Guide to files in the Freedom of Press Writings

The collection Freedom of Press Writings (Danish: Trykkefrihedens Skrifter) gives an unusual deep insight into what happened during some turbulent three years in the 18th century society in Denmark. High and low expressed their opinions in an outburst of pamphlets and writings about many different issues between mainly 1770 and 1773. Most of this would have been lost had it not been for the diligent collector Bolle Willum Luxdorph:

> https://www.kb.dk/en/inspiration/freedom-press-writings

The collection was digitized with fundings from the Carlsberg Foundation, which also funded the important scientific work 'Grov Konfækt' about this collection published in September 2020 by authors Henrik Horstbøll, Ulrik Langen and Frederik Stjernfelt.

The digital collection was first presented on the Royal Danish Library's text portal on the 14th September 2020, the official date for celebrating 250 years of freedom of press in Denmark:

> https://tekster.kb.dk/tfs

In September 2021 the improved version 2.00 of the digital collection has been made available as files/datasets in the Royal Danish Library's Open Access repository: https://loar.kb.dk The entire collection is Public Domain and is therefore free of known restrictions under copyright law.

The datasets in the digital collection are:

- 28,643 facsimiles, TIFF format
- 28,643 ALTO-files, one for each facsimile
- Registry, in TEI and Microsoft Excel
- Scripts for extraction of plain text files from the ALTO files
- 28,643 plain text files with original lines
- 28,643 plain text files with words at line endings joined where there is hyphenation (following common practice in the 1770 decade)
- 1,097 facsimiles with noticeable illustrations. It is a subset of the facsimiles in a lower resolution, in JPEG format

All text files use the UTF-8 character set.

Except for plain text all text files use UNIX/Linux end-of-lines (a LF character). In the plain text files we have used Windows end-of-lines (CR + LF) since it seems preferable by most researchers.

The ALTO files were obtained through some iterations using mainly the ABBYY Fraktur OCR application since parallel OCR with Google Tesseract did not provide better results.

These OCR results, however, were littered with misidentifications since the nature of the physical collection poses a severe challenge to today's best OCR software:

> Poor paper quality, often poor print quality, unequal fraktur (black letter) fonts, a mix of fraktur and antiqua, big variations in spelling, unequal page layouts, multiple languages (primarily Danish, German and French but also Latin, English and Swedish and minor segments in Hebrew and Greek), unusual constructed proper nouns (especially the important pseudonyms of the then anonymous proponents like Philopatreias, Misokakias, Philodanus and many more).

Therefore, after the OCR process, a large manual proofreading process was performed on the ALTO files to correct most errors. However, the process was bound by financial and time restrictions.

It should be noted that some documents are handwritten, which gave nonsense (or empty) OCR results. In most cases these OCR results have not been corrected. Furthermore, approx. 1000 documents in German have not undergone proofreading.

Names of folders and files in the collection

Folder names
Example of a folder name:

> 2_017_130017168194

The folder name consists of three segments separated by underscore characters (_):

The **first segment**, e.g. 2, is the series number, which can be 1, 2, 3 or 4.

Mr. Luxdorph defined the series 1 and 2. Mr. Luxdorph later added a separate volume, which we have assigned the virtual series number 3 in the digitization process.

The contemporary editors Kanneworff and Gyldendal published an important overview/registry of the writings/pamplets, which is typically seen as part of the collection, and is therefore a part of the digital collection that we have assigned the series number 4.

The **second segment**, e.g. 017, is the volume number within a series.

The **third segment**, e.g. 130017168194, is the Royal Danish Library's "bar code number", which indicates the physical location of the volume on the library's premises.

File names
Example of file names:

130017168194_2_017_009_004.tif	Facsimile, TIFF format
130017168194_2_017_009_004.xml	ALTO file
130017168194_2_017_009_004.txt	Plain text file

Each file represents one page. Note that a single page appears in multiple files with the same name, except that the different file name extensions tell the type of contents.

Disregarding the extension, the file name consists of five segments separated by underscore characters (_):

The **first segment**, e.g. 130017168194, is the Royal Danish Library's "bar code number", which indicates the physical location of the volume on the library's premises.

The **second segment**, e.g. 2, is the "series number", which can be 1, 2, 3 or 4 explained above.

The **third segment**, e.g. 017, is the volume number within a series.

The **fourth segment**, e.g. 009, is the document (pamphlet etc.) number within a volume. During the scanning process, all documents were given numbers sequentially from the first page to the end of the volume.

The exception are non-document pages within a volume. In this case, the fourth segment is "peri" (i.e. peritext) instead of a number.

The **fifth segment**, e.g. 0004, is the page number within a single document (pamphlet etc.).
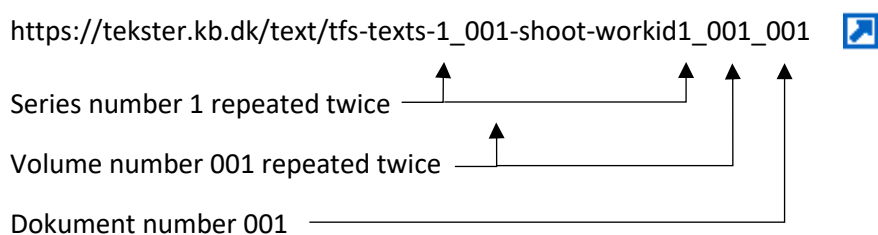
Note that the ALTO files contain information about identified text styles, and the position (in pixels) on the corresponding facsimile of each line and word. However, in case of pages that needed major corrections/additions the position information might be incorrect (repositioning in pixels was not part of the proofreading process).

Relationship between the files and the library's web portals

There is a 1:1 relationship between the file names in loar.kb.dk and the URL's used in the library's web portals.

Example

Presentation of XML and TIFF files named 130017168305_1_001_001_001.{xml,tif}
(located in folder 1_001_130017168305) in the library's text portal:

> https://tekster.kb.dk/text/tfs-texts-1_001-shoot-workid1_001_001

Series number 1 repeated twice

Volume number 001 repeated twice

Dokument number 001

Note that the "bar code" number (physical location) is not used in tekster.kb.dk

tekster.kb.dk displays all texts and facsimiles within a document on a single web page. Therefore, the page numbers do not appear tekster.kb.dk's URL.

Peritext pages are not included in tekster.kb.tk

The facsimiles (original high resolution TIFF in loar.kb.dk) including peritext are available as JPEG's in any resolution (less or equal to the original resolution) on the Royal Danish Library's image server through URL's.

Above example, first facsimile:

> https://kb-images.kb.dk/public/tekstportal/tfs/1_001_130017168305/130017168305_1_001_001_001/full/400,/0/default.jpg

> Contrary to tekster.kb.dk the full folder name and the full file name appear in the URL.

> As defined in the IIIF image API protocol, the"400," segment define that the produced JPEG has 400 pixels' width. You may substitute with another pixel value, or "full" to fetch the maximum resolution.

Other information

The digital collection (46 physical volumes) is spreading over these years:

Year published	Count	
1760	1	
1770	20	
1771	345	
1772	303	
1773	58	
1774	35	
1775	9	
1776	1	(Copper engraving)
Unknown/doubt	148	
Total	**920**	

In the old registry (series 4) there are **79** separat documents/prints.

Thus, the Royal Danish Library's digital collection Freedom of Press Writings amounts to **999** separate documents. They were scanned from **49** physical volumes.

Not part of the present digital collection

Mr. Luxdorph provided a volume with folios, which is not part of the Royal Danish Library's collections. It is part of the Frederiksborg Museum's collections.

Luxdorph's 47th volume contains prints of 32 woodcuts, 12 copper engravings and 21 smaller engravings.

Bo Krantz Simonsen (boks@kb.dk)
Royal Danish Library
December 2021