

OCR text-layer & metadata extraction:



Image/chart/map recognition and extraction: