

ETD discovery services in Lithuania

Vilius Kuciukas, Antanas Streimikis, Arunas Franckevicius, Aleksandras Targamadze

Kaunas University of Technology

Abstract

The creation of Lithuanian ETD began in 2004 when financial support from UNESCO for this purpose was received. At present 14 Lithuanian universities are taking part in this process and more than 4420 ETD's are submitted. The activities are coordinated by Kaunas University of Technology. ETD's are submitted in the repository of Lithuanian Academic Electronic Library (eLABa), the basis of which is *Fedora* repository open source software. Content model of eLABa repository includes books, research journals and articles, conference materials, research reports and ETD's. ETD's together with metadata in MARC21 XML format and other additional files are stored in the eLABa repository. There are a few possible ways how ETD's can be discovered using services of: eLABa portal, library system OPAC of Lithuanian Academic Libraries Network (LABT), portal of Lithuanian Virtual Library (LVB), union catalog of NDLTD and *Google*. eLABa portal information discovery services are based on *Lucene* software. ETD metadata are harvested with the help of OAI-PMH protocol and after they are submitted to LABT ETD union catalog. Here the search of information is possible with the help of ALEPH OPAC WEB interface. LABT ETD union catalog is integrated into LVB portal, the basis of which is MetaLib software. Metadata are also presented to international information search systems, such as NDLTD and *Google*. Access authorization to the full-text documents of ETD's is done via XACML policies. The access rights to the full-text documents are described when the license agreements with the authors are signed. ETD's can be linked with the information in other electronic resources using SFX context sensitive services. Also SRU/SRW services are created with the help of which ETD can be integrated into other external information systems.

Keywords: e-publishing, e. documents repository, digital library

Introduction

In the second quarter of the year 2007, Lithuanian Network of Academic Libraries (LABT, <http://www.labt.lt>) has combined 58 academic libraries – that of Lithuanian Academy of Sciences, those of universities (16), research institutes (22) and colleges (20). The processes in these libraries (cataloguing, circulation, serials control, acquisitions, interlibrary loan, access to the Online Public Access Catalogue (OPAC) and others) have been automated applying the ALEPH, the worldwide known software for libraries (<http://www.exlibris.co.il/aleph.htm>), which provides the base for LABT Library Information System (BIS). The LABT BIS Network has accumulated over 1,55 million bibliographic record titles and over 4,21 million corresponding bibliographic record items.

A user, with the help of public access to the internet, may perform a search on a bibliographic record (document) in e-catalogues, and after the item required has been found, take advantage of virtual services being provided, for instance, order a bibliographic record or prolong the usage term, find out information on items to be returned to the library or when a desirable copy is to be returned, etc. Still, the bibliographic record in question generally is present in print version, and the user should take it from the library or use on the library premises.

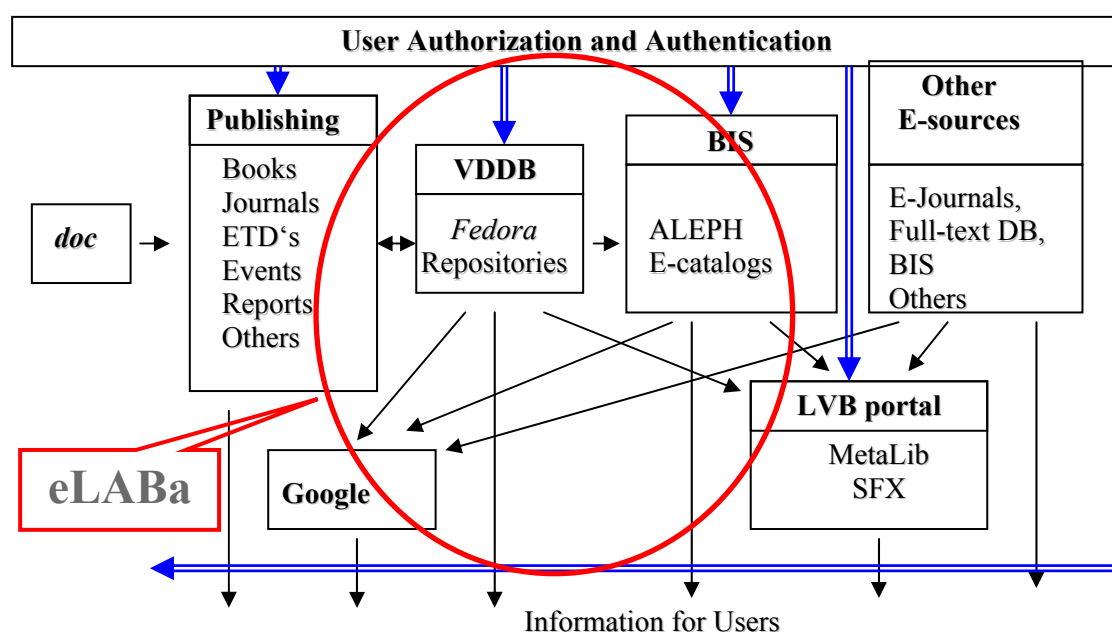
Usage of the internet, the Web, communication tools and search systems allows readers not to make a trip to the library and wait for a desirable document to be sent. Every user would reasonably want to read at the time and in the place acceptable to him or her. To meet demands of Lithuanian readers, in the first place those of scientists and students, the creation of full-text document databases (VDDDB) has been launched. One of VDDDB bases is the Lithuanian ETD document database (Franckevičius et al., 2006; Štreimikis et al., 2005; Targamadžė et al., 2004, Targamadžė et al., 2004a, Valiulis et al., 2006), which operates as a constituent of Lithuanian Academic Electronic Library (eLABa).

ETD in context of eLABa

eLABa is being created during the processes of LABT project development and implementation of the projects financed by European Union Structural Funds – “Development of Lithuanian academic e-publishing system“ (<http://leid.sf.library.lt>), “Lithuanian virtual library and creation of a full-text documents database“ (<http://lvb.sf.library.lt>), and “Lithuanian science and study e-document accumulation and delivery to the readers“ (<http://edok.sf.library.lt>). These three projects comprise the e-publishing project group.

The main goal of eLABa – to create the environment and means and using them to prepare, accumulate and store for long periods of time Lithuanian science and study e-documents and deliver them to the users via different outputs.

eLABa is being created employing the information infrastructure (Picture 1) that has been created in the context of the LABT project. A publication manuscript prepared by an author (in Picture 1 – *doc*) during the process of publishing becomes an e-document that shall be submitted into eLABa repositories (in Picture 1 – VDDb repositories, hereinafter – eLABa repositories) alongside with the metadata describing this document.



Picture 1. Information Infrastructure of eLABa

The publishing process may be carried out in two ways – traditional publishing or e-publishing. In the case of traditional publishing, the result of publishing procedures comes out to be a print version of a document, and the e-document shall be an exact copy of this print version.

The result of e-publishing is an e-document, and its print version may be prepared under user's request. To ensure e-publishing, the prevailing e-publishing systems may be applied, for instance: OJS (<http://pkp.sfu.ca/?q=ojs>), OCS (<http://pkp.sfu.ca/?q=ocs>), XEROX DocuShare (<http://docushare.xerox.com/>), VTLS Valet for ETDs (<http://www.vtls.com/Products/>) and the like.

The metadata of e-documents that have been submitted into the eLABa repositories come into e-catalogues of LABT BIS (using ALEPH software). Such e-catalogues are employed by the eLABa system as the main resource for information search.

The users may run search on information in several ways, using:

- eLABa repository search system (portal) directly,
- LABT BIS OPAC service,
- the portal of Lithuanian Virtual Library (LVB),
- *Google*, *Google Scholar* or any other internet-wide information search system known worldwide.

LVB combines various e-information sources and provides the reader with a unified and personified user interface for information search and access to e-information and virtual services that are related to it. Information submission for search via *Google*, *Google Scholar*, and other search systems, provides favorable conditions for dissemination of Lithuanian science publications on the world's internet (to be implemented in 4th quarter of 2007).

In eLABa repositories, e-documents from various science and study fields and of various trends, and their metadata are accumulated and stored up to the date of expiration of license agreements that have been entered into with the authors of the documents that may be made public.

Administrative eLABa data, which cannot be made public, are accumulated and stored up to the expiration date of legal relationship with a copyright subject or a registered user. The following e-documents of five types, which have been submitted by eLABa data suppliers, are already being accumulated, stored and published in eLABa:

- books (monograph's, textbooks, training books and other science or study publications that are or may be published in print version as a book) (hereinafter – Books);
- reviewed periodic or onetime science publications and science popularization publications in print, CD versions or published in virtual environment, and, as well, single scientific and science popularization articles, which have been published in print, CD versions or virtual environment in periodic or onetime science and science popularization publications (hereinafter – Journals)
- papers for scientific or methodical conferences, seminars and other science and study events (hereinafter – Events);
- reports on scientific research and experimental development work and projects (hereinafter – Reports);
- graduation studies, doctoral dissertations and their abstracts (hereinafter – ETD documents).

The data used to describe metadata of e-documents are obtained from the following resources (Štreimikis et al., 2007, Vilūnas, Glosienė, 2006):

- classifiers;
- international and national library-bibliographic classification systems;
- registers;
- databases that function at the level of the Ministry of Education and Science and institutions that have been established by the Ministry;
- information systems functioning at the level of the Ministry of Education and Science and institutions that have been established by the Ministry.

Lithuanian ETD's on-line

Structure of ETD objects

In eLABa repositories e-documents are accumulated that have been selected and grouped according to classification principles set on the basis of Lithuanian and international bibliographic practice, and in compliance with the eLABa conception – to accumulate and provide the user with e-documents of Lithuanian science and study that have been divided into five different collections (Books, Journals, Events, Reports and ETD), which are described in the previous chapter.

Grouping of documents into collections is related to the special software being applied to e-document submitting into eLABa.

The special software consists of two parts:

- eLABa-SPEC – a part of the special software intended for qualified e-publishing specialists, who submits into eLABa previously published e-documents into five collections (see above). It can be also ETD documents that have been published as books and have their own ISBN.
- eLABa-ETD – a part of the special software intended for postgraduates and doctorate students who submits into eLABa their ETD's (defended Master's studies, doctoral dissertations, their abstracts and the like).

E-documents are submitted to the eLABa repositories according to the type (kind) of an e-document from each collection. Depending on these elements, e-document metadata files are generated and other files required as well, which will be stored and accumulated in eLABa repositories as integral e-objects.

A description of an e-object is submitted into eLABa repository as FOXML file. According to the type (kind) chosen of e-document, an eLABa data supplier (qualified e-publishing specialist, postgraduate or doctorate student), with the help of special program, forms e-document files required, fills out the metadata fields in the form provided, and creates the FOXML file that corresponds to the object schemes being described and descriptions of the metadata.

All metadata of an e-document are stored in eLABa repositories in two formats: simple Dublin Core (DC) and expanded MARC21, which constitutes the archive eLABa format eLABaMARC. Both formats (DC and eLABaMARC) are described using the XML language, thus they are submitted into the eLABa repositories as XML metadata files.

eLABa e-documents structure is formed by three types of files: main, additional and service. The names of all files are made of according to certain rules that allow those names to code.

The following are examples of main file names of e-documents that are stored in eLABa repositories, when those names are linked with the corresponding five collections of e-documents:

- BOOK – Book (available formats: PDF – for the internet; PDF – for publishing and XML – for long term preservation).
- ARTIC – Article (available formats: PDF – for the internet; PDF – for publishing and XML – for long term preservation).
- PAPER – Paper – (available formats: PDF – for the internet and XML – for long term preservation).

- REPOR – Report (available formats: PDF – for the internet; PDF – for publishing and XML – for long-term preservation).
- ETD – ETD document– (available formats: PDF – for the internet; PDF – for publishing and XML – for long term preservation).

The following is description of possible additional file names and contents of ETD documents that are accumulated in eLABa repositories:

- ABSTR – Abstract – the abstract of an ETD document (XML format for the internet).
- CONTEK – Contents – the contents of an ETD document provided at the beginning/end (XML format for the internet).
- COVER – Cover – the cover of an e-document (JPG format for the internet).
- DESCRK – Description – description of an ETD document (XML format for the internet)
- ESSPG – Important pages – extremely important pages of an ETD document that provide help in revealing its uniqueness, originality and the like (PDF format for the internet).
- RELAT – Related e-document – this is an additional e-document that is related to the ETD document being submitted, for instance: a review, an annex, a publication related directly and the like. (PDF format for the internet).

The following is description of service file names and file contents of all documents that are supposed to be submitted into eLABa repositories, including ETD documents as well:

- INFO – Submit information – information on the e-document that has been submitted into the eLABa repository, for instance, technical data specific to the e-document and to the supplier that has submitted the document and specific to the eLABa data administrator (XML format for the internet).
- COPYR – Copyright – copyright document, defining authors' rights for a certain type of e-documents (XML format, with formation of Creative Commons license required).
- POLICY – Authorization – data of e-document authorization (XML format in compliance with XACML rules).

Authentication and authorization

During stages of search and submission of e-documents, including ETD documents as well, protection against unauthorized access (hereinafter – eLABa safety) to eLABa system and to all data present in it, comes out to be an extremely important issue. Two components are employed to ensure the eLABa safety, i.e. authentication and authorization. The following introduces a more detailed description of authentication and authorization of eLABa system.

Authentication

Every authentication mechanism requires catalogue (database) of system users. In eLABa system specification a centralized users' catalogue is provided. To implement it, a popular open code software OpenLDAP (<http://www.openldap.org>) has been applied and data schemes have been created concerning eLABa users and data related to them that is stored in eLABa.

To conduct the eLABa users' authentication, a single portal for joining has been provided, which is usually called Single Sign-on (SSO). This allows working with all components of eLABa system after you have registered yourself in eLABa system, for instance, working with the components intended for e-document submitting into eLABa repositories (eLABa-ETD, eLABa-SPEC), with those for e-document search (ALEPH, Metalib, SFX) and the like.

SSO system is intended to use as a superstructure on the centralized users' catalogue, which has been implemented with OpenLDAP and as a unique entry point to all components of eLABa system. Thus, SSO provides eLABa with the following main advantages:

- the user needs only one account, with which he or she logs in only one time and may employ all eLABa components that have been described in SSO domain;
- increased eLABa safety, as user's name and password are not sent to the systems that are present in SSO domain (system components);
- more abstract authentication is guaranteed that allows, if needed, new users' catalogue systems to be included into eLABa, for instance, Microsoft Active Directory, Network Information System, Oracle RDBMS, etc.);
- it is easier to improve and supervise the eLABa components that have been already created and create new ones, as for new, supposed to be included into SSO domain eLABa components it shall not be required to implement the authentication mechanism;
- centralized user management is guaranteed, and the like.

When choosing a concrete SSO for eLABa system, several common SSO systems have been analyzed. One of them – *Ex Libris* company's Patron Directory Service (PDS) that has already been integrated into MetaLib and ALEPH systems functioning as eLABa components. Having conducted the more thorough analysis of SSO systems applied worldwide, it has been decided as well to test the SSO system *Shibboleth* (<http://shibboleth.internet2.edu/>) that has been created by consortium *Internet2* (<http://www.internet2.edu/>). It is a much more complex system, and it performs more functions in comparison to PDS. *Shibboleth* has been created on the basis of open standard SAML (<http://www.oasis-open.org/committees/security/>) that has been specified by OASIS (<http://www.oasis-open.org>) organization. The following are advantages of *Shibboleth* option in comparison to PDS:

- for authentication in more modern software *Fedora* versions in eLABa repositories, the *Shibboleth* system is intended to be employed;
- for authentication in MetaLib and ALEPH higher versions in other eLABa components, PDS with *Shibboleth* system are intended to be employed as well;
- *Shibboleth* is an open code software, therefore it shall be provided for free;
- MAMS (<https://mams.melcoe.mq.edu.au/zope/mams/>) project survey has shown, that *Shibboleth* system and SAML become more and more applicable to the process of Academic Institutional Repository creation worldwide.

Authorisation

It is foreseen in eLABa regulations (Dël Lietuvos...,2006), that metadata, which describes e-documents, including also ETD documents, shall be available without restriction to all users of the internet, and full-text documents may have the following availability conditions:

- a) freely available over the internet,
- b) available over the institution's intranet only,
- c) available to the registered eLABa users only,
- d) unavailable to eLABa users,
- e) full-text may be used for science, study and self-education purposes only (this condition should be indicated with *a*, *b* and *c* conditions).

Three availability conditions (*b*, *c* and *d*) have time restriction that shall be specified by the author (maximum 5 years), and after the time period a.m. has expired (and without author's objection), the condition "freely available over the internet" shall be assigned to the full-text.

Saved in service files, other information, which cannot be made public, shall be stored alongside with full-text and metadata in eLABa repositories, for instance, information on persons who have confirmed submitting the ETD document into eLABa repository, data on License Agreement and other official information. These data shall be used for eLABa administration only and shall not be made public.

In compliance with Model License Agreement, approved by the Minister of Education and Science of the Republic of Lithuania (Dël Lietuvos...,2006), every author, including authors of ETD documents, shall enter into License Agreements with copyright subjects, confirming their agreement to submit their works into eLABa repositories under conditions (above mentioned). One item of signed original License Agreement in print version shall be kept at the library of the institution. When e-signature is legitimated in eLABa system, license agreements shall be stored in eLABa repositories as official information.

eLABa repository software (based on *Fedora*) supports XACML (<http://www.oasis-open.org/committees/xacml/>) – data access control language, specified by OASIS (<http://www.oasis-open.org/>) organization, allowing flexible authorization of documents being stored in eLABa repositories. Therefore, the XACML language is applied widely to authorization policy implementation concerning all eLABa data. When applying the XACML language, availability rules are described (including temporary restrictions as well), implementing three availability restriction conditions (above mentioned) at full-texts of ETD documents. Other eLABa data authorization rules are described using the XACML language as well, for instance, when establishing conditions allowing eLABa system administrators of certain level to have access to all objects that eLABa repositories contain.

To describe availability conditions "available on the Institution's intranet only" and "available to the registered eLABa users only" applying the XACML language, IP addresses of corresponding institutions shall be needed (if needed, addresses of their branches as well) that may vary during development process concerning intranets of the institutions. Therefore, it shall not be purposeful to introduce directly into *Fedora* POLICY file of an e-object those IP addresses that may vary in the future, when the file has been prepared in accordance with XACML rules and is stored as the constituent of the e-object. To find solution to the present problem, some improvements to *Fedora* software have been performed on the basis of *Java Servlet Filters* (<http://java.sun.com/products/servlet/Filters.html>). Special program filters have been created allowing automated identification and authentication of the users according to their IP addresses indicated in external sources regarding *Fedora* e-objects.

Local search services

Not only ETD document metadata may represent search results, but a full-text (s) document as well. Additional files also, if they exist in an ETD e-object. Still a full-text e-document may not be available any time. Availability rights are established according to license agreement that has been entered into with the author. Availability conditions for full-text e-documents have been described in the previous chapter. To facilitate e-document search for internet users, special package tools for *Fedora* e-object search, mapping and navigation in collections have been created. These tools are applied to all search possibilities functioning in eLABa system and are described below: to eLABa portal, LABT BIS ALEPH OPAC and LVB portal.

eLABa portal is intended for on-line search of e-objects in eLABa repositories applying *Fedora GSearch* and *Lucene* software.

LABT BIS (ALEPH) is employed for implementation of OPAC functions of Lithuanian ETD union e-catalogue. The metadata of ETD documents are converted to MARC21 format and, applying OAI-PMH protocol, are transferred automatically from eLABa repositories into Lithuanian ETD union e-catalogue that functions in ALEPH system. Readers may run information search over the internet in this e-catalogue using the OPAC service of ALEPH system. To ensure the effective information search in the ETD union e-catalogue, special search indexes have been created in ALEPH system.

To facilitate scientific information search for readers of Lithuanian science and study community, the LVB portal has been created. The basis for LVB portal creation – MetaLib and SFX program products. The following main information resources have been integrated into the LVB portal (the number of resources that have been integrated into LVB in 2006 is provided in brackets):

- e-catalogues of Lithuanian and foreign libraries;
- freely available and subscribable bibliographic and full-text databases;
- bibliographic databases of scientific publications of Lithuanian Science and Study institutions;
- eLABa full-text databases;
- indexes of virtual catalogues and disciplines and other e-resources.

The ETD union e-catalogue that has been created in ALEPH system is integrated into LVB portal as single e-resource, applying ALEPH-X protocol. In the LVB portal, the readers may run the information search in several ways: in ETD union e-catalogue only, or in selected several e-resources simultaneously, combining them into single search query.

Dissemination of Lithuanian ETD's world wide

Dissemination of Lithuanian ETD documents through the internet is basically performed using the following protocols and technologies:

- HTTP;
- OAI-PMH;
- Z39.50;
- SRU/SRW.

Lithuanian ETD documents search and accessibility services provided to the readers via the internet, in their totality consist of mentioned above eLABa portal, LABT BIS ALEPH OPAC and LVB portal. Using the OAI-PMH protocol, ETD metadata from eLABa repositories are harvested into Lithuanian ETD union e-catalogue that has been created and functions in the ALEPH system. During harvesting process into Lithuanian ETD union e-catalogue, the ETD metadata are converted from eLABaMARC to MARC21 metadata standard.

Similarly, ETD metadata are submitted for harvesting into the NDLTD union e-catalogue. During harvesting into NDLTD, ETD metadata are converted from eLABaMARC to ETD-MS metadata standard. Via OAI-PMH protocol of Lithuanian ETD metadata, the service of harvesting is accessible publicly and may be applied to other search systems on the internet and union catalogues.

The service of Lithuanian ETD metadata transfer via Z39.50 protocol has been configured and made public in LABT BIS (ALEPH) environment; this service may be used in other world's bibliographic information search systems.

Currently, the SRU/SRW service for Lithuanian ETD metadata transfer is being designed, which has been foreseen in eLABa specification; the purpose of this service – to ensure closer cooperation between Lithuanian ETD and other information systems from Lithuania and abroad.

Conclusions

- The experience of creating Lithuanian ETD document database was very valuable while creating Lithuanian Academic Electronic Library (eLABa). Lithuanian ETD document database is integrated into eLABa and it became the constituent part of eLABa.
- Lithuanian ETD document database is being created in the context of eLABa based on the open source software (*Fedora* and other) and worldwide known systems acquired by LABT: ALEPH, MetaLib ir SFX. This allows to combine:
 - the reliability and efficiency in the areas where there are clear and settled processes and standards;
 - flexibility and innovative of IT solutions in the areas where new atypical decisions must be made.
- A lot of work was made while creating eLABa content model and metadata formats. That makes a good background ensuring search and spread of ETD's of good quality.
- Library system ALEPH carries out the main function of information search engine and this allows to ensure high quality and effectiveness of ETD information search (also as in eLABa). The portal of Lithuanian Virtual Library (LVB) supplies additional services of information search and presentation in the context of common resources of Lithuanian science and studies.
- Well known OAI-PMH and Z39.50 standards are used for spreading information all over the worldwide. Lithuanian ETD's metadata is harvested to NDLTD union catalogue. We have big expectations regarding the spreading of information of Lithuanian ETD through *Google* and *Google Scholar*.
- Authentication and authorization of access to ETD's (as eLABa in total) is considered to be very important. Common database of the users, LDAP and other well known standards are used for authentication. Authorization is solved using XACML and is based on the license agreements with the authors.

References

1. Dėl Lietuvos mokslo ir studijų elektroninių dokumentų informacinės sistemos (eLABa) įsteigimo (2006): Lietuvos Respublikos švietimo ir mokslo ministro 2006 m. liepos 14 d. įsakymas Nr. ISAK-1506. *Valstybės žinios*, 2006, Nr. 81-3212.
2. FRANCKEVIČIUS, Arūnas; KUČIUKAS, Vilius; ŠTREIMIKIS, Antanas; et al. (2006). The integration of Lithuanian ETD into the National Digital Library. In *ETD 2006* [online]: Unlocking Access: The 9th International Symposium on Electronic Theses and Dissertations, Quebec City, Quebec, Canada, June 7 - 10, 2006. Québec, [cited 2007-05-15], p. [1-10]. Available from World Wide Web: <http://www6.bibl.ulaval.ca:8080/etd2006/pages/papers/Poster_Streimikis.pdf>.
3. ŠTREIMIKIS, Antanas; KUČIUKAS, Vilius; ir TARGAMADŽĖ, Aleksandras. (2005). Lithuanian ETD Project. In *ETD 2005* [online]: Evolution through Discovery: 8th International Symposium on Electronic Theses & Dissertations, 28-30 September, 2005, Sydney, Australia. Sydney, [cited 2007-05-15], p. [1-6]. Available from World Wide Web:

- <<http://adt.caul.edu.au/etd2005/papers/127Streimikis.pdf>>.
4. ŠTREIMIKIS, Antanas; KUČIUKAS, Vilius; ŽURAUSKAS, Stanislovas; et al. (2007). Lietuvos akademinė e. biblioteka (eLABa). *Informacijos mokslai*, t. 40, p. 9-24.
 5. TARGAMADZĖ, Aleksandras; ŠTREIMIKIS, Antanas; KUČIUKAS, Vilius, et al. (2004). Lithuanian Electronic Theses and Dissertations (ETD) project as a Pilot Project for the Baltic States. In *Diversity in Unity: Baltic Libraries in the European Union: Proceedings of the 7th Congress of Baltic Librarians*, September 30–October 2, 2004, Jūrmala, Madona Region, Latvia. Riga, p. 96-102.
 6. TARGAMADZĖ, Aleksandras; ŠTREIMIKIS, Antanas; VALIULIS, Algirdas Vaclovas; et al. (2004). The creation of electronic text databases in Lithuanian higher education institutions. In *4th Global Congress on Engineering Education in collaboration with King Mongkut's University of Technology Thonburi Bangkok*, Thailand, 5-9 July, 2004, Melbourne, p. 219-222.
 7. VALIULIS, A.V.; TARGAMADZĖ Aleksandras; KUČIUKAS, Vilius; et al. (2006). E-publishing at Lithuanian higher education institutions. In *5th Global Congress on Engineering Education in collaboration with Polytechnic University Brooklyn*, New York, 17-21 July, 2006, Melbourne, p. 151-154.
 8. VILIŪNAS, Giedrius; ir GLOSIENĖ, Audronė. (2006). Institucinės talpyklos ir naujoji mokslinės komunikacijos infrastruktūrų sankloda. *Informacijos mokslai*, t. 36, p. 53-67.