

2021

Six Steps Toward Improving Discoverability of Ph.D. Dissertations

Ljubomir Paskali

University of Novi Sad, Serbia, ljubomir.paskali@gmail.com

Lidija Ivanović

University of Novi Sad, Serbia, lidija.ivanovic@uns.ac.rs

Dragan Ivanović

University of Novi Sad, Serbia, dragan.ivanovic@uns.ac.rs

Dušan Surla

University of Novi Sad, Serbia, surla@uns.ac.rs

Follow this and additional works at: <https://scholarworks.uaeu.ac.ae/j-etd>



Part of the [Collection Development and Management Commons](#), [International and Comparative Education Commons](#), [Scholarly Communication Commons](#), and the [Scholarship of Teaching and Learning Commons](#)

Recommended Citation

Paskali, Ljubomir; Ivanović, Lidija; Ivanović, Dragan; and Surla, Dušan (2021) "Six Steps Toward Improving Discoverability of Ph.D. Dissertations," *The Journal of Electronic Theses and Dissertations*: Vol. 1 , Article 2.

Available at: <https://scholarworks.uaeu.ac.ae/j-etd/vol1/iss1/2>

This Article is brought to you for free and open access by Scholarworks@UAEU. It has been accepted for inclusion in The Journal of Electronic Theses and Dissertations by an authorized editor of Scholarworks@UAEU. For more information, please contact j-etd@ndltd.org.

Six Steps Toward Improving Discoverability of Ph.D. Dissertations

Ljubomir Paskali,¹ Lidija Ivanović,¹ Dragan Ivanović,¹ and Dusan Surla¹

¹University of Novi Sad, Serbia

Abstract

This study proposes six steps that scientific institutions should undertake to increase the visibility and accessibility of their dissertations; these steps were implemented in the digital library of the University of Novi Sad. An analysis was conducted thereafter to evaluate the success, and it was found that the six steps and associated strategies were successful, with 400,000 downloads having been performed since the digital library was operational. Although this study presents six steps for improving electronic thesis and dissertation (ETD) discoverability in the digital library at the University of Novi Sad, these steps can easily be customized and implemented for ETD digital libraries at any scientific institution.

Keywords: digital library, Ph.D. dissertations, University of Novi Sad, discoverability, export, search

Six Steps Toward Improving Discoverability of Ph.D. Dissertations

Previous studies have examined the purpose, benefits, and drawbacks of storing dissertations in digital libraries compared with the classical approach of keeping hard copies in the physical library. This study thus examined the six steps used by the University of Novi Sad, Serbia, to develop a searchable e-theses digital library that could export the e-theses to other digital library networks such as the Networked Digital Library of Theses and Dissertations (NDLTD) union catalog, the Digital Access to Research Theses - Europe (DART-Europe), the National Repository of Dissertations in Serbia (NaRDuS), Open-AIRE+, and Open Access Theses and Dissertations (OATD).

The remainder of this paper is organized in the following manner: Section 2 illustrates the literature review related to this field, and Sections 3–8 describe the six following steps in the establishment of the searchable e-theses digital library:

- the development of digital dissertation libraries data model at the University of Novi Sad (Section 3);
- the design of a workflow specification for dissertation submission and defense (Section 4);
- the workflow specification implementation (Section 5);
- the acquisition of the metadata for defended e-theses to put into the digital library (Section 6);
- the implementation of the digital library search features (Section 7); and
- the export of the data to the digital library networks and repositories using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) (Section 8).

The results were then verified by analyzing the download logs stored in the digital library, and the benefits derived from the implemented digital dissertations library are discussed. Section 9 concludes the paper and outlines some directions for further research. Moreover, the implementation steps could be customized and applied to digital libraries of electronic thesis and dissertation (ETD) at any scientific institution to improve Ph.D. dissertation discoverability.

Literature Review

Fox (2001) claimed that the primary goal of the Networked Digital Library of Theses and Dissertations (NDLTD) organization was to assist future and contemporary doctoral students deal with science in accordance with the digital age in which they lived and that storing e-theses (theses and dissertations in digital form) in digital libraries improved the education of future doctoral students by facilitating easier access to previous knowledge and research, thus promoting cooperation between universities and researchers and improving dissertation writing. Copeland (2008) discussed the ways in which the availability of e-theses raised research visibility, thereby concluding that the increase in digital e-theses libraries and institutional repositories was remarkably improving research access.

Several previous studies have examined the advantages and disadvantages of digital e-theses libraries (Brown, 2010; Hall, 2002; Goodfellow et al., 2012); the analyses in these studies were mainly based on university students and teaching staff surveys. It was observed that compared to keeping the hard copies of dissertations in conventional libraries, the main advantage of digital e-theses libraries was in the ease of access to the scientific results. Consequently, special attention was paid to this aspect during the modeling and implementation of the digital library described in this study.

Therefore, the main objective of the University of Novi Sad digital defended Ph.D. dissertation library was to maximize the discoverability of the dissertations' scientific results, that is, to maximize the scientific results dissemination to improve access for future Ph.D. students. In addition to scientific institution digital dissertation publishing research policies based on open access licenses, the dissertation knowledge visibility could be increased using a digital library dissertation search application and by exporting dissertation metadata to other systems (Ivanović, 2014). To this end, dominant systems and standards were analyzed to determine the dissertation search solutions, develop implementation web service search guidelines (Ivanović et al., 2013), and enable the exporting of data from the digital library to as many other systems as possible to increase international dissertation visibility. The digital library reviewed in this paper consequently became a member of the NTLDT, DART-Europe, NaRDuS, OpeanAIRE+, and OATD networks. Although there were already existing open source digital library development platforms available, such as DSpace and EPrints, due to specific local requests from the faculty and the University of Novi Sad, it was decided to develop a specific digital library within the Current Research Information Systems of the University of Novi Sad (CRIS UNS system), which was a subsystem of the CRIS UNS research information system that has been under development since 2008.

Data Model Creation

The CRIS UNS system (www.cris.uns.ac.rs) is the University of Novi Sad's research information system; hence, the digital dissertation library had to be integrated into this system. The CRIS UNS was based on the Common European Research Information Format (CERIF) compatible data model, which was in turn based on the MARC 21 format, a set of codes and content designators defined for encoding machine-readable records (Ivanović et al., 2011) that

provides interoperability between the CERIF compatible research information systems and the MARC 21 compatible library information systems. This model includes all attributes and entities prescribed by the CERIF standard, preserves the existing relationship between the CERIF model entities, and enables the multilingual input of certain data. The CERIF compatible data model maps data related to the publications, patents, products, researchers, projects, events, and institutions from the CERIF model to MARC 21 bibliographic and authority records.

For example, the *cfResPubl* entity for the CERIF data model was meant for the storage of published scientific research results of any type, including dissertations. One instance of the CERIF entity *cfResPubl* and its linked entities (*ResultPublicationBiblNote*, *ResultPublicationTitle*, *ResultPublicationSubtitle*, *ResultPublicationNameAbbrev*, *ResultPublicationKeywords*, and *ResultPublicationAbstract*) are mapped to a single MARC 21 bibliographic record instance, which is more fully described in Ivanović et al. (2011). The MARC 21 record is stored in the CERIF compatible model as a string that represents the MARC 21 record as outlined in the ISO 2709 standard (http://www.iso.org/iso/catalogue_detail.htm?csnumber=7675), with the data fields and subfields MARC 21 record values being indexed using the *Apache Lucene* information retrieval library in the CRIS UNS system to enable bibliographic record searches of the metadata held in the data fields and subfields. The MARC 21 format is a rich data format that accords with the FRBR (*Records Requirements for Bibliographic Records* – www.ifla.org/en/publications/functional-requirements-for-bibliographic-records). As the metadata for all published scientific results, which includes theses and dissertations in the CRIS UNS model, are stored in the MARC 21 format, these results can be deduced using the library information prescribed by the rich MARC 21 bibliographic format.

The analysis of the dominant systems and standards that deal with Ph.D. dissertations (Ivanović, 2014) revealed that the most used protocol for the data exchanges supported by these systems was the OAI-PMH protocol, and the most used formats for metadata dissertation descriptions were the Dublin Core, ETD-MS, MARC 21, and CERIF. The digital dissertation library integration with the CRIS UNS system included defining the MARC 21 mapping to the Dublin Core and ETD-MS records and extending the CRIS UNS data model with metadata as prescribed by the University of Novi Sad Ph.D. studies rulebook. The introduced metadata prescribed by the University of Novi Sad rulebook are extended abstract, physical dissertation descriptions (chapters/pages/literature/tables/pictures/graphs/appendices), the Universal Decimal Classification(UDC), the scientific discipline, the scientific board dissertation acceptance date, the defense date, and the holding data. Therefore, the CRIS UNS system dissertation metadata set is a union of the metadata sets prescribed by CERIF, MARC 21, the Dublin Core, and the ETD-MS formats, which were extended by the metadata being used. Additionally, the metadata dissertation set was extended to support data exporting using the OAI-PMH protocol; it included the attribute *creator*, *dateOfCreation*, *modifier*, and *dateOfLastModification* to store information about which user created the record, when the record was created, which user changed the record, and when the record was changed, respectively. The date of creation and the last modification date were required to respond to the OAI-PMH *ListRecords* request requirements to export only those records that were processed in a specific period.

In digital library models, there are no special attributes for storing the paths to the digital dissertations as the paths are held within the MARC 21 record stored in accordance with the ISO2709 standard. In the subfield *u* of the MARC 2 856 data fields, 1 record holds the URL to the *Java Servlet* to download the digital dissertation. For example, this URL connects to the

servlet:

[http://www.cris.uns.ac.rs/DownloadFileServlet/Disertacija140069152600136.pdf?controlNumber=\(BISIS\)85348&fileName=140069152600136.pdf&id=2011](http://www.cris.uns.ac.rs/DownloadFileServlet/Disertacija140069152600136.pdf?controlNumber=(BISIS)85348&fileName=140069152600136.pdf&id=2011), with the digital thesis or

dissertation stored in the server's file system.

Notably, as the digital library model is based on a CERIF compatible data model, it contains all metadata prescribed by the CERIF standard, not just the dissertation metadata. Therefore, this model contains the metadata on the researchers, the institutions, the research projects, and the other relevant entities of scientific research activity prescribed by CERIF. Additionally, a system based on the digital library model can exchange metadata about the theses and dissertations using the OAI-PMH protocol in *CERIF*, *MARC 21*, *Dublin Core*, and the *ETD-MS* formats. As previously mentioned, in addition to storing metadata about Ph.D. dissertations, the proposed model can store metadata about other scientific results such as monographs, papers published in monographs, papers published in journals, and papers published in proceedings. Metadata about the published results can also be exchanged with other systems using the OAI-PMH protocol in the previously mentioned formats, and the data model contains all metadata from the former digital library of the University of Novi Sad, the DIGLIB UNS, which existed until 2012, the metadata for which were in accordance with the Ph.D. studies of the University of Novi Sad rulebook. Therefore, the data migration from the DIGLIB UNS system to the CRIS system was successfully performed, with the CRIS UNS system that is now the integrated research information system and institutional repository for the University of Novi Sad and the unique repository for all published scientific University of Novi Sad research, all of which is visible in other large systems that have similar protocols and formats.

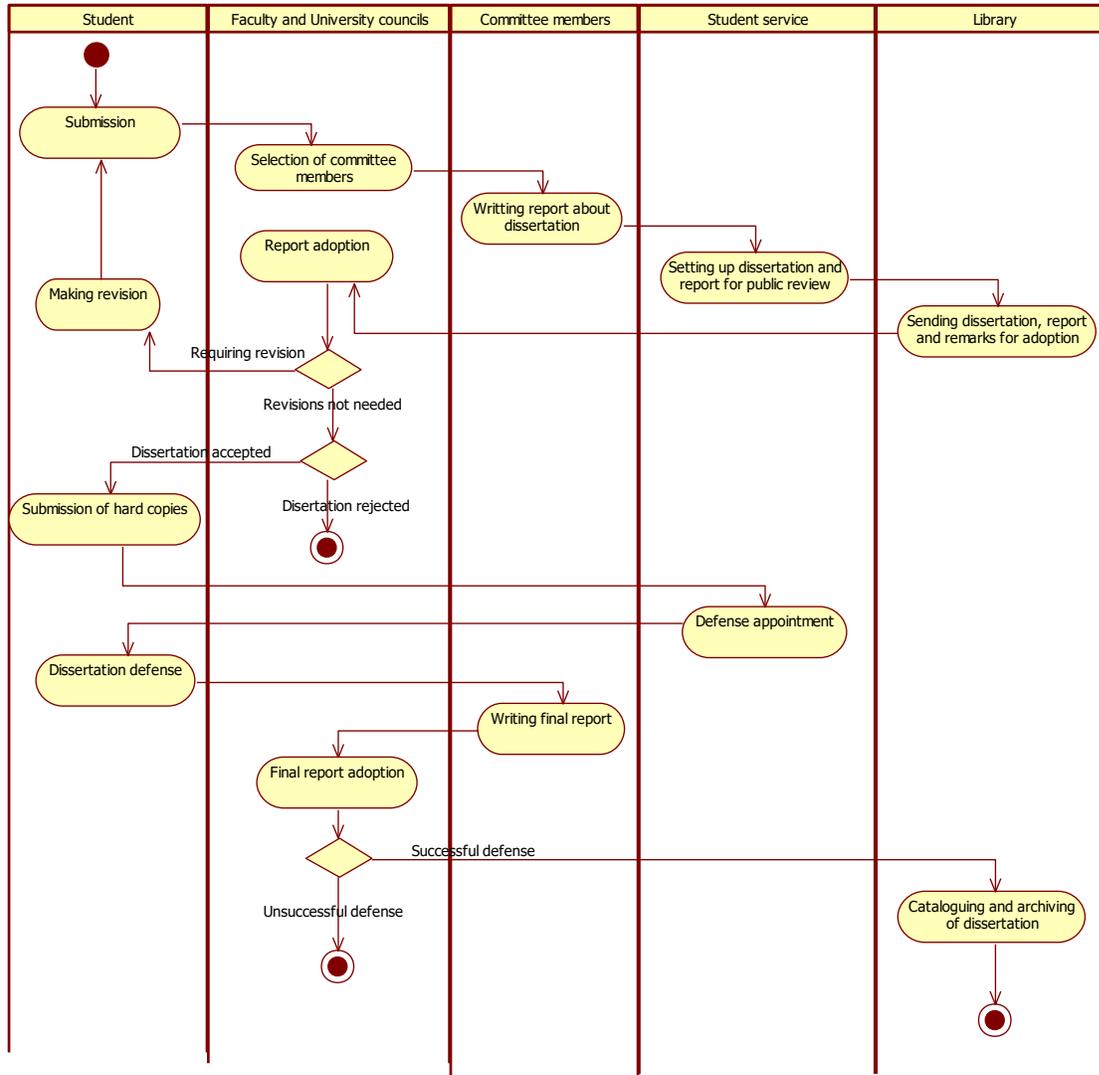
Submission and Dissertation Defense Process

The University of Novi Sad document entitled 'Ph.D. studies regulations' implemented in the digital library discussed in this study describes the current Ph.D. dissertation submission and defense process as well as the post-defense cataloging, as shown in Figure 1, which is a Unified Modeling Language (UML) activity diagram created using the StarUML tool.

Ph.D. studies are organized based on the 14 University of Novi Sad faculties. The dissertation submission and defense procedures have several well-defined steps that involve other students, the faculty and university councils, committee members, student services, and the library. The first step is to be given approval by the University Senate, which is a positive suitability evaluation of a Ph.D. candidate, a Ph.D. topic, and an advisor. Once approved, the Ph.D. candidate (student) can start their research. After the Ph.D. dissertation is complete and before it is finally accepted, it goes through faculty councils and the University Senate, evaluation and dissertation defense committees, Ph.D. student service officers, and the librarians employed at the faculty and Central University Library, as shown in Figure 1.

Figure 1

Dissertation Submission and Defense Process



The Ph.D. student first submits their dissertation (submission activity), with one hard copy being sent to student services. Since 2005, Ph.D. students must deliver a printed and electronic version (CD) of the dissertation to the student services in the faculty in which the dissertation is to be defended. After dissertation submission, student services send a request to the faculty council to establish an evaluation committee for the defense of the dissertation (selection of committee member activity). After the committee selection, one copy of the

dissertation is given to each committee member. Thereafter, the committee members write an evaluation report and submit the report to the student services at the relevant faculty (dissertation report writing activity), with the Ph.D. dissertation evaluation report and an electronic version of the dissertation being made publicly available on the official faculty and university website (<http://www.cris.uns.ac.rs/publicTheses.jsf>), (public review of the dissertation and report activity), which includes defining some basic metadata such as the author's name, the dissertation language, title, and the scientific field of the dissertation. At this point, the faculty student services start the dissertation cataloging process.

The dissertation is publicly available for 30 days, during which time remarks can be sent to the Central University Library. After 30 days, the Central University Library librarian collects and sends all remarks (if any) and a report (sending dissertation, report, and remarks for adoption activity) to the faculty and university boards, who then consider all the received documents and make a decision about whether to accept the committee report (report adoption activity), reject the report (with explanation), or request a review of the dissertation (request revision activity). If a review is required, the student has to review the dissertation based on the submitted remarks (make revision activity) and resubmit the dissertation, after which the process starts again. If the dissertation is rejected, the dissertation cannot be defended.

If the positive dissertation report is accepted, the student delivers a prescribed number of dissertation hard copies, the keyword documentation prescribed by the University, and a CD with an electronic version of the dissertation (submission of hard copies activity) to student services. Then, the dissertation defense is scheduled by student services (make defense appointment), a committee selected by the University Senate, and the accepted dissertation is publicly defended in front of the committee (dissertation defense). The committee then submits a

report for adoption by the faculty and university councils that states that: “The student defended the Ph.D. dissertation” or “The student did not defend the Ph.D. dissertation” (writing final report and final report adoption). If the report that concludes that the “student defended the Ph.D. dissertation” is adopted by the councils, a librarian completes the Ph.D. dissertation cataloging by entering all necessary metadata into the digital library (cataloging and archiving dissertation).

Some research has claimed that publishing research under open access licenses improves the impact and research citations in the world science community (Lawrence, 2001; Harnad & Brody, 2004; Anderson et al., 2001; Antelman 2004). In Serbia, the Ph.D. dissertation copyright belongs to the student who wrote the dissertation and not to the institution at which the dissertation was defended. Therefore, the legal department of the University of Novi Sad drew up a statement that allows students to publish their dissertations under one of the six Creative Commons license levels (hereinafter referred to as an open access statement) and a legal statement that states that the copyright is retained by the students (hereinafter referred to as a non-open access statement). When submitting the printed dissertation copies, the Ph.D. students (submission of hard copies activity) have to sign open access or non-open access statements related to the use of their Ph.D. dissertation. Since December 2013, the university management has recommended that Ph.D. students sign open access statements; however, signing non-open access statements was allowed until December 2014, after which all Ph.D. candidates had to sign open access statements in accordance with the new Serbian Higher Education Regulations. The statement contains information about the dissertation author; title; institution in which the dissertation was defended; and the defense date, which the Ph.D. author signs and dates. An open access statement authorizes the Central University Library to publish the dissertation under the open access license; therefore, the dissertations of the Ph.D. students who signed open access

statements are publicly available on the University of Novi Sad website, and the dissertations of the Ph.D. students who did not sign the open access statements are only available on the official website for 30 days during the public review.

Process Implementation

The digital Ph.D. dissertation library was implemented as an extension of the existing research management system web application at the University of Novi Sad, for which a set of 'best-of-breed' opensource libraries written in Java programming language were used. To ensure a *user-friendly* interface, Web 2.0 technologies were applied; to enable the addition of new language user interfaces without needing to change the application source code, the messages visible on the user interface are located in external files. The application user interface allows users to input data without needing to know the adopted CERIF or MARC 21 standards. The system can also be extended using a MARC 21 records editor, which gives the librarian access to the full Ph.D. dissertation cataloging features provided by MARC 21.

The University of Novi Sad research management system already had basic functionality for adding Ph.D. dissertation metadata before the integration of the digital library. However, as this basic set just included the metadata prescribed by the CERIF standard, it was extended to include metadata from the union of metadata for dissertations prescribed by the CERIF standard, Dublin Core, and the ETD-MS formats, and metadata prescribed by the University of Novi Sad Ph.D. studies rulebook. The application user interface was also changed to support these metadata additions and a component added to convert these metadata to accord with the MARC 21 format system data model. These changes were included in the existing system architecture, that is, the system architecture was not changed from the extensions to the supported Ph.D. dissertation metadata. However, system architecture changes were required to the *FileServer*

component to support the Ph.D. dissertations and supplemental material digital uploads. The digital Ph.D. dissertations, the open access statements, and the committee reports (supplemental materials) are all stored in the server's file system. Although this component allows for the storing of files with any extension, to accord with the needs of digital libraries, a restriction to PDF file format only was implemented by enabling the select only PDF files in the application user interface. The *FileServer* component is also used to store other digital scientific publications, such as papers published in journals, monographs, and conference proceedings. As web browsers do not have direct access to the server file system directory in which the digital files are stored, the files can be only reached using a *Java Servlet*. Although it is a recommendation, scientific publications do not have to be published under open access licenses; therefore, digital file access control was implemented within the source code of the *Java Servlet*. As mentioned, Ph.D. candidates could sign non-open access statements until December 2014, after which they had to sign open access statements. The *FileServer* component enables the long-term preservation of the knowledge stored in dissertations and is also able to extract the full dissertation texts using the Apache PDFBox library to enable a search of the digital library by dissertation content.

Before the integration of the digital library into the research management system, metadata dissertation input was only allowed by the dissertation authors and system administrators. However, after the integration, the system was extended to the librarian and the student service officers in the institutions in which they are affiliated, who after logging on to the system are now able to add and edit submitted or defended dissertations. This extension introduces a new type of user collaboration into the dissertation cataloging process, which

demands that dissertation records be locked for further changes when the dissertation is up for public review or when the dissertation cataloging is completed by a user.

After a Ph.D. candidate submits their dissertation (submission activity shown in Figure 1), dissertation defense committee members are selected by the university council (selection of committee members), which write a dissertation report (writing dissertation report). Then, the dissertation along with the report are forwarded to the Ph.D. student services where an officer inputs the dissertation and report in the digital library and sets it up for public review (setting up dissertation and report for public review) and a minimal set of dissertation metadata—*author's first and last name, dissertation language, title, research area, and the faculty where dissertation is going to be defended*—being automatically entered on the user's account. Figure 2 shows part of the form for the input of the dissertation metadata with the mandatory metadata marked with an asterix *.

Figure 2

Form for Dissertation Metadata Input

The screenshot displays a web form titled "Thesis/dissertation data" with a close button in the top right corner. The form is organized into a "Metadata" section. It contains several input fields, some of which are mandatory (indicated by an asterisk *):

- Author***: A dropdown menu with the text "Please choose personal name form".
- Type***: A dropdown menu with "PhD dissertation" selected.
- Language***: A dropdown menu with "Please select a language" selected.
- Original title***: A large text area with a "Title translations" button to its right.
- Subtitle**: A text area with a "Subtitle translations" button to its right.
- Alternative title**: A text area with an "Alternative title translations" button to its right.

Below the text areas is a rich text editor toolbar with various icons for text formatting (bold, italic, underline, strikethrough, text color, background color), alignment, and other editing functions. At the bottom of the form, there are two buttons: "<< Back" and "Next >>".

After the input of the digital dissertation and report metadata, the officer has a few options for the dissertation record, including the option to set up the dissertation for public review. By selecting this option, the metadata *date for setting up the public review* is automatically set on the current date, with the dissertation and committee report being available on <http://www.cris.uns.ac.rs/publicTheses.jsf>. All Ph.D. dissertation remarks are delivered to the Central University Library, and after 30 days, the system automatically removes the dissertation from the public review dissertations list and puts it on the finished public review dissertations list at the aforementioned website. Thereafter, the system sends a notification email to the Central University Library librarians that the public review is finished for the dissertation. The librarian then checks whether any remarks have been received during the public review period and sends a notification to the responsible body that the public review is finished and that further procedures can be actioned, attached to which are the dissertation, the report, and the list of remarks received during the dissertation public review (sending dissertation, report, and remarks for adoption). The dissertation and report are then adopted (or not) by the faculty and university councils (report adoption), and if the dissertation and the councils remarks are deemed valid, a notification on those remarks is sent to Ph.D. student services. The student service officer then notifies the Ph.D. student and their advisor that they need to revise the dissertation and resubmit, that is, they should start the process from the beginning. The faculty and university councils can also reject a dissertation. Both previously described situations are finished by locking the dissertation record for further changes and logging the information that the process is complete. These dissertations are not searchable through the digital library search form but are visible to the system administrators, librarians, and student service officers.

However, the most common case is that the public dissertation review is finished without any remarks, after which the faculty and university councils accept the report and dissertation, the Ph.D. student delivers a prescribed number of hard copies, the keyword documentation prescribed by the university, and a digital dissertation on CD to student services (submission of hard copies) and then signs an open access or non-open access statement about the use of the dissertation. The dissertation defense time is then scheduled (defense appointment), and the Ph.D. student defends their dissertation in front of the committee board (dissertation defense). If they successfully defend their dissertation, it is approved in a committee board report (writing final report), and the student is awarded a “doctor of science” in a certain scientific field. The faculty librarian then completes the dissertation cataloging by entering all remaining metadata and scanning and uploading the signed student license statement (Figure 3).

Figure 3

Example of the Upload for the Dissertation Documents

The screenshot shows a web application window titled "Thesis/dissertation data". It contains two main sections for document uploads:

- Public review - electronic version:**
 - Thesis/Dissertation: DisertacijaLidjajlvanovic.pdf (Remove button)
 - Committee board report: Lidja Izvestaj o oceni doktorske disertacije.pdf (Remove button)
- Final thesis/dissertation version:**
 - File: DisertacijaLidjajlvanovic.pdf (Remove button)
 - Scanned statement: img334.pdf (Remove button)
 - Statement type: Please, select an option (dropdown menu)
 - Compliance of printed and electronic version: img335.pdf (Remove button)

At the bottom of the window, there is a "Next >>" button.

The Central University Library librarian controls all metadata and digital files and locks the dissertation against further changes (see Figure 3 for the list of options including archiving). The dissertation cataloging is then complete (cataloging and archiving of dissertation), and the dissertation becomes searchable. If the student signed the open access statement, the dissertation is exported to other systems and can be downloaded for reading.

Data Acquisition

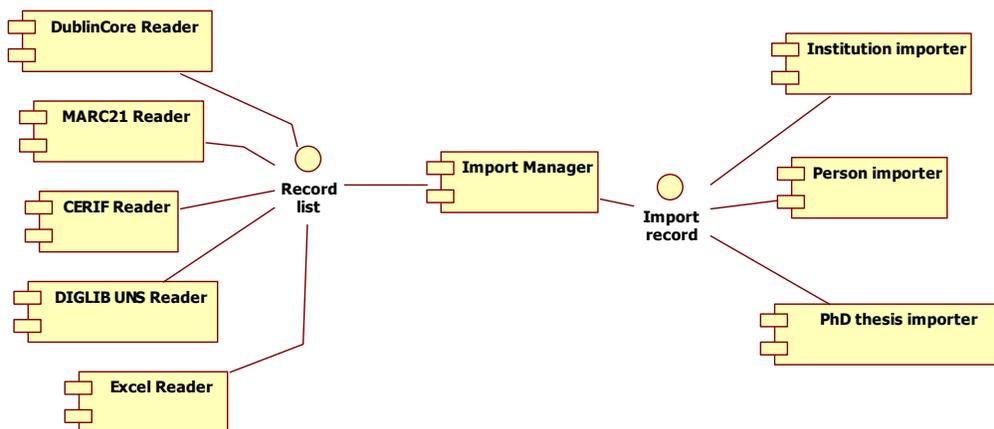
A digital library of all theses defended at the University of Novi Sad (bachelor theses, master theses, Ph.D. dissertations) called DIGLIB UNS existed before the digital library presented in this study was implemented. However, the DIGLIB UNS was not widely adopted by the faculties or the university and contained just a small part of all dissertations defended at the University of Novi Sad—about 200 theses in total. Besides the DIGLIB UNS, a register of all dissertations metadata had also been kept in an excel file (.xls) format, which had 3,500 University Novi Sad dissertations; however, there were no digital files with the full dissertation texts. The importation of the metadata to the digital library presented in this study was necessary to enable the metadata migration from DIGLIB UNS and from the excel record of dissertations to the new digital library (Ivanović & Surla, 2012), for which a new software module to import data was developed that could be easily extended with plugins to provide support for the importation of dissertations from various data sources as well as for the importation of other scientific research results. Therefore, this module was also useful for the CRIS UNS system, which, besides dissertations, stored other types of scientific research results and also enabled the human-controlled importation of data through an interactive user process that consolidated the data. When the same object already existed in the CRIS database, the module showed that there

is a potential duplicate object, with the user who started the importation then needing to decide what to do with the new record.

The module imports a complete network of objects related to the target objects that should be imported into CRIS. For example, besides the target objects (dissertations), it also imports the institutions in which the dissertations were defended and related persons, such as the authors, advisors, and board members. The module is then implemented using the Java platform, JSF technology, and the open source libraries written in Java. The module architecture is shown in Figure 4 using an UML components diagram created using the StarUML tool.

Figure 4

Component Diagram for the Importation Data Module



The left side of the diagram shows the list of supported input formats. The reader components transfer the input data from the various formats (Dublin Core, MARC 21, CERIF, DIGLIB UNS, and excel) to a list of objects representing the records that should be imported into the digital library.

The right side of the diagram connects the module with the rest of the system. The importer components perform the following functions:

- Check whether an object for importation (institution, person, or dissertation) already exists in the system and shows all objects stored in the system database in which the metadata has a similarity with the metadata of the object for importation; the similarity criteria are mentioned below:
 - the Levenshtein distance (edit distance) between two words must be less than or equal to the integer quotient, which is obtained by dividing the number of characters of the longer word by five;
 - if the shorter of the metadata (with fewer words) has more than five words, then the previous criteria are valid for 80% of the words; and
 - that Cyrillic and Latin scripts are equal.
- Store the object's metadata using the following algorithm:
 - if there is no dilemma as to whether the same object already exists in the system database, the importer component stores the object as a new record in the database;
 - whenever there is a dilemma as to whether the same object already exists in the database, the module shows a potential duplicate object, and the user who started the importation has to decide what to do with the object's metadata:
 - leave the metadata that already exist in the database;
 - overwrite the existing metadata with the new object metadata;
 - perform a merge of the metadata controlled by the user; or
 - store the object's metadata as a new record.

The `Import_Manager` component coordinates the whole data import process:

- invokes the appropriate reader component to create a list of objects (records) from an input metadata format; and
- invokes an appropriate import component that imports the object for each object from the list.

The data import module is extensible with a new reader component that provides support for the importation of a new metadata format to represent dissertations, with this extension not requiring the addition of a new importer component. The existing reader components can also be extended to provide support for the transformation of the input metadata format to an extended list of objects (not limited to institutions, persons, and dissertations), that is, the reader components can be extended to provide support for the importation of other scientific research results such as journal papers (papers – the target objects and journals and persons – the related objects), papers published in conference proceedings (papers – the target objects, conferences, proceedings, and persons – the related objects). If the data importation module is extended to support the importation of new types of scientific results (e.g., papers published in journals), a new importer component (e.g., *Journal_Importer*) needs to be added to check the existence and import new type of objects into the system database (e.g., journal).

Besides the previously described systematically supported data migration from various sources, in mid-2014, the University started acquiring the metadata and digital versions of the defended dissertations before the digital library was put into operation. A researcher at the University of Novi Sad can enter the metadata about their dissertation and upload the digital form of the dissertation through the system using their own user account. After the data is entered by the authors, a librarian checks the data accuracy using the dissertation hard copy held in the faculty library and then contacts the author to sign an open access or non-open access

statement for the use of the dissertation. After the author signs the statement, the librarian scans and uploads it to the digital library. Finally, the Central University Library's librarian checks the accuracy and completeness of dissertation metadata and locks the dissertation record against further changes, which makes the dissertation searchable (Section 7), and if the author has signed the open access statement, downloadable for reading and exportable to other systems using the OAI-PMH protocol (Section 8). The dissertations defended before the digital age and therefore not available in digital form are scanned by a faculty librarian using the automatic Robotic Book Scan (<http://www.qidenus.com/product/robotic/>) located in the Central University Library. At the beginning of 2020, there were more than 500 scanned dissertations, and the faculty librarians had begun entering data for dissertations defended at the University by authors not employed at the University. If a CD with a digital version of a dissertation exists, the digital file is uploaded to the system; otherwise, the dissertation hard copy is scanned. A strategy to convince the authors of those dissertations to sign open access statements is currently in progress.

Digital Library Search

We are living in the digital age, drowning in data, but starving for knowledge (Kielgast & Hubbard, 1997). As mentioned, the discoverability of Ph.D. dissertations can be improved using a web application to search the digital library, for which a web page was developed using the *JavaServer Faces* and *RichFaces* libraries (Ivanović et al., 2013).

The web page is open access, (<http://www.cris.uns.ac.rs/searchDissertations.jsf>), with the web page user interface being available in Serbian and English. The web page enables dissertation searches by metadata and by the dissertation full text extracted from the PDF versions of the dissertations. The web application business logic for the digital library search was

implemented without modifying the digital library software architecture and includes the processing and execution of user queries.

The search has three distinct modes:

- Making complex queries using the application user interface elements: Full text; Title; Abstract; Keywords; Author; Advisor, Board chair, Board member, with the Boolean operators AND, OR, AND NOT being used to create complex queries. The search results can be limited to a specific period anywhere between 1956 and 2020, and the list of results can be filtered by faculty.
- Searching researcher databases (authors/advisors/boards' members). The search can be conducted by the first and last names of the researchers. Besides the basic personal researcher data column (first and last name, affiliation, position, and title), there is also a link to retrieve the dissertation metadata and links for retrieving the dissertation metadata about the advisors, board chair, or board members.
- Making Lucene queries, which can be used by people with a basic knowledge of the Lucene query language.

The search is insensitive to Cyrillic or Latin scripts, which are the two scripts used in the Republic of Serbia. Additionally, the search is insensitive to various terms (synonyms) used in the six languages originating from Serbo–Croatian, the official language of former Yugoslavia: Serbian, Croatian, Slovenian, Bosnian, Montenegrin, and Macedonian.

Retrieved dissertations that match an executed query can be presented in several formats. The basic dissertation metadata set is displayed in the Harvard style reference by default; additional dissertation metadata and metadata about the dissertation are also in the MARC 21 format, the Dublin Core format, and the ETD-MS format, which can also be displayed through

the user interface. For all dissertations that have an uploaded digital dissertation and an open access statement signed by the author, there is a download link for the dissertation and information about the license. For each dissertation that has an uploaded digital dissertation but do not have an open access statement signed by the author, there is a link to send an email request to the Central University Library to obtain an electronic version of the dissertation. The Central University Library then communicates with the author of the dissertation and the user who sends the request to obtain the dissertation. This communication is performed out of the system using email and phone communication.

Data Exportation

As mentioned, the discoverability of Ph.D. dissertations could be improved by exporting dissertation metadata using OAI-PMH protocol standards to well-known international systems in accordance with the CERIF, Dublin Core, ETD-MS standards, which would avoid or reduce duplicate inputs on various platforms; increase metadata quality, reliability, and reusability; and increase the service quality based on these metadata.

The OAI-PMH protocol has been a very popular exchange data protocol, with several popular digital library and institutional repository networks, such as NDLTD (<http://search.ndltd.org/>), DART-Europe (<http://www.dart-europe.eu/basic-search.php>), NaRDuS (<http://nardus.mpn.gov.rs/>), OATD (<http://oatd.org/>), and OPENAire (<https://www.openaire.eu/>), being based on the OAI-PMH protocol. These networks improve the discoverability of digital documents stored in the databases of network members as these networks have a central node and hundreds of peripheral nodes all containing digital documents. To become a network member, a digital repository must enable the export of metadata using the OAI-PMH protocol. The central network node server periodically takes the metadata from digital documents stored in

the peripheral nodes, which, besides the title, author, and other basic metadata, contain an URL that enables the downloading of the digital document from the peripheral node storage. A central catalog created in this way can usually be searched using a purpose-developed application such as www.dart-europe.eu/basic-search.php. The OAI-PMH compliance of a potential network member is checked using specialized software developed by the network developers or using an open access web application to check for OAI-PMH compatibility such as *OAI Repository Explorer* (<http://re.cs.uct.ac.za/>).

From an analysis of available libraries for the OAI-PMH server-side implementation (Ivanović, 2014), a University of Novi Sad Digital Library exporting feature was implemented for which the *OAICat* library was selected, which is an open source, configurable, open-architecture library that exports data using the OAI-PMH protocol. The library comprises four components: *OAIHandler*, *Catalog*, *Factory*, and *Crosswalk*. The *OAIHandler* and *Factory* components were not changed for the OAI-PMH data export from the University of Novi Sad Digital Library; however, components of the *Catalog* and *Crosswalk* were extended:

Catalog – The *OAICat* library contains the *ExtendedJDBCOAICatalog* entity that supports the catalogs stored in relational database systems. *SQL Select* queries that the data requested from a database can be configured as shown in Listing 1. After executing a *SQL Select* query to fetch the data, the *ExtendedJDBCOAICatalog* entity invokes an *extendItem* method to fetch additional data from another data source. To export data from the digital library, a new entity *CRISUNSJDBCOAICatalog* was created to extend the *ExtendedJDBCOAICatalog* entity and override the *extendItem* method. The redefined method converts an ISO 2709 string representation of a MARC 21 record to an object representation of a MARC 21 record.

Crosswalk – This component was extended with four new entities designed to transform the MARC 21 object representation data to MARC 21 XML, Dublin Core XML, ETD-MS XML, and CERIF XML records. These entities were implemented as classes that extended the OAICat abstract class *Crosswalk* and overrode the *public String createMetadata(Object nativeItem)* method. The metadata format transformations represent the mapping implementation presented in Ivanović et al. (2012).

Besides the previously described OAICat library extensions, a configuration file to specify the parameters and *SQL Select* queries to export the dissertations was created (Listing 1).

Listing 1

OAICat Configuration

```
OAIHandler.styleSheet=/interoperability/oaicat.xml  
  
AbstractCatalog.secondsToLive=3600  
  
AbstractCatalog.oaiCatalogClassName=rs.ac.uns.ftn.informatika.bibliography.interoperability.oa  
ipmh.  
  
CRISUNSJDBCOAICatalog  
  
AbstractCatalog.recordFactoryClassName=rs.ac.uns.ftn.informatika.bibliography.interoperabilit  
y.oaipmh.JDBCRecordFactory  
  
AbstractCatalog.granularity=YYYY-MM-DD  
  
ExtendedJDBCOAICatalog.maxListSize=20  
  
ExtendedJDBCOAICatalog.jdbcDriverName=com.mysql.jdbc.Driver
```

```
ExtendedJDBCOAICatalog.jdbcURL=jdbc:mysql://localhost/cris?useUnicode=true&characterEncoding=UTF-8
```

```
ExtendedJDBCOAICatalog.jdbcLogin=****
```

```
ExtendedJDBCOAICatalog.jdbcPasswd=****
```

```
ExtendedJDBCOAICatalog.isPersistentConnection=true
```

```
ExtendedJDBCOAICatalog.identifierQuery=SELECT r.RECORDID, r.DATEMODIFIED, r.ARCHIVED FROM MARC21RECORD r, FILE_STORAGE f WHERE r.ARCHIVED in (100,2) AND r.RECORDID=f.RECORDID AND f.LICENSE is not null AND f.LICENSE not like 'Usage forbidden' AND f.LICENSE not like 'Temporary available' AND r.RECORDID<=\\i' AND r.RECORDID in (SELECT RECORDID FROM MARC21RECORD_CLASS WHERE CFCLASSID in (SELECT CFCLASSID1 FROM CFCLASS_CLASS WHERE CFCLASSSCHEMEID like 'oaipmhSetRelations' AND CFCLASSID like 'belongs to' AND CFCLASSID2 like 'theses'))
```

```
ExtendedJDBCOAICatalog.rangeQuery=SELECT r.RECORDID, r.RECORDSTRING, r.DATEMODIFIED, r.ARCHIVED FROM MARC21RECORD r, FILE_STORAGE f WHERE r.ARCHIVED in (100,2) AND r.RECORDID=f.RECORDID AND f.LICENSE is not null AND f.LICENSE not like 'Usage forbidden' AND f.LICENSE not like 'Temporary available' AND (r.DATEMODIFIED BETWEEN '\\f' AND '\\u') AND r.RECORDID in (SELECT RECORDID FROM MARC21RECORD_CLASS WHERE CFCLASSID in (SELECT CFCLASSID1 FROM CFCLASS_CLASS WHERE CFCLASSSCHEMEID like 'oaipmhSetRelations' AND CFCLASSID like 'belongs to' AND CFCLASSID2 like 'theses')) LIMIT \\a,\\b
```

```
ExtendedJDBCOAICatalog.rangeSetQuery=SELECT r.RECORDID, r.RECORDSTRING,  
r.DATEMODIFIED, r.ARCHIVED FROM MARC21RECORD r, MARC21RECORD_CLASS  
c, FILE_STORAGE f WHERE r.ARCHIVED in (100,2) AND r.RECORDID\=f.RECORDID  
AND f.LICENSE is not null AND f.LICENSE not like 'Usage forbidden' AND f.LICENSE not  
like 'Temporary available' AND r.RECORDID\=c.RECORDID AND (r.DATEMODIFIED  
BETWEEN '\f' AND '\u') AND c.CFCLASSID in (SELECT CFCLASSID1 FROM  
CFCLASS_CLASS WHERE CFCLASSSCHEMEID like 'oaipmhSetRelations' AND  
CFCLASSID like 'belongs to' AND CFCLASSSCHEMEID2 like 'oaipmhSet' AND  
CFCLASSID2 like '\s' AND CFCLASSID2 like 'theses') LIMIT \a,\b
```

```
ExtendedJDBCOAICatalog.setQuery=SELECT ct.CFCLASSID, ct.CFTERM FROM  
CFCLASSTERM ct WHERE ct.CFCLASSSCHEMEID like 'oaipmhSet' AND  
ct.CFLANGCODE like 'eng' AND ct.CFCLASSID like 'theses'
```

```
ExtendedJDBCOAICatalog.setSpecQuery=SELECT ct.CFCLASSID, ct.CFTERM FROM  
CFCLASSTERM ct WHERE ct.CFCLASSSCHEMEID like 'oaipmhSet' AND  
ct.CFLANGCODE like 'eng' AND ct.CFCLASSID\='\i'
```

```
ExtendedJDBCOAICatalog.setSpecItemLabel=ct.CFCLASSID
```

```
ExtendedJDBCOAICatalog.setSpecListLabel=ct.CFCLASSID
```

```
ExtendedJDBCOAICatalog.setNameLabel=ct.CFTERM
```

```
JDBCRecordFactory.repositoryIdentifier=CRISUNS
```

JDBCRecordFactory.identifierLabel=r.RECORDID

JDBCRecordFactory.datestampLabel=r.DATEMODIFIED

CRISUNSJDBCRecordFactory.deletedLabel=r.ARCHIVED

Identify.repositoryName=CRIS UNS

Identify.adminEmail=mailto:chenejac@[uns.ac.rs](mailto:chenejac@uns.ac.rs)

Identify.earliestDatestamp=2009-01-01

Identify.deletedRecord=persistent

Crosswalks.marc21=rs.ac.uns.ftn.informatika.bibliography.interoperability.oaipmh.MARC21JDBC2MARCXML

Crosswalks.oai_dc=rs.ac.uns.ftn.informatika.bibliography.interoperability.oaipmh.MARC21JDBC2oai_dc

Crosswalks.oai_etdms=rs.ac.uns.ftn.informatika.bibliography.interoperability.oaipmh.MARC21JDBC2oai_etdms

As mentioned, the University of Novi Sad Digital Library was operational in December 2013. When students submit a dissertation, they define a license for the use of the dissertation, with the two options that were available up to December 2014: publishing the dissertation under one of the six levels of the *Creative Commons license* and publishing the dissertation under a non-open access license. After December 2014, in accordance with the new Serbian Regulations on Higher Education, Ph.D. candidates now must sign an open access statement. The dissertation

metadata published under the open access licenses are exported using the OAI-PMH protocol to the networks shown in Table 1.

Table 1

Exported Data

Title	URL	Number of exported dissertations
NDLTD	http://search.ndltd.org/search.php?q=publis+her%3A%22University+of+Novi+Sad%22	1847
DART-Europe	http://www.dart-europe.eu/browse-results.php?institution=578	1871
NaRDuS	http://nardus.mpn.gov.rs/	1837
OATD	https://oatd.org/oatd/search?q=publisher%3A%28Novi%20AND%20Sad%29	1848
OPENAire +	https://explore.openaire.eu/search/dataprovider?datasourceId=openoar::ae1d2c2d957a01dcb3f3b39685cdb4fa	1825

Discussion: Benefits of the Digital Library

The University of Novi Sad Digital Ph.D. Dissertation Library encompasses the complete process from dissertation submission to defense to cataloging. This digital library was integrated

with the university research management system to create a unified central catalog of all university published scientific research output.

The digital library was operational in December 2013, and now, it includes all dissertations defended after 2013 and a repository of digital metadata library of all dissertations defended at the University of Novi Sad since its foundation in 1955, the information from which was migrated from the dissertation excel records and the DIGLIB UNS system using the metadata importation module. The University legal office developed an open access license statement to be signed by all dissertation authors for the publication of the dissertations. The University management also recommends that Ph.D. authors sign the open access statement for their and the university's benefit. The university management also initiated the scanning and uploading of dissertations that were not available in digital format and acquired open access license statements from the authors.

At the time of writing, the digital library had metadata on 5,984 dissertations, 2,305 of which have an uploaded digital dissertation form (ETD), with 1,908 of these being published under an open access license. More than 95% (790 of 822) of the dissertations defended after the digital library went live (December 2013) are published under open access licenses. To improve the discoverability of those dissertations, a web page was implemented to make searching the digital library (Section 7) easier and allow for the exporting of the associated metadata using the OAI-PMH protocol (Section 8).

The digital library logs messages regarding the Ph.D. dissertation digital file downloads, which includes information about the download date and time, the title, and other bibliographic information about the dissertation. The log messages also include information on how the user found the downloaded dissertation, that is, through a web search page of the digital library

(www.cris.uns.ac.rs/searchDissertations.jsf) or through another web portal at some of the other networks shown in Table 1. An example of a logged message is shown in Listing 2.

Listing 2

A Downloaded Message

```
[ INFO] 05.08.2014. 00:30:59 (FileManagerServlet:handleDownload)
Download file: (BISIS)77387| 230| disertacija.pdf| DART-Europe (https://www.google.rs/) |
93.87.188.22 | record: Molnar Jelena, Efekti odabranih unapredenih procesa oksidacije i
koagulacije na sadržaj prirodnih organskih materija u vodi, Prirodno-matematički fakultet u
Novom Sadu, Univerzitet u Novom Sadu, 2011
```

Table 2 shows the results of an analysis conducted on those messages.

Table 2

Digital Library Logs

Title	Number of downloads
Web page for search	324,242
NDLTD	9,030
DART-Europe	14,070
NaRDuS	38,819
OATD	11,124
OpenAIRE+	12,129

The analysis revealed that there had been around 400,000 downloads since the digital library went live, with more than 320,000 being from the web form searches described in Section 7. The Central University Library has also received many requests for print or digital versions of the non-open access Ph.D. dissertations—the metadata for which are shown in the web form search results. Many of these requests have come from researchers in neighboring countries (Croatia, Montenegro, Bosnia, and Herzegovina) because the search is independent of the Serbo-Croatian language dialects used in these countries. More than 85,000 dissertation downloads were invoked from the network portals, as shown in Table 2.

The analysis indicated that the main goal of this study had been achieved. The implementation of the searchable Ph.D. dissertation digital library, the promotion of dissertation publishing under open access licenses, and the exporting of the dissertation metadata using the OAI-PMH protocol to internationally well-known digital libraries networks have increased the discoverability of the knowledge stored in the dissertations defended at the University of Novi Sad and therefore has increased the rating of the University in the world scientific community.

Conclusion

This study described the implementation of the digital defended dissertations library at the University of Novi Sad; the digital library was created to increase the discoverability of the dissertations from the University of Novi Sad. The discoverability of the knowledge stored in these dissertations and their impact on the world scientific community are increased by open access licenses and a searchable digital library, which also exports metadata to well-known digital library networks. The implemented strategy was proven to be successful based on the analysis of messages logged by the digital library. The efforts made by the university management to promote the publication of the dissertations under open access licenses have been successful, as more than 95% of Ph.D. candidates since 2014 have published their dissertations under open access licenses. This new strategy was prescribed in December 2014, which included a rule that all future dissertations had to be published under one of the six open access Creative Commons license levels, which was also in accordance with the new Regulation on Higher Education prescribed by the Serbian government (“Sl. glasnik RS,” no. 76/2005, 100/2007 - 97/2008, 44/2010, 93/2012, 89/2013, 99/2014). This digital dissertation library implementation plan could be customized for any scientific institution to improve the Ph.D. dissertation discoverability.

Future work on the digital library includes data acquisition, with the final goal being to digitize all Ph.D. dissertations defended at the University of Novi Sad since its foundation in 1955 (5,984 so far) and convince authors to publish their dissertations under an open access license. Actions to achieve this goal have already been started, with university librarians having already digitized more than 500 dissertations using the Robotic Book Scan (<http://www.qidenus.com/product/robotic/>). The mining of user queries is also planned to improve the digital library services and the web search page user interface. The development of an automatic recommendation system based on user search history and profiles are in progress in collaboration with the University of Malta and the University of Cyprus.

References

- Anderson, K., Sack, J., Krauss, L., & O'Keefe, L. (2001). Publishing online-only peer-reviewed biomedical literature: Three years of citation, author perception, and usage experience. *Journal of Electronic Publishing*, 6(3). <https://doi.org/10.3998/3336451.0006.303>
- Antelman, K. (2004). Do open-access articles have a greater research impact? *College & Research Libraries*, 65(5), 372–382. <https://doi.org/10.5860/crl.65.5.372>
- Brown, J. (2010). *Literature review of research into attitudes towards electronic theses and dissertations (ETDs)*. UCL Eprints. <http://eprints.ucl.ac.uk/20424/>.
- Copeland, S. (2008). Electronic theses and dissertations: Promoting 'hidden' research. *Policy Futures in Education* 6(1), 87–96. <https://doi.org/10.2304/pfie.2008.6.1.87>
- Fox, E. A. (2001). *Overview of a guide for electronic theses and dissertations*. Digital Libraries and Virtual Workplaces: Important Initiatives for Latin America in the Information Age, Washington. <http://eprints.cs.vt.edu/archive/00000624/01/OASndltd14refs.pdf>
- Goodfellow, L. M., Macduff, C., Leslie, G., Copeland, S., Nolfi, D., & Blackwood, D. (2012). Nurse scholars' knowledge and use of electronic theses and dissertations. *International Nursing Review*. 59(4), 511–518. <https://doi.org/10.1111/j.1466-7657.2012.01008.x>
- Hall, S. (2002). Electronic theses and dissertations: enhancing scholarly communication and the graduate student experience. *Science & Technology Libraries*, 22(3-4), 51–58. https://doi.org/10.1300/J122v22n03_06
- Harnad, S., & Brody, T. (2004). Comparing the impact of open access (OA) vs. non-OA articles in the same journals. *D-Lib Magazine*, 10(6). <https://doi.org/10.1045/june2004-harnad> .
- Ivanović, L. (2014). *Modeling and implementation of digital library of PhD dissertations (PhD Thesis)*

[http://cris.uns.ac.rs/DownloadFileServlet/Disertacija140069152600136.pdf?controlNumber=\(BISIS\)85348&fileName=140069152600136.pdf&id=2011](http://cris.uns.ac.rs/DownloadFileServlet/Disertacija140069152600136.pdf?controlNumber=(BISIS)85348&fileName=140069152600136.pdf&id=2011).

Ivanović, L., Ivanović, D., & Surla, D. (2012). A data model of theses and dissertations compatible with CERIF, Dublin Core and EDT-MS. *Online Information Review*, 36(4), 568–586. <https://doi.org/10.1108/14684521211254068>

Ivanovic, L., Ivanovic, D., Surla, D., & Konjovic, Z. (2013). User interface of web application for searching PhD dissertations of the University of Novi Sad. *In Intelligent systems and informatics (SISY), 2013 IEEE 11th International Symposium on* (pp. 117–122). IEEE.

Ivanović, D., Surla, D., & Konjović, Z. (2011). CERIF compatible data model based on MARC-21 format. *The Electronic Library*, 29(1), 52–70. <https://doi.org/10.1108/02640471111111433> .

Ivanović, L., & Surla, D. (2012). A software module for import of theses and dissertations to CRISs, *Proceedings of the CRIS 2012 Conference, Prague, June 6–9, pp. 313–322*, http://www.eurocris.org/Uploads/Web%20pages/CRIS%202012%20-%20Prague/CRIS2012_34_full_paper.pdf

Kielgast, S., & Hubbard, B. A. (1997). Valor agregado à informação: da teoria à prática. *Ciência da informação*, 26(3). <https://doi.org/10.1590/S0100-19651997000300007>

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature*, 411(6837), 521. <https://doi.org/10.1038/35079151> .