OPEN DATA POLICY OF ELNET CONSORTIUM

The strategic goal of the ELNET Consortium (hereinafter ELNET) for 2019-2022 is to increase access to the data in the online catalogue ESTER, article database ISE and Estonian subject thesaurus EMS, and to develop new services. One way for achieving this goal is to make the data in the ELNET datasets accessible as open data.

The making available of the ELNET Consortium open data aims to make the data in the datasets accessible in machine-readable format in order to support the reuse and interbase cross-usage of data by as many target groups as possible.

Offering the data managed by ELNET as open data enables to use them in linked data or semantic web and share with other organisations the best practices how to modify data and publish it as linked data.

The aim of the ELNET open data policy is to regulate and describe the publication of the data in ELNET datasets as open data.

> **Note:** Only the data having the consent of all the ELNET Consortium members and partners is published as open data. For the use of other data, please contact the particular library.

The ELNET open data policy is a strategic document, the open data working group regularly assesses whether it is up-to-date and proposes the necessary amendments. The policy is approved by the General Meeting of the ELNET Consortium.

Concepts and abbreviations

Concept	Explanation
Open data	Open data means the data in a machine-readable format that is made available for public use and that has no limits to usage or sharing. Therefore, personal data with restrictions on access and data, the sharing of which is limited by law, are not considered open data. Open data is: ● available free to all persons for any use; ● digital, machine-readable and cross-usable with other data; ● licensed for unrestricted use and further dissemination.
Dublin Core	One of the standards for creating a list of metadata elements for interdisciplinary description of information resources.
BIBFRAME	*Bibliographic Framework.* A model of bibliographic description. BIBFRAME was created to replace MARC standards and to use the principles of linked data in order to make bibliographic data more useful both within the library community and beyond.
EDM	*Europeana Data Model.* The formal specification of Europeana data model, listing the classes and properties used for presenting data in Europeana and support the functionality of Europeana.
IFLA LRM	*IFLA Library Reference Model.* A high-level conceptual reference model designed to be used in linked data environments and to support and promote the use of bibliographic data and authority data.
Human-readable format	A format understandable for human beings (and also computers) where the presentation of the data enables people to read it in a natural way, e.g. file format allowing plain text content.
Personal data	Any data concerning an identified or identifiable natural person, regardless of the form or format in which such data exist.
JSON	*JavaScript Object Notation.* A lightweight data interchange format based on open standard.

Linked data	Structured data which is described and published in a manner enabling to automatically link them. Linked data may or may not be open data.
Machine-readable format	Machine-readable format means a file format structured so that software applications can easily identify, recognise and extract[1] specific data, including individual statements of fact, and their internal structure.
MARC21	Data exchange format which presents, in conformity with rules, a unit of bibliographic and authority data, enabling to fully identify the publication under description. The records are usually prepared into an online catalogue or bibliographic database. MARC21 bibliographic record: https://www.loc.gov/marc/bibliographic/
MARCXML	MARCXML describes MARC21 format in XML.
XML	*Extensible Markup Language.* A platform independent human- and machine-readable markup language for describing data and data structures.
RDA	*Resource Description and Access.* A ruleset for describing publications and creating authority data. RDA is based on the Functional Requirements for Bibliographic Records (FRBR).
RDF	*Resource Description Framework.* The standard model of semantic web concept, for the conceptual describing or modelling of resources (data or other entity) and data exchange in the web. Extends linking structure by showing the links of objects with URI, based on triplets subject – predicate – object.
SUTRS	Simple Unstructured Text Record Syntax (widely used in data exchange format Z39.50).
URI	Uniform Resource Identifier. A compact string of characters identifying an abstract or physical resource. The string refers to the object (file, programme, service) on the internet.

---

[1] Under the Green Paper on Open Data and the Estonian open data policy, open data has to be presented in at least 3-star format in order to meet the requirement of machine-readability: its usage is governed by free license, it can be used by any software applications, it follows open standards.

Introduction

Open data is data in machine-readable format that has been made available to all for free and public use, with no restrictions on its use and dissemination.

A system of formats and codings or a five-star rating system of open data has been set for assessing the „user-friendliness" or extent of machine-readability of open data[2].

The open data of ELNET is:
- available to all for free for any purpose of usage;
- digital, machine-readable and cross-usable with other data;
- up-to-date, i.e. regularly updated;
- licensed and with no restrictions on use and further dissemination.

Databases and datasets made available as open data

The following databases and data contained in them are made available as open data by the ELNET Consortium:
- Estonian Subject Thesaurus EMS (Annex 1)
- Database of Estonian articles ISE (Annex 2)
- Online catalogue ESTER
  - Descriptions of publications (Annex 3)
  - Data of items (Annex 4)

The data composition of each database and datasets made available as open data are described in the annexes of this document.

ELNET makes data available as 4-star open data.

---

[2] *   OL (open license) – the data is in any format under an open license
**   OL+RE (machine readable) – the data is in structured format under an open license, e.g. Excel
***   OL+RE+OF (open format) – the data is displayed in open format, e.g. CSV, XML.
****   OL+RE+OF+URI – the objects/things in data are identified with URIs (andmetes olevaid objekte/asju identifitseeritakse URI-dega)
***** OL+RE+OF+URI+LD (linked data) – the data is linked with other data via URIs.

Formats and standards

The following data formats are supported:
- CSV
- XML
- MARC21 (machine-readable)
- JSON

The following data standards and data models are supported:
- Dublin Core (DC)
- Simple Unstructured Text Record Syntax (SUTRS)
- Europeana Data Model (EDM)
- MARC21

NB! In different databases data formats may vary. See more in the annexes of this document.

Methods, protocols and channels of data aggregation

The following are supported:
- Z39.50 protocol
- Sierra REST API
- HTTP/HTTPS
- OAI-PMH
- FTP

NB! In different databases the data exchange channels and protocols may vary. See more in the annexes of this document.

Frequency of data updates

The update frequency of each single database and its datasets is described in the annexes of this document.

Licenses

The data has been made available by license Creative Commons Zero (CC0) [ver. 1].

Personal data

The data is made available under the privacy policy of the respective ELNET Consortium database, each privacy policy is available on the website of the corresponding database.

Contact

The information is available on the website at https://data.elnet.ee.
All queries on open data are welcome at elnet@elnet.ee.

--

Note

The annexes of this document are regularly updated by the Head of the ELNET Consortium Open Data Working Group and the General Administrator of the information system ESTER.

VISION 2030

- ELNET Consortsium accedes to the [Estonian Open Government Data Portal](#) in 2020.

- RDA (*Resource Description and Access*) will be taken into use as the cataloguing standard for publications, and IFLA LRM (*IFLA Library Reference Model*) will be taken into use as the conceptual data model.

- The ELNET Consortium will make data available as 5-star[3] open data and offers services via SPARQL ENDPOINT.

- In order to facilitate the availability of linked open data, the Consortium takes into use internationally recognised ontologies and vocabularies, and provides both bibliographic and authority records with permanent identifiers and URIs.

- In order to start producing data as linked data and thus give up the conversion of data from MARC format into formats supporting linked data, the Consortium explores alternative solutions for bibliographic description, e.g. BIBFRAME.

---

[3] *   OL (open license) – the data is in any format under an open license
**   OL+RE (machine readable) – the data is in structured format under an open license, e.g. Excel
***   OL+RE+OF (open format) – the data is displayed in open format, e.g. CSV, XML.
****   OL+RE+OF+URI – the objects/things in data are identified with URIs
***** OL+RE+OF+URI+LD (linked data) – the data is linked with other data via URIs.

ANNEX 1. ESTONIAN SUBJECT THESAURUS EMS

The records of terms in the Estonian Subject Thesaurus are compiled on the basis of Anglo-American Cataloguing Rules (AACR2).

DATASETS made available as open data:
- chronological terms
- geographic terms
- form/genre terms
- topical terms
- the entire thesaurus

CHANNELS:
- API - https://ems.elnet.ee/index.php
- OAI-PMH - https://data.elnet.ee
- HTTP - https://data.elnet.ee

DATA AGGREGATION, FORMATS and FREQUENCY OF UPDATES:

	API	OAI-PMH	HTTP
MARC21 (ISO 2709)	*real-time*	*once a month*	*once a month*
MARC21XML	*real-time*	*once a month*	*once a month*
MARC21 (human-readable)	*real-time*	-	-

LISA 2. DATABASE OF ESTONIAN ARTICLES ISE

Bibliographic descriptions in the Database of Estonian Articles ISE have been compiled under the following standards and guidelines:

- IFLA Guidelines for the application of the ISBDs to the description of component parts
- MARC 21 Format for Bibliographic Data
- Anglo-American Cataloguing Rules (AACR2)
- MARC21 analytical description guidelines for ELNET Consortium member libraries

DATASETS made available as open data:

- ISE databases, each separately (e.g. social sciences, linguistics)
- thematic databases, each separately (e.g. Bibliography of Military Matters)[4]
- the entire ISE (except thematic databases)

CHANNELS:

- Parameters of data exchange Z39.50
- Sierra REST API
- X-record
- OAI-PMH - https://data.elnet.ee
- HTTP - https://data.elnet.ee

DATA AGGREGATION, FORMATS and FREQUENCY OF UPDATES:

	Z39.50	REST API	X-RECORD	OAI-PMH	HTTP
MARC21 (human-readable)	*real-time*	*real-time*	-	-	*once a month*
MARC21XML	*real-time*	*real-time*	-	*once a day*	*once a month*
XML (MARC21)	-	-	*real-time*	-	-
MARC21-JSON	-	*real-time*	-	-	-
SUTRS	*real-time*	-	-	-	-
Dublin Core	*real-time*	-	-	*once a day*	*once a month*
EDM				*once a day*	

---

[4] Exception: the ELNET open data portal contains a reference to the open data portal of the National Library of Estonia, which makes available the thematic databases compiled by the National Library. These are: Bibliography of Estonian Legal Publications, database of reproductions "Reprod", Database of the President of Estonian Republic

Not made available as open data:

Deleted records	BCODE3=d
Hidden records	BCODE3=n

Before data export, the records must be processed as follows, in the order indicated below:

MARC field	RULE
001	delete
003	delete
907	- enter data in field 001 (presumed that the previous 001 has been deleted from records) - delete full stop in front of the record - after that delete field 907
003	generate ErISE *(NB! The previously present 003 must be deleted from records)*
9xx	Delete

ANNEX 3. ONLINE CATALOGUE ESTER | BIBLIOGRAPHIC DESCRIPTIONS (BIBLIOGRAPHIC RECORDS)

The following standards and guidelines have been followed while creating the bibliographic descriptions for punlications in the online catalogue ESTER.

- MARC 21 Format for Bibliographic Data
- Anglo-Americal Cataloguing Rules (AACR2)
- ELNET Consortium Cataloguing Manual

DATASETS made available as open data:

- Databases based on material type (for exceptions see below "Not made available as open data")
- E-resources (on the basis of BIBLEVEL)
- National publications – the ELNET open data portal contains a reference to the National Library of Estonia open data portal which publishes the entire Estonian national bibliography as open data

CHANNELS:

- Parameters of data exchange Z39.50
- Sierra REST API
- X-record
- OAI-PMH - https://data.elnet.ee
- HTTP - https://data.elnet.ee

DATA AGGREGATION, FORMATS and FREQUENCY OF UPDATES:

	Z39.50	REST API	X-RECORD	OAI-PMH	HTTP
MARC21	*real-time*	*real-time*	-	-	*once a month*
MARCXML	-	-	*real-time*	-	-
MARC21XML	-	*real-time*	-	*once a day*	*once a month*
MARC-JSON	-	*real-time*	-	-	-
SUTRS	*real-time*	-	-	-	-
EDM	-	-	-	*once a day*	-
Dublin Core	*real-time*	-	-	*once a day*	*once a month*

Not made available as open data:

Deleted records	BCODE3=d
CIP records	Leader/17=8 BCODE3=e
Hidden records	BCODE3=n
ISN registry records - hidden	Leader/17=z BCODE3=i
Hidden order records	BCODE3=c
Records of withdrawals	BCODE3=m
Records prepared for gadgets	ltype=x
Publications with restricted access	field 506 |a*Access restricted...*
British Library	under code in field 040: *Uk*
Bibliographic records related only to order records, without copies	

Bibliographic records bulk downloaded from licensed databases will not be made available as open data:

CAB	740 0# CAB eBooks Front File (2014-present)
EBSCO	740 0# Academic Complete 740 0# EBSCO eBooks
Elsevier	740 0# Elsevier e-books
Estonian University of Life Sciences e-books	740 0# EMÜR e-raamatud
NAXOS	740 0# Naxos Music Library
OverDrive	740 0# OverDrive e-books
Oxford Scholarship Online	740 0# Oxford Scholarship Online 740 0# Oxford scholarship online.|pEconomics and finance 740 0# Oxford scholarship online.|pPolitical science

ProQuest Ebook Central	740 0# ProQuest Ebook Central Collection 740 0# Ebook Central|5RR 740 0# ProQuest Ebook Central e-books
ScienceDirect	740 0# ScienceDirect ebooks
Springer	740 0# Springer e-books 740 0# Springer e-books.|pBiomedical and Life Sciences 740 0# Springer e-books.|pEarth and Environmental Science 740 0# Springer e-books.|pLecture notes in computer science
Taylor & Francis	740 0# Taylor & Francis eBooks
Wiley	740 0# Wiley ebooks

Before data export, the records must be processed as follows, in the order indicated below:

MARC field	RULE
001	delete
003	delete
035 (Tartu)	delete
035 (Tartu-L)	delete
089	delete
880	delete
907	- enter data in field 001 (presumed that the previous 001 has been deleted from records) - delete full stop in front of the record - after that delete field 907
003	generate ErESTER *(NB! The previously present 003 must be deleted from records)*
9xx	delete

ANNEX 4. ONLINE CATALOGUE ESTER | ITEMS

CHANNELS:
- OAI-PMH - https://data.elnet.ee
- HTTP - https://data.elnet.ee

DATA AGGREGATION, FORMATS and FREQUENCY OF UPDATES:

	OAI-PMH	HTTP
MARC21XML	*daily*	*daily*
MARC21	*daily*	*daily*
CSV	*daily*	*daily*

DATA COMPOSITION:

FIELD	DATA FIELD DESCRIPTION
No of bibliographic record	81
LOCATION	79
STATUS	88 – values selectively: - 2 – IN 2nd FLOOR READ. ROOM - 3 – IN 3rd FLOOR READ. ROOM - 4 – NO LENDING - 9 – IN RESERVE COLLECTION - - – AVAILABLE - w – FOR USE IN LIBRARY - c - UNDEFINED - e – FOR INSTITUTE STUDENT - u – ON NEW BOOKS STAND - v – REFERENCE COLLECTION
Call number	c c - 090 c - 091 (except ERM)
VOLUME	v
ISBN	h

NOT MADE AVAILABLE as open data:

FIELD	DATA FIELD DESCRIPTION
LOCATION	79, values: - www-items (hidden ICODE2=n) - KV - TlnKR - Tallinna Linnaarhiiv (Tallinn City Archive)
ICODE1	59, values: - 9999 (hides item)
ICODE2	60, values: - d – to be deleted - m – to be withdrawn - n - hidden
TYPE OF ITEM	61, values: - 183 - gadget (TÜR)
CALL NUMBER	c, values: - 091 (ERM)
DUE DATE	65, values: - dd.mm.yyyy (due date, items checked out)
STATUS	88, values: - 7 – in deleting - 8 – in digitising - a – archival copy - d – on exhibition - f – not arrived - g - reserved - h – packed for moving - i - inventory - j - sold - k – in processing - m - lost - n – invoice issued - o – in use by library - p – no access - q – in binding - r – in repair - s – on search - z – claims returned - x - deleted - ! – on holdshelf

	- $ - lost and paid