

# Data sets

## 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.
- [Classics Library](#) — A collection of classic Finnish fiction from 19th and 20th centuries.
- [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.
- [Digi collection texts and metadata](#) — Metadata of digitized collections texts and metadata
- [Digitalia data packages](#)
- [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.
- [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.
- [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.
- [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.
- [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.
- [Finnish journals -1939](#) — Digitised collection of generic journals in Finland until end of 1939.
- [Finnish newspapers' layout analysis \(METS package\) 1771-1917](#) — The layout analysis files from digitisation for Finnish newspapers, years 1771-1917.
- [Finnish newspapers 1771-1939](#) — Digitised collection of newspapers published in Finland from the 18th century up until 1939.
- [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.
- [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.
- [Illustration base type classifier model file](#) — Illustration base type classifier model file for newspaper, journal etc. illustration categorization.
- [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.
- [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.
- [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.
- [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.
- [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.
- [Raita: Early Finnish Recordings](#) — Raita is a collection of digitized early Finnish sound recordings.
- [Technical Ephemera Collection](#) — Digitised collection of technical ephemera (selection of brochures, ads, leaflets, price catalogues and instruction guides)
- [Tesseract3 Finnish fraktur model](#) — Tesseract 3 Finnish fraktur model
- [Uusi Suometar \(1457-4721\) ALTO XML](#) — ALTO XML files of newspaper Uusi Suometar (1457-4721) years 1869-1917.
- [Uusi Suometar \(1457-4721\) REOCR ALTO XML](#) — The REOCR'd ALTO XML files of newspaper Uusi Suometar (1457-4721) years 1869-1917.

## Metadata

Metadata repositories contain bibliographic metadata records. When the documents described are available online, the metadata provides a link to them.

- [Arto](#) — ARTO is an aggregation of metadata on Finnish periodical and monograph articles. Formerly a discrete database, it is now part of Melinda.
- [Fennica](#) — Fennica - the Finnish National Bibliography is a database dedicated to Finnish publication activities.
  - [Repository description: Fennica](#)
  - [Tietovarantokuvaus: Fennica](#)
- [Fennica-LD](#) — This is the Linked Data version of Fennica - the Finnish National Bibliography.
- [Finna.fi](#) — Finna is a search service that aggregates metadata from Finnish archives, libraries and museums.
- [Melinda](#) — Melinda is the National Metadata Repository (union catalog).
- [Viola](#) — Viola is the Finnish national discography and the national bibliography of sheet music.

## Vocabularies

Controlled vocabularies such as thesauri, ontologies and authority files are referenced from bibliographic metadata. We provide access to controlled vocabularies mainly through the Finto <http://finto.fi> service.

- [Allärs - Allmän tesaurus på svenska](#) — Allärs is the Swedish translation of YSA - General Finnish Thesaurus.

- [FGF - Finnish genre and form vocabulary](#) — FGF (in Finnish known as SLM) is a bilingual (Finnish + Swedish) vocabulary for describing genres and forms of literature (both fiction and non-fiction) and music.
- [Finnish Corporate Names](#) — The Finnish corporate names data set is used by the National Library of Finland in the description of the national bibliography Fennica.
- [ISIL Identifiers of Finnish Libraries](#) — An ISIL identifies a library, an archive, a museum or a related organization, or one of its subordinate units.
- [KOKO Ontology](#) — KOKO is a collection of Finnish core ontologies, which have been merged together.
- [Metadata thesaurus](#) — Metadata thesaurus contains terms and expressions required in describing materials. The thesaurus is also suitable for selecting headings displayed in user interfaces.
- [PLC - Finnish Public Libraries Classification System](#) — The classification system for Finnish public libraries PLC (known by its Finnish acronym YKL) is an adaptation of the Dewey Decimal Classification.
- [SEKO - Finnish Medium of Performance Thesaurus](#) — SEKO is a thesaurus in Finnish which covers instruments, voices etc. used in the performance of musical works.
- [UDC Summary](#) — The Multilingual Universal Decimal Classification Summary.
- [YSA - General Finnish Thesaurus](#) — YSA is a general thesaurus in Finnish which covers all fields of research and knowledge, which contains the most common terms and geographical names used in content description.
- [YSE - concept suggestions for YSO](#) — YSE is a collection of concepts that have been suggested for inclusion in YSO, but they have not yet been accepted.
- [YSO - General Finnish Ontology](#) — General Finnish Ontology YSO is a trilingual ontology consisting mainly of general concepts.
- [YSO places](#) — YSO places is a multilingual gazetteer.