

Fulltext

Fulltext data sets contain both the actual contents (often digitized and OCRd), as well as metadata about the documents.

Books

Copyright free books that the National Library has digitised from its collections.

User interface

[Doria](#)

APIs

[Doria OAI-PMH](#) (use parameter `set=com_10024_109240`)

[Doria OpenSearch](#) (use parameter `scope=10024/109240`, e.g. [search laplanders](#))

Individual documents may be downloaded from Doria.

License

CC0 for most titles, with few exceptions as CC-BY

Classics Library

A collection of classic Finnish fiction from 19th and 20th centuries.

User interface

[Doria](#)

APIs

[Doria OAI-PMH](#) (use parameter `set=col_10024_88083`)

[Doria OpenSearch](#) (use parameter `scope=10024/88083`, e.g. [search rakkaus](#))

Individual documents may be downloaded from Doria.

License

CC0

Collection Catalogues

Digitized catalogues and card files of the National Library collections. Collections are not fully catalogued in the library databases, hence the old card files and catalogues can provide supplemental information on the collections.

User interface

[Doria](#)

APIs

[Doria OAI-PMH](#) (use parameter `set=com_10024_111861`)

[Doria OpenSearch](#) (use parameter `scope=10024/111861`, e.g. [search machine](#))

Individual documents may be downloaded from Doria.

License

CC0

Digi collection texts and metadata

Metadata of digitized collections texts and metadata

Description

Metadata of digitized collections texts and metadata

User interface

[digi.kansalliskirjasto.fi](#) Collections

Data downloads

[Digi.kansalliskirjasto.fi/opendata](#) -page

Select file: Digi collection texts and metadata [v1](106.2 kB)

APIs

License

[Terms of use](#)

Digitalia data packages

Digitalia (2017-2019)

[Uusi Suometar \(1457-4721\) REOCR ALTO XML](#)

[Uusi Suometar \(1457-4721\) ALTO XML](#)

Dissertations of the Royal Academy of Turku

This collection contains 4173 digitized dissertations that were defended at the Royal Academy of Turku between 1642 and 1828. The collection also includes a number of Pehr Kalm's dissertations.

User interface

[Doria](#)

APIs

[Doria OAI-PMH](#) (use parameter [set=col_10024_50699](#))

[Doria OpenSearch](#) (use parameter [scope=10024/50699](#), e.g. [search aquae](#))

Individual documents may be downloaded from Doria.

License

CC0

Ephemera Collection

A digitised collection of ephemera from the legal deposit collections of the National Library of Finland. Subject matters include tourism, protection of animals, war-time rationing, women's movement, etiquette, sports, board games and vehicles. Publication dates range from early 19th century to 1944.

User interface

[Doria](#)

APIs

[Doria OAI-PMH](#) (use parameter [set=com_10024_85119](#) except for the board games use [set=col_10024_121989](#))

[Doria OpenSearch](#) (use parameter [scope=10024/85119](#) or [10024/121989](#) for the board games, eg.g. [search chrysler](#))

Individual documents may be downloaded from Doria.

License

CC0

Fenno-Ugrica

Fenno-Ugrica is a digital collection of publications in Uralic languages. The Fenno-Ugrica collection includes more than 1500 monographs and over 110 newspaper and journal titles in 20 languages. The collection also features word lists, which are generated from the digitized and edited books by language. Zip-files with full-text and images are included with some of the titles.

User interface

[Fenno-Ugrica](#)

APIs

[Fenno-Ugrica OAI-PMH](#), and direct link to the [OAI-interface](#)

[OpenSearch](#), e.g. [search](#) on the books collection

Individual documents may be downloaded from Fenno-Ugrica.

License

Public domain based on due diligence agreement, Certificate is available in <http://s1.doria.fi/ohje/img-603112949-0001.pdf>

Fin-Clariah dataset - Copyright-free Finnish newspapers and periodicals

Digitised collection of copyright-free newspapers and periodicals published in Finland. This dataset is available via Allas-service in CSC via Fin-clariah project.

Description

Digitised collection of copyright-free newspapers published in Finland. This dataset is available via Allas-service in CSC via Fin-clariah project. See detailed instructions [here](#).

Dataset id links from Fin-Clariah dataset to metadata records can be found below.

User interface

[Newspapers](#) at digi.kansalliskirjasto.fi

Data downloads

- Newspapers until 31.12.1918
- Journals until 31.12.1912
- Copyright free books

APIs

[Digi OAI-PMH](#)

https://digi.kansalliskirjasto.fi/interfaces/OAI-PMH?metadataPrefix=oai_dc&set=col-861&verb=ListIdentifiers

License

[Terms of use](#)

Finnish Civil War And Independence

A selection of ephemera from the events of 1917 and 1918 in the midst of Finnish civil war. The collection offers documents on the Red Guards, the White Guard, inserts for the newspapers, declarations and food supply.

User interface

[Doria](#)

APIs

[Doria OAI-PMH](#) (use parameter [set=com_10024_111871](#))
[Doria OpenSearch](#) (use parameter [scope=10024/111871](#), e.g. [search mannerheim](#))
Individual documents may be downloaded from Doria.

License

CC0

Finnish journals -1939

Digitised collection of generic journals in Finland until end of 1939.

Description

[Detailed description](#) (in Finnish)

Note that 1921-1939 is opened by agreement with Kopiosto and National Library for year 2023.

User interface

[Journals](#) at [digi.kansalliskirjasto.fi](#)

Data downloads

Zip packages, which custom XML contains metadata, ALTO XML, and raw text of a page.

Data package contains journal material until 1910.

APIs

[Digi OAI-PMH](#)

License

[Terms of use](#)

Finnish newspapers' layout analysis (METS package) 1771-1917

The layout analysis files from digitisation for Finnish newspapers, years 1771-1917.

Description

Zip packages, which contain METS XML for each binding. METS xml standard contains layout information of the materials and technical processing information.

Note! Due to improvements in materials, the few years back created ALTO XML export packages are not fully in sync with the METS information. I.e. some binding id's that exist in ALTO exports can be missing from METS, which have been generated in early September 2018.

User interface

[Newspapers](#) at [digi.kansalliskirjasto.fi](#)

Data downloads

<https://digi.kansalliskirjasto.fi/opendata/submit> Pick (Other)

APIs

License

[Terms of use](#) (in Finnish).

Finnish newspapers 1771-1939

Digitised collection of newspapers published in Finland from the 18th century up until 1939.

Description

Note that materials of 1918-1939 is opened by agreement with Kopiosto and National Library for year 2023

User interface

[Newspapers](#) at [digi.kansalliskirjasto.fi](#)

Data downloads

[Zip packages, which custom XML contains metadata, ALTO XML, and raw text of a page.](#) Data packages contain material of newspapers until end of 1917 and journals until end of 1910.

APIs

[Digi OAI-PMH](#)

[Digi OpenURL](#)

License

[Terms of use](#)

Fragmenta Membranea Collection

The Fragmenta membranea collection contains the vast majority of the remains of books written and used in the eastern parts of medieval Sweden, the Diocese of Turku. The Fragmenta membranea database contains 9,319 digitized parchment leaves meaning 18,638 pages which come from approximately 1,500 different medieval manuscripts.

User interface

[Fragmenta membranea](#)

APIs

[Fragmenta OAI-PMH](#)

[OpenSearch](#) (e.g. [search Gloria](#))

Individual documents may be downloaded from Fragmenta Membranea.

License

CC0

History of the books

A broad collection of books and other texts from the 18th and 19th centuries ranging from devotional books and broadside to educational material and fiction. There are also catalogues from book actions.

User interface

[Doria](#)

APIs

[Doria OAI-PMH](#) (use parameter `set=com_10024_144026`)

[Doria OpenSearch](#) (use parameter `scope=10024/144026`, e.g. [search rakkaus](#))

Individual documents may be downloaded from Doria.

License

CC0

Illustration base type classifier model file

Illustration base type classifier model file for newspaper, journal etc. illustration categorization.

User interface

- `nlf_basetype_classifier.pb`
- `nlf_basetype_classifier_labels.txt`

Some examples of concept here: <https://blogs.helsinki.fi/digitalia/?s=tensorflow&submit=Search>

Classifier model file can be used with TensorFlow (https://www.tensorflow.org/guide/saved_model)

When using the file, please cite:

<https://digi.nationallibrary.fi> , *Digital Collections of National Library of Finland, Illustration classifier model file of Digitalia, 30.9.2019.*

Data downloads

<http://digi.kansalliskirjasto.fi/opendata>

API

-

License

[Terms of use](#)

Manuscript collection

Digitised material from the Manuscript Collection. Versatile material includes Medieval and sixteenth-century manuscript books, Mannerheim's Fragment Collection, Paul Scheel's letter collection, parchment Letters, J.J. Tikkanen's sketch books and Väinö Raitio's musical manuscripts. Also the main card index of the Manuscript Collection is available.

User interface

[Doria](#)

APIs

[Doria OAI-PMH](#) (use parameter `set=com_10024_109242`)

[Doria OpenSearch](#) (use parameter `scope=10024/109242`, e.g. [search Mannerheim](#))

Individual documents may be downloaded from Doria.

License

CC0

Maps and Atlases of Finland

A collection of digitized maps about Finland ranging from the 16th century to 20th century. Map types include Town maps, general maps of provinces and regions, nautical charts, town and parish maps, and Atlases.

User interface

[Doria](#)

APIs

[Doria OAI-PMH](#) (use parameter `set=com_10024_78800`)

[Doria OpenSearch](#) (use parameter `scope=10024/78800`, e.g. [search Turku](#))

Individual documents may be downloaded from Doria.

License

CC0

Nordenskiöld Map Collection, The

A selection of digitized maps from the Nordenskiöld Collection. The maps depict the development of Western countries' geographical knowledge. They cover all continents, with a particular emphasis on Arctic areas. There is an almost complete series of the Geographica, the classic cartographic work by Claudius Ptolemy, as well as a considerable number of works related to the discovery of America.

User interface

[Doria](#)

APIs

[Doria OAI-PMH](#) (use parameter [set=com_10024_97216](#))

[Doria OpenSearch](#) (use parameter [scope=10024/97216](#), e.g. [search Belgium](#))

Individual documents may be downloaded from Doria.

License

CC0

OCR Ground Truth Package for Finnish Fraktur

Package contains 450 page images and ALTO XML files for each page, with the proofreading done by the Finnish native speakers.

Description

The pages of fraktur range from the year 1836 until 1910. The package can help in creating own postcorrection algorithms for OCR text recognition.

There is also an Excel file for all of the 471 903 words, which contains result given to the word by Tesseract and FineReader. If a tool hasn't found corresponding word, then the given cell is empty, so select the words in the Excel, which you need.

NB! The ground truth package does not contain the data for the 1918 due to copyright reasons.

User interface

Data downloads

<http://digi.kansalliskirjasto.fi/opendata>

API

-

License

[Terms of use](#) (in Finnish).

OCR Ground Truth Package for Swedish Fraktur

Package contains page images and ALTO XML files for each page, with the proofreading done in Swedish by the Finnish native speakers.

Description

The pages of fraktur range from the year 1771 until 1915. The package can help in creating own post correction algorithms for OCR text recognition.

Note1. the tiff files exif metadata lacks resolution information, so if the coordinates of ALTO do not match, be aware that images has been done either 200 or 300 dpi.

Note2. The ground truth package does not contain the data of the 1918 or later due to copyright reasons.

User interface

Data downloads

<http://digi.kansalliskirjasto.fi/opendata>

API

-

License

[Terms of use](#)

Raita: Early Finnish Recordings

Raita is a collection of digitized early Finnish sound recordings.

User interface

[Doria](#)

APIs

[Doria OAI-PMH](#) (use parameter [set=col_10024_66373](#))

[Doria OpenSearch](#) (use parameter [scope=10024/66373](#), e.g. [search Verdi](#))

Individual documents may be downloaded from Doria.

License

CC0

Technical Ephemera Collection

Digitised collection of technical ephemera (selection of brochures, ads, leaflets, price catalogues and instruction guides)

Description

[Detailed description](#) (in Finnish)

User interface

[Ephemera](#) at digi.kansalliskirjasto.fi

Data downloads

-

API

-

License

Tesseract3 Finnish fraktur model

Tesseract 3 Finnish fraktur model

User interface

Copy the file to the Tesseract's TESSDATA directory.

You can utilize the file in Tesseract via:

```
tesseract input.jpg out -l fi_frak_nlf
```

See also tesseract 3 in Github: <https://github.com/tesseract-ocr/tesseract>

Data downloads

<http://digi.kansalliskirjasto.fi/opendata>

API

-

License

[Terms of use](#)

Translocalis clippings 1820-1885

Translocalis clippings 1820-1885

Description

Translocalis is a digital database for reader letters written in different locations and published in Finnish papers up to the year 1885. The Translocalis database contains 72 000 reader letters from Finland and abroad. In the name of the collection, trans refers to an object going over or through something and localis refers to a space or location. Combined, Translocalis expresses something more than local, which these local letters represented. The Finnish-speaking press started out on a nation-wide level and became more regional and local only during the latter half of the 19th century.

User interface

<https://digi.kansalliskirjasto.fi/collections?id=742>

Data downloads

Zip package, which contains

- individual text file for each clipping text
- Log of downloaded clippings
- Clippings metadata excels 1820-1883 and 1884-1885

Folder Structure:

translocalis_data/year/ISSN/txt/388373_2973653_1457-4403_1868-04-09_15_page-2

- File name is formulated by article_id, binding_id, issn, publishing date (YYYY-MM-DD), issue, and page on which clipping has been taken.

Clippings metadata (2 files)

(part until 1883): translocalis_clippings_export_1820_1883.xlsx

(1884-1885): translocalis_clippings_export_1884_1885.xlsx

Fields in metadata excels:

Main title - name of the newspaper

ISSN - (International Standard Serial Number), i.e. newspaper identifier

Date - Publishing date of the number

Issue - Issue number of newspaper (can be empty, or contain A, Bs e.g. for extra editions)

URL - link to the clipping page

Title - Title of the article

Keywords - keywords of clipping (see [Wiki](#) for full explanations of translocalis_xyz fields)

Category - category of the article

Subject

Notes - if any extra notes

Created - when clipping was created

OCR - text contents of the clipping

APIs

[Digi OAI-PMH](#)

[Digi OpenURL](#)

License

[Terms of use](#)

Uusi Suometar (1457-4721) ALTO XML

ALTO XML files of newspaper Uusi Suometar (1457-4721) years 1869-1917.

Description

ALTO XMLs as they have been produced in the digitisation.

Year 1918 excluded.

User interface

<https://digi.kansalliskirjasto.fi>

Data downloads

<https://digi.kansalliskirjasto.fi/opendata>

API

-

License

[Terms of use](#)

Uusi Suometar (1457-4721) REOCR ALTO XML

The REOCR'd ALTO XML files of newspaper Uusi Suometar (1457-4721) years 1869-1917.

Description

ALTO XMLs as they have been produced in the digitisation.

Year 1918 excluded.

User interface

<https://digi.kansalliskirjasto.fi>

Data downloads

<https://digi.kansalliskirjasto.fi/opendata>

API

-

License

[Terms of use](#)