Notes on Operations

Integration of a Research Management System and an OAI-PMH Compatible ETDs Repository at the University of Novi Sad, Republic of Serbia

Lidija Ivanović, Dragan Ivanović, and Dušan Surla

This paper discusses the extension of the Current Research Information System (CRIS) at the University of Novi Sad, Republic of Serbia, to incorporate electronic theses and dissertations (ETDs). Data describing ETDs is entered using a web application that enables researchers to input their own data through a webpage without knowing the standards on which the system is based. The ETDs repository can exchange data with CRIS institutional repositories and Networked Digital Library of Theses and Dissertations members. In this way, the international visibility of theses and dissertations created at the University of Novi Sad is enhanced without duplicating data entry in various systems. This approach has been verified and tested on a dataset of theses and dissertations at the University of Novi Sad.

Lidija Ivanović (lidija.ivanovic@pef.uns.ac.rs) is a teaching assistant, Faculty of Education, University of Novi Sad, Sombor, Republic of Serbia; Dragan Ivanović (chenejac@uns.ac.rs) is an assistant professor, Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Republic of Serbia; and Dušan Surla (surla@uns.ac.rs) is a professor ermeritus, Faculty of Sciences, University of Novi Sad, Novi Sad, Republic of Serbia.

Public access to theses and dissertations via the Internet is important for the development of a knowledge-based society. A knowledge-based society relies on the knowledge of its citizens to drive entrepreneurship, innovation, and vitality of that society's economy. A knowledge-based society possesses a community of scholars, researchers, research networks, engineers, technicians, and businesses engaged in research and the production of high-technology goods and provision of services. It forms a national innovation and production system, which is integrated into international networks of knowledge production. Its communication and information technological tools make vast amounts of human knowledge easily accessible. This paper describes a test bed project at the University of Novi Sad (UNS), Republic of Serbia, which aims to improve international access to UNS research. The approach described here can inform projects at other institutions.

One approach to achieving a knowledge-based society can be through depositing electronic dissertations and theses (ETDs) in a freely accessible digital repository. Assigning appropriate metadata to ETDs can improve discoverability by increasing their visibility. Furthermore, visibility of ETDs can be increased by putting the digital object or its descriptive metadata (or both) into systems containing theses and dissertations, such as digital libraries, research management systems, institutional repositories (IRs), the Networked Digital Library of Thesis and Dissertations (NDLTD), DART-Europe E-thesis portal, Digital Repository Infrastructure for European Research (DRIVER), and others. These initiatives

and related terms are explored in detail later in this paper.

Current Research Information System (CRIS) at the University of Novi Sad (UNS), Republic of Serbia, is a Common European Research Information Format (CERIF)–compatible research management system that has been in development since 2008 at UNS.[1] CERIF is "a comprehensive metadata standard and data exchange model that can be used for a very broad range of purposes involving the management and exchange of research data" developed by the European Organization for International Research Information (www.eurocris.org).[2] This system has been extended at UNS with a module for storing ETDs. The primary motivation for this expansion of CRIS UNS has been to increase the international visibility of theses and dissertations by UNS scholars. Increasing the visibility of ETDs can be achieved in the following ways:

- Exchanging data between the CRIS UNS system and other research management systems according to the CERIF standard.
- Exchanging data between the CRIS UNS system and IRs in Dublin Core (DC) format according to the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).
- Membership in the NDLTD network, i.e., exchanging data between CRIS UNS and other members of the NDLTD network in the Interoperability Metadata Standard for Electronic Theses and Dissertations (ETD-MS) format according to OAI-PMH protocol.

Another motivation for building this module was the creation of a unique system with all relevant data for scientific research activities. The system described here is a research management system integrated with an IR. The system architecture enables easy integration with library information systems, which are based on MARC 21format and also can hold metadata about ETDs.[3]

The goal of the integrated system, developed at UNS in accordance with CERIF, DC, ETD-MS, and OAI-PMH, is to avoid or reduce duplicated inputs on the two platforms and increase metadata quality, reliability, and reusability.

Literature Review

Scientific research is an important component of knowledge. Much contemporary scientific research along with its associated metadata is available in digital format via various means, such as digital libraries, research management systems, IRs, and publishers' platforms. Open access to scientific research enhances further development of science.[4] Maximizing the visibility of scientific research is essential for scientific advancement. Visibility can be enhanced by putting research into IRs that are OAI-PMH interoperable.[5] The OAI-PMH protocol was primarily developed as a low-barrier method for interoperability between metadata repositories and provides an interoperability framework based on metadata harvesting by defining two classes of participants: data providers that expose metadata and service providers that harvest metadata. The IR's metadata schema has a key role in increasing interoperability of the repository, i.e., maximizing visibility of theses and dissertations that are stored in the repository.[6]

A rich metadata schema enables establishing relations between various systems that contain scientific research. Collaboration between those systems has been discussed in recent years. According to Joint, sharing data between institution repositories and research management systems to avoid duplication of efforts is necessary.[7] Joint recommends a single point of entry to research articles regardless of whether it is through an institution repository or a research management system. He notes that three institutions (Glasgow University, Southampton University, and Kingston University) have already implemented this approach.

Krause suggests the creation of a virtual library that aims to enable users to gain integrated access to all relevant information in their special scientific field, irrespective of the location of metadata and digital form of documents.[8] A virtual library includes a single point for creation of queries that are sent to all systems that are part of the virtual library and integrates results retrieved from the systems.

"NARCIS: The Gateway to Dutch Scientific Information," by Dijk and colleagues, describes the National Academic Research and Collaborations Research System (NARCIS) portal (www.narcis.nl), which provides access to all scientific research information in the Netherlands.[9] That system is an integration of the Netherlands research management system and the Digital Academic Repositories in the Netherlands (DARENET).[10] Olivier describes collaboration between the research management system and the digital library at Pretoria University.[11]

The general objective of the CRIS-IR group is "to work out an optimal solution for the interoperability of Research Management Systems on the one hand and Institutional Repositories on the other, on a European scale, taking into account all relevant aspects."[12] The aim of the Current Research Information Systems and Open Access Repository (CRIS/OAR) interoperability project is to increase the interoperability between research management systems and open access repositories "by defining and proposing a metadata exchange format for publication information with an associate common vocabulary."[13] The aim of integrating systems that contain scientific research is to

maximize visibility of scientific research and avoid duplicating input of the same metadata in various systems. Many libraries worldwide store metadata in the MARC 21 format. Those libraries have electronic services that enable downloading metadata about bibliographic records related to scientific research, thus the interoperability of a repository of scientific research with those libraries is important for increasing visibility. Frequently, electronic services will enable metadata exchange in DC, but the fact that DC is not strict a standard can cause problem in metadata interoperability. However, the metadata defined by this format are a subset of metadata defined by MARC 21. When searching databases of scientific research, users can better express their information needs if the research is described in a richer set of metadata. A CERIF-compatible data model based on the MARC 21 format makes CRIS systems interoperable with library information systems.[14] In this model some CERIF data are stored in the MARC 21 format data model. As noted earlier, CERIF defines a data model that enables interoperability between CRIS implementations; MARC 21 is a standard for storing data for library systems. That model includes all entities and attributes of the CERIF data model and preserves the existing references between the CERIF data model entities. Furthermore, that model enables input of multilingual data prescribed by the CERIF standard. The MARC 21 format is rich in metadata and enables more detailed description of entities in CRIS systems. A MARC 21 record can store all metadata prescribed by DC and ETD-MS format.[15] An information system based on the CERIF-compatible data model can exchange data with other systems using XML documents (which have XML schemas prescribed by CERIF standard) and can exchange data with LIS based on the MARC 21formats and with IRs based on DC or ETD-MS format.

Context: CRIS UNS

CRIS UNS is a CERIF-compatible research management system under development since 2008 at UNS. The first phase of CRIS UNS development was the implementation of a system for entering metadata about published scientific research including papers published in journals, papers from scientific conferences, monographs, and papers published in monographs.

CRIS UNS is built on the CERIF-compatible data model based on the MARC 21 format described in the previous section. CRIS UNS was implemented as a web application based on "best-of-breed" open-source components written in Java.[16] The system has three-tier architecture. Three-tier architecture contains a client tier (the presentation logic, including simple control and user input validation), middle tier (the business processes logic and the data access), and data tier (the data server provides the business data).[17] Any web browser supporting HTML 4 and JavaScript can be used for application access.

The server side of the system is executed within the Apache Tomcat (http://tomcat.apache.org) application server. Apache Tomcat is an open-source software implementation of the Java Servlet and JavaServer Pages (JSP) technologies. A servlet is "used to extend the capabilities of servers that host applications accessed via a request-response programming model. Although servlets can respond to any type of request, they are commonly used to extend the applications hosted by Web servers."[18] JSP technology "provides a simplified, fast way to create dynamic web content."[19]

The presentation tier is developed using the JavaServer Faces (JSF) development environment (www.jcp.org/en/jsr/detail?id=252) and RichFaces (www.jboss.org/richfaces) library. JSF technology simplifies building user interfaces for JavaServer applications, and RichFaces is a library of Ajax-enabled components for JSF. The Apache Lucene (http://lucene.apache.org) library is an open-source information retrieval library written in Java and is used for indexing and searching text contents. Text indexing and query processing include a Cyrillic to Latin transliteration algorithm. All index entries are stored as Latin text, thus enabling the use of both scripts in searching. On the other hand, database contents hold information as it was entered by the user, preserving the correct script. Cyrillic to Latin transliteration is unambiguous. This means that every character of Cyrillic has an appropriate character in the Latin (or Roman) alphabet—and that every word written using Cyrillic can be unambiguously translated to a word using Latin characters. The MySQL (www.mysql.com) database management system is used for data preservation.[4] The system data model and architecture enable easy integration of the system with LIS and interoperability with other CERIF-compatible national CRIS systems.

Published results from the system are available to anonymous users via the Internet. Moreover, the system is in accordance with the CERIF standard and meets requirements prescribed by the Republic of Serbia Ministry of Science and Technological Development in the field of scientific research results evaluation. Therefore the system data model is extended with necessary entities.[20] The system is implemented as a web application that enables authors to input metadata about their own research without knowledge of the CERIF standard and the MARC 21 format.

Research Method

The first step in this project was analysis of various systems that contain metadata about theses and dissertations. The following are international initiatives:

- NDLTD (www.ndltd.org) is an international organization that aims to create a worldwide network of ETDs. Each digital repository that is a network member has to enable metadata exchange in the ETD-MS format (developed by DNLTD) in accordance with OAI-PMH.[21]
- DART-Europe E-Thesis Portal (www.dart-europe.eu/contributors/how.php) aims to collect details of the open access research theses stored in Europe's digital repositories (doctoral and master theses). It collects metadata in DC using OAI-PMH.
- DRIVER (www.driver-community.eu) is an international organization co-funded by the European Commission with the goal of creating a network of freely accessible digital repositories with content across all academic disciplines. Each digital repository that is a network member has to enable metadata exchange in DC in accordance with the OAI-PMH protocol.

In addition, many academic and research institutions and research communities may implement and manage the following approaches to collecting, preserving, accessing, and disseminating research:

- IRs are online systems that collect, preserve, and disseminate the intellectual output in digital form of an institution. IRs may use open-source software, such as DSpace (www.dspace.org) and Fedora (http://fedora-commons.org), or hosted, proprietary software, such as Digital Commons (http://digitalcommons.bepress.com) and SimpleDL (www.simpledl.com). Many IRs support the exchange of data in DC via OAI-PMH.
- A CRIS is a database of other

information system for storing data on current research (e.g., data about institutions, researchers, research projects, equipment, published results, etc.). The European Union encourages the development of national research management systems in accordance with the CERIF standard.[22] CERIF-compatible research management systems are called CRIS. Due to specific local or national requirements, CRIS systems are built on different modifications (or extensions) of CERIF data model.[22]

- A library information system (LIS) is a software system for acquiring, cataloging, and circulating library holdings. LIS are built on various bibliographic standards; most are based on MARC 21 formats.

Across these systems, different standards and protocols—CERIF, OAI-PMH, DC, ETD-MS, and MARC—enable interoperability.

After analysis was completed, a comprehensive metadata set was defined to develop a repository that is compatible with all previously mentioned systems. Then the authors extended the CRIS UNS data model to store all metadata about ETDs as well as the ETDs as digital objects. Finally, the authors expanded CRIS UNS with a module for storing ETDs along with associated metadata. An object-oriented method was used for the module modeling. Object-oriented modeling creates models using object-oriented diagrams (class diagram, sequence diagram, etc.), which is the starting point for implementing a system using object-oriented programming language. The modeling was carried out using the Sybase PowerDesigner tool that supports OMG's Unified Modeling Language (UML) 2.0 (www.omg.org/spec/UML/2.0). The module model can be obtained by contacting the authors. The implementation

was realized using "best-of-breed" open-source components written in Java. After the authors developed the module, it was verified and tested on EDTs by researchers in the Faculty of Sciences, UNS. After migration of the existing dataset containing ETDs along with associated metadata from the DIGLIB UNS system to the CRIS UNS system, UNS researchers verified the migrated data about their theses and dissertations and supplied additional data; the CRIS UNS metadata set is richer than the DIGLIB UNS metadata set. These steps are covered in detail in the following sections.

Data Model Definition

After analysis of various systems that contain metadata about theses and dissertations (NDLTD, DART-Europe E-thesis portal, DRIVER, IRs, CRISs, LIS, DIGLIB UNS), a comprehensive metadata set was defined to create a repository that is compatible with various ETDs systems. DIGLIB UNS is the IR at UNS and contains theses and dissertations from the university. This system allows input of metadata about theses and dissertations as required by the UNS rule book, which defines key words that all the university theses and dissertations must have. Table 1 presents the list of metadata elements selected for CRIS UNS and indicates their presence or absence in CERIF, DC, and ETF-MS. This metadata set unites metadata describing EDTs, drawing from all standards used in the DIGLIB UNS (diglib.uns.ac.rs).[23] The set of metadata about EDTs adopted for the CRIS-UNS system unites the metadata sets prescribed by CERIF, DC, and ETD-MS format, extended by metadata that are used in DIGLIB UNS to meet the needs of the UNS.

Data Model Extension

As already stated, the CRIS UNS data model holds data about scientific

research in MARC 21 format. MARC 21 records are stored using an attribute of the MARC 21 record entity that holds a string representing a MARC 21 record serialized according to the International Standards Organization (ISO) 2709 standard, which sets out the format for information exchange.[24] Upon serializing the MARC 21 record in an ISO 2709 string, the record is stored in the database and its contents are indexed using the Apache Lucene information retrieval library. MARC 21 records can be classified using the entity MARC 21 Record_Class: master thesis, PhD dissertation, and so on. Also, that entity can be used for the definition of the scientific field and scientific discipline of the research, such as mathematics, computer sciences, biology, information systems, and artificial intelligence. Using that entity, records can be divided in sets and the OAI-PMH "ListRecords" requirement, which mandates the ability to download only records that belong to a defined set, can be met.

The earlier CRIS UNS data model was extended by adding four attributes to the MARC 21 Record entity. These added attributes are *creator*, *dateOfCreation*, *modifier*, and *dateOfLastModification*. Date of creation and date of the last modification are necessary to meet all requirements prescribed by the OAI-PMH protocol; the OAI-PMH ListRecords request must be able to download only records that are processed in a certain period.

Furthermore, the previous CRIS UNS data model is extended by adding the *File_Storage* entity that is intended to hold data related to the digital form of theses or dissertations. Each instance of the *File_Storage* entity is connected to an instance of the *MARC 21 Record* entity that holds bibliographic metadata about the thesis or dissertations. The *uploader* attribute holds the e-mail address of the user who uploaded the digital content.

Table 1. Metadata about Theses and Dissertations Adopted for the CRIS-UNS System

CRIS-UNS	CERIF	Dublin Core	ETD-MS
author	+	+	+
advisor	-	-	+
chair	-	-	+
committee member	-	-	+
title	+	+	+
alternative title	-	-	+
subtitle	+	-	-
keywords	+	+	+
abstract	+	+	+
extended abstract	-	-	-
note	+	-	+
language	-	+	+
ISBN	+	-	-
physical description	+	-	-
UDC	-	-	-
publisher	+	+	+
publication date	+	+	+
record type	-	+	+
content format	-	+	+
URI	+	+	+
access rights	-	+	+
thesis type	+	-	-
name of author degree after defense	-	-	+
level of education	-	-	+
scientific field	-	-	+
scientific discipline	-	-	-
accepted by competent scientific institution on	-	-	-
institution	+	+	+
defended on	-	-	-
holding data	-	-	-

The attributes *fileName*, *mime*, and *length* store metadata describing the digital content that is stored in a folder of the file system of the CRIS UNS server. The folder is not directly accessible through the Internet, but digital contents can be downloaded using a Java servlet. In this way, access to digital content is controlled, i.e., the Java servlet controls who can download digital content.

Table 2 shows mappings of adopted metadata about theses and dissertations shown in table 1 to the extended CRIS UNS data model. The first column holds names of metadata and the second column holds location in MARC 21 bibliographic record. The first three characters of a MARC 21 record present a field code; the next two characters present the first and the second indicator, respectively; and the last character presents a subfield code. The character "#" indicates that indicator is not defined. The last column shows some notes

Table 2. Mappings of Metadata to Data Model

Metadata	MARC 21	Note
author	1001# a	All data about *authors/advisors/chair/committee members* are stored in a MARC 21 authority record; relation of thesis or dissertation with the authority record is established using the subfield 0 of data field 100/700 of MARC 21 bibliographic record. The subfield e of data field 100/700 holds relationship type: *author, mentor, thesis/dissertation defend board chair, thesis/dissertation defend board member.*
advisor	7001# a	
chair	7001# a	
committee member	7001# a	
title	24500 a	Translations of those metadata are stored in the field 880 as described in "CERIF Compatible Data Model Based on MARC 21 Format."*
alternative title	2460# a	
subtitle	24500 b	
keywords	653 ## a	
abstract	5203# a	
extended abstract	520 ## a	
note	500 ## a	
language	008	Language is stored using three letters from 35th to 37th character positions of the control field 008. Character positions starts from 0.
ISBN	020 ## a	
physical description	300 ##	Physical description is stored using subfields of the data field 300.
UDC	080 ## a	
publisher	260 ## b	The metadata holds a value *author's reprint* or name of the appropriate institution.
publication date	260 ## c	Year of publication are additionally stored in character positions 7–10 of the control field 008.
record type	LDR	Record type is stored in 6th character position of the leader of MARC 21 record. Character positions starts from 0.
content format	856 ## q	The metadata holds one of the following values: pdf, doc, docx, odt.
URL	856 ## u	The subfield holds the URL of a thesis or dissertation in digital form.
access rights	540 ## a	
thesis type	655 #4 a	Also stored using the *MARC 21Record_Class* entity of the CRIS UNS data model.
name of author degree after defense	502 ## a	Name of degree is prescribed at the institution where author defends his or her thesis or dissertation. For example: master of electrical engineering, doctor of technical sciences, etc.
level of education	502 ## b	The element holds level of education: bachelor, master, doctoral, post-doctoral, etc.
scientific field	65024 a	Also stored using the *MARC 21Record_Class* entity of the CRIS UNS data model.
scientific discipline	65014 a	
accepted by competent scientific institution on	502 ## g	The metadata are stored in the subfield g in the following format: 502 ## $gTheme of thesis or dissertation accepted on date.
institution	502 ## c	That subfield holds the name and address of the institution. All data about institutions are stored in a MARC 21 authority record, the relation of thesis or dissertation with the authority record is realized using entity *MARC 21Record_MARC 21Record.*
defended on	502 ## g	The metadata are stored in the subfield g in the following format: 502 ## $gThesis or dissertation defended on *date*.
holding data	852 ## a	

* Dragan Ivanović, Dušla Surla, and Zora Konjović, "CERIF Compatible Data Model Based on MARC 21 Format," *Electronic Library* 29, no. 1 (2011): 52–70.

about metadata and methods of their storing.

CRIS UNS Extension with ETDs

The next phase of the development of CRIS UNS was to extend it with a subsystem that enables uploading ETDs and inputting their metadata. The authors identified the basic

information requirements of this subsystem as the following:

- Uploading ETDs. The system supports pdf, doc, docx, and odt file formats. Furthermore, the system has to backup files and provides long-time preservation of those files.
- Migrating existing data from

DIGLIB UNS to the system.

- Entering all metadata about EDTs that that CERIF standard prescribes and all metadata that are necessary for exchange in accordance with the OAI-PMH protocol within NDLTD. User interface has to be as simple as possible so that it can be used by users without
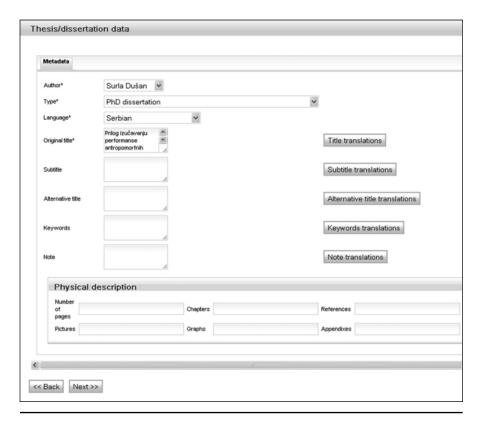
Figure 1. Form for Input of Metadata

the knowledge of standards and protocols.

- Exchanging metadata about EDTs with other CRIS systems. In this way, researchers from European countries using national CRIS systems can find EDTs from the CRIS UNS system.
- Exchanging metadata about EDTs in accordance with the OAI-PMH protocol. In this way, theses and dissertations from CRIS UNS can be visible through a various IRs as well as through web applications for searching the NDLTD Union Catalogue: SCIRUS ETD Search (www.ndltd.org/service providers/scirus-etd-search), VTLS Visualizer (www.vtls.com/products/visualizer), etc.

The system architecture was extended with a file server component that manages storing and downloading files from the server's file system. This component also is used to preserve digital contents of other scientific research, such as papers published in journals, monographs, and papers published in conference proceedings. This digital content is not freely accessible and access to those digital materials is controlled through the Java servlet. The file server component will be integrated with an open-source solution for long-term file preservation such as Lots of Copies Keep Stuff Safe (LOCKSS) (www.lockss .org). The file server component also extracts textual content from uploaded files using open-source Apache Tika library (http://tika.apache.org). After extraction, text goes through a Cyrillic to Latin transliteration algorithm and then is indexed using the Apache Lucene library. Query processing also includes a Cyrillic to Latin transliteration algorithm and

thus enables the use of both scripts (Cyrillic and Latin) in searching.

Furthermore, the system user interface is extended with user forms for uploading ETDs and entering metadata about ETDs. All textual user interface elements are stored in external files that facilitate the translation of the user interface to other languages. The first step is uploading the digital content, which uses a dialog that prompts the user to find the file to be added from his or her own computer. After uploading the digital content, the next step is input of the metadata listed in table 1. The form for input of metadata is shown in figure 1. Translations of multilingual metadata can be entered using this form and invoking (clicking on) the boxes to the right (e.g., Title translations, Subtitle translations, and so on).

All data about authors, advisors, chair, and committee members are stored in a MARC 21 authority record. The relation of a thesis or dissertation with the authority record is established using the subfield "0" of the MARC 21 record field 100/700. The subfield "0" contains the control number of the authority record that stores data about a researcher (thesis author, mentor, and so on). Subfield "e" of the field 100/700 holds the relationship type between a thesis and researcher (relation is established by subfield "0"), e.g., author, mentor, thesis or dissertation defense board chair, or thesis or dissertation defense board member. This approach to establishing relationships allows various reports to be generated, such as

- thesis and dissertations in which a researcher has been a mentor, thesis defense board chair, or thesis defense board member; and
- thesis and dissertations in which researchers from some departments have been a mentor, thesis defense board chair, or thesis defense board member.

Because some metadata are multilingual, information retrieval measures (precision, recall, and F-measure) are improved, i.e., visibility of ETDs are increased. Furthermore, visibility of ETDs is improved by using fuzzy search that is enabled through Apache Lucene library. Fuzzy search retrieves all theses and dissertations that meet a set of criteria that define similarity. For example, similarity criteria for two strings (string from a query and string from a thesis or dissertation title stored in the CRIS UNS database) are defined as follows:

- Each word in one string does not differ by more than two letters from a word in another string.
- If one string contains more than five words, the previous criterion is satisfied for at least 80 percent of the words.
- It is case insensitive and Cyrillic-Latin script insensitive (i.e., lower case and upper case are equal, as well as Cyrillic and Latin scripts).

Data Verification

This application was verified and tested on data about theses and dissertations of researchers employed at Faculty of Sciences, UNS. After migration of the existing dataset containing ETDs along with associated metadata from the DIGLIB UNS system to the CRIS UNS system, researchers from the University of Novi Sad verified and supplied migrated data about their theses and dissertations. The Faculty of Sciences employs more than 300 researchers and has written approximately 900 master theses and 500 PhD dissertations through 2011. The test set included metadata about all 1,400 theses and dissertations. Hard-copies of all 1,400 theses and dissertations can be found in the faculty library. In time of this writing, 200 of them also can be found in digital form. Transforming

the remaining 1,200 from hard-copy to ETDs by scanning is in progress. Researchers did not complain about the migrated data or the user interface. Adding theses and dissertations from the additional fourteen UNS faculties is also in progress. After this process is finished, an additional effort to consolidate data will be necessary; this will include such activities as removing duplicated items and consolidating scientific fields and disciplines.

Conclusion

This paper describes the implementation of a digital repository of EDTs within the CRIS UNS system. Metadata about theses and dissertations are stored in the MARC 21 bibliographic format. The implementation is based on open-source components. The system architecture allows an easy transition to other bibliographic standards and easy integration with LIS based on the adopted bibliographic standard.

The system can exchange ETDs metadata with other CRIS systems, IRs, the NDLTD network members, and LIS. Interoperability with previous stated systems maximizes visibility of ETDs from the repository without duplicate entry of ETDs metadata in various systems. Metadata are entered once, but metadata are stored in various systems across the Internet. High international visibility of theses and dissertations of researchers from University of Novi Sad enhances the further development of science and raises public awareness of UNS research.

The system for inputting of ETDs has been verified and tested on a dataset containing EDTs by researchers at Faculty of Sciences, UNS. The addition of ETDs from additional fourteen UNS faculties is in progress. After this process is finished, further effort to consolidate data will be necessary; this will include such activities as removing duplicated items and consolidating scientific fields and disciplines.

After this step, web services for data exchange will be made available for public access. Finally, an audit will be performed to assess whether the visibility of scientific research from UMS has increased after this repository implementation.

References and Notes

1. EuroCRIS, CERIF 2008—Final Release (1.2), www.eurocris.org/Index.php?page=CERIF2008&t=1 (accessed Nov. 16, 2011). CERIF 1.3 Release was available for preview until early December 2011, www.eurocris.org/Index.php?page=CERIF-1.3&t=1 (accessed Nov. 16, 2011).
2. CERIFy, What is the CERIFy Project? What is CERIF?, http://cerify.ukoln.ac.uk/node/196 (accessed Nov. 14, 2011).
3. Library of Congress, MARC Standards, MARC 21 Formats, www.loc.gov/marc/marcdocz.html (accessed Nov. 16, 2011).
4. Steve Lawrence, "Free Online Availability Substantially Increases a Paper's Impact," *Nature* 411 (May 2001): 477, www.nature.com/nature/debates/e-access/Articles/lawrence (accessed Nov. 14, 2011); Stevan Harnad and Tim Brody, "Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals," *D-Lib Magazine* 10, no. 6 (2004), www.dlib.org/dlib/june04/harnad/06harnad.html (accessed Aug. 22, 2011) ; Kristin Antelman, "Do Open-Access Articles Have a Greater Research Impact?" *College & Research Libraries* 65, no. 5 (Sept. 2004): 372–82.; Kent Anderson et al., "Publishing Online-Only Peer-Reviewed Biomedical Literature: Three Years of Citation, Author Perception, and Usage Experience," *Journal of Electronic Publishing* 6, no. 3 (Mar. 2001), http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0006.303 (accessed Aug. 22, 2011); Gunther Eysenbach, "Citation Advantage of Open Access Articles," *PLoS Biology* 4, no. 5 (May 2006): 692–98, www.plosbiology.org/article/info:doi/10.1371/journal

.pbio.0040157 (accessed Aug. 22, 2011).

5. Mohammad Hanief Bhat, "Interoperability of Open Access Repositories in Computer Science and IT—An Evaluation," *Library Hi Tech* 28, no. 1 (2010): 107–18.

6. Eun G. Park and Marc Richard, "Metadata Assessment in E-Theses and Dissertations of Canadian Institutional Repositories," *The Electronic Library* 29, no. 3 (2011): 394–407; Sevim McCutcheon et al., "Morphing Metadata: Maximizing Access to Electronic Theses and Dissertations," *Library Hi Tech* 26, no. 1 (2008): 41–57.

7. Nicholas Joint, "Current Research Information Systems, Open Access Repositories and Libraries: ANTAEUS," *Library Review* 57, no. 8 (2008): 570–75.

8. Jürgen Krause, "Current Research Information As Part of Digital Libraries and the Heterogeneity Problem Integrated Searches in the Context of Databases with Different Content Analyses," in *Gaining Insight from Research Information: 6th International Conference on Current Research Information Systems*, ed. Wolfgang Adamczak and Annemarie Nase, 21-31 (Kassel, Germany: Kassel University Press, 2002), www.uni-kassel.de/hrz/db4/extern/dbupress/publik/abstract_en.php?978-3-933146-84-7 (accessed Nov. 14, 2011).

9. Elly Dijk et al., "NARCIS: The Gateway to Dutch Scientific Information," in *Digital Spectrum: Integrating Technology and Culture: Proceedings of the 10th International Conference on Electronic Publishing held in Bansko, June 14–16, 2006*, ed. Bob Martens and Milena Dobreva, 49–57 (Sofia: FOI-COMMERCE, 2006), http://elpub.scix.net/data/works/att/233_elpub2006.content.pdf (accessed Nov. 15, 2011).

10. Astrid van Wesenbeeck, "Digital Academic Repositories in the Netherlands: Built with the DARE Program (2003-2006)" (presentation, Valencia, Spain, June 20, 2006), http://cde.uv.es/documents/2007-VANWESENBEECK.pdf (accessed Nov. 14, 2011).

11. Elsabé Olivier, "Open Scholarship and Research Reporting in Tandem: Creating More Value" (presentation, The African Digital Scholarship & Curation Conference, May 12–14, 2009, Pretoria, South Africa), www.ais.up.ac.za/digi/docs/olivier_paper.pdf (accessed Aug. 22, 2011).

12. EuroCRIS, Operation Work Plan for the CRIS-IR Task Group, www.eurocris.org/Index.php?page=CRIS-IR_workplan&t=1 (accessed Nov. 14, 2011).

13. KE: Knowledge Exchange, CRIS/OAR Project, www.knowledge-exchange.info/Default.aspx?ID=340 (accessed Nov. 15, 2011).

14. Dragan Ivanović, Dušla Surla, and Zora Konjović, "CERIF Compatible Data Model Based on MARC 21 Format," *Electronic Library* 29, no. 1 (2011): 52–70.

15. Lidija Ivanović, Dragan Ivanović, and Dušan Surla, "A Data Model of Theses and Dissertations Compatible with CERIF, Dublin Core, and ETD-MS," *Online Information Review* (forthcoming).

16. Dragan Ivanović et al., "A CERIF-Compatible Research Management System Based on the MARC 21 Format," *Program: Electronic Library & Information Systems* 44, no. 3 (2010): 229–51; Gordana Milosavljević et al., "Automated Construction of the User Interface for a CERIF-Compliant Research Management System," *Electronic Library* 29, no. 5 (2011): 565–88; Aleksandar Kovačević et al., "Automatic Extraction of Metadata from Scientific Publications for CRIS Systems," *Program: Electronic Library & Information Systems* 45, no. 4 (2011): 376–96.

17. Ariel Ortiz Ramires, "Three-Tier Architecture," *Linux Journal* 75 (July 1, 2000), www.linuxjournal.com/article/3508 (accessed Aug. 22, 2011).

18. The H2EE Tutorial, What Is a Servlet? http://java.sun.com/j2ee/tutorial/1_3-fcs/doc/Servlets2.html#75087 (accessed Nov. 19, 2011).

19. Oracle, JavaServer Pages Technology, www.oracle.com/technetwork/java/javaee/jsp/index.html (accessed Nov. 19, 2011).

20. Dragan Ivanović, Dušan Surla and Miloš Racković, "A CERIF Data Model Extension for Evaluation and Quantitative Expression of Scientific Research Results," *Scientometrics* 86, no. 1 (2011): 155–72.

21. Networked Digital Library of Theses and Dissertations, ETD-MS: An Interoperability Metadata Standard for Electronic Theses and Dissertations, version 1.00, rev. 2, www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.html (accessed Nov. 16, 2011); Open Archives Initiative, The Open Archives Initiative Protocol for Metadata Harvesting, Protocol version 2.0 of 2002-06014, www.openarchives.org/OAI/openarchivesprotocol.html (accessed Nov. 16, 2011).

22. EuroCRIS, CERIF 2008—Final Release 1.2.

23. Dušan Surla et al., "Overview of Implementation of the Networked Digital Library of Theses and Dissertations," *Infoteka* 5, no. 1–2 (2004): 75–86.

24. International Standards Organization, *International Standard: ISO 2709, Information and Documentation—Format for Information Exchange = Information et Documentation—Format por l'échange d'information*, 4th ed. (Geneva, Switzerland: ISO Copyright Office, 2008).