

SOFTWARE SYSTEMS FOR INCREASING AVAILABILITY OF SCIENTIFIC-RESEARCH OUTPUTS

Dragan Ivanović¹

Abstract. This paper describes the basic approaches for publishing scientific-research outputs using modern ICT technologies and web applications such as institutional repositories, the NDLTD, and CRIS. Also, the CRIS which has been developed at the University of Novi Sad is described. That system unifies the mentioned approaches and it is compatible with all of them.

AMS Mathematics Subject Classification (2010): 68U35

Key words and phrases: institutional repositories, NDLTD, research management systems, CRIS UNS

1. Introduction

One of the basic postulates of knowledge society is the availability of knowledge. Due to the development of ICT, it is evident that the process of learning and research has been changed in recent years. Earlier, knowledge was available to researchers usually in print form contrary to today when a lot of knowledge can be found in digital form using various web applications [30, 39, 31]. By 2010, published scientific-research outputs in digital form along with associated metadata are available via various web applications such as digital libraries, research management systems, institutional repositories, information systems of publishing activity (Springer, Emerald), etc. The exchange of metadata between those systems is necessary in order to increase the availability of scientific content. The relation between research management systems, library information systems and institutional repositories has been discussed recently [26]. The paper [10] describes the NARCIS (*National Academic Research and Collaborations Research System*) portal representing an access point to all scientific research information in the Netherlands. That system is an integration of the Netherlands research management system and DARENET (Digital Academic Repositories in the Netherlands). Collaboration between the system SICRIS (Slovenian research management system) and system COBISS.SI (Slovenian library system) is described in the paper [45]. Furthermore, the paper [38] describes the collaboration between the research management system and the digital library at Pretoria University. Also, there are similar collaborations between a research management system and a digital library at *Glasgow University, Southampton*

¹Faculty of Technical Sciences, Novi Sad, e-mail: chenejac@uns.ac.rs

University and Kingston University. Imperial College Digital Repository is a digital repository of the scientific-research outputs of *Imperial College London*. That system is described in [1]. The CRIS UNS system represents a unique digital repository of published results of researchers from the University of Novi Sad. That system is compatible with institutional repositories, members of the NDLTD (*Networked Digital Library of Thesis and Dissertations*) [36] and research management systems based on the CERIF data model.

The rest of the paper is structured as follows. Firstly, basic approaches for publishing scientific-research outputs using modern ICT technologies and web applications (institutional repositories, the NDLTD and research management systems) are described. After that, the CRIS UNS system is described and its main features are listed. The last section concludes the paper and outlines some directions for further investigations.

2. Institutional repositories

An institutional repository is a software system for collecting, storing and retrieving digital contents which represent outputs of researchers employed in a scientific institution. Those outputs can be papers published in the journal, papers presented at a scientific conference, master theses, PhD dissertations, etc. The existence of a scientific result in digital form in an institutional repository increases the availability of that result. Besides the ability of downloading digital archives, most of those systems enable downloading metadata about them in the Dublin Core format according to the OAI-PMH protocol. Today, there are many software systems that represent institutional repositories.

The EPrints [15] is the most popular institutional repository software platform. EPrints open source software is developed at the School of Electronics and Computer Science, University of Southampton, United Kingdom. It is a flexible system for building high quality repositories. It is recognized as the fastest and easiest way to set up repositories of research outputs. The main features of this system are:

- Archive Documents, Multimedia and Data
- Multi-Language Support
- OAI Compliant (Dublin Core format)

Research outputs can be grouped by Years, Subjects, Authors and Divisions. Figure 1 shows a web page of the demonstration repository which is created using EPrints software. The web page shows the research outputs that exist in the repository and that are published in 2005.

A separate web page shows details of selected output along with the URL which enables downloading of the digital archive.

Today, there are 269 known repositories with 519,952 records created using the EPrints software platform.

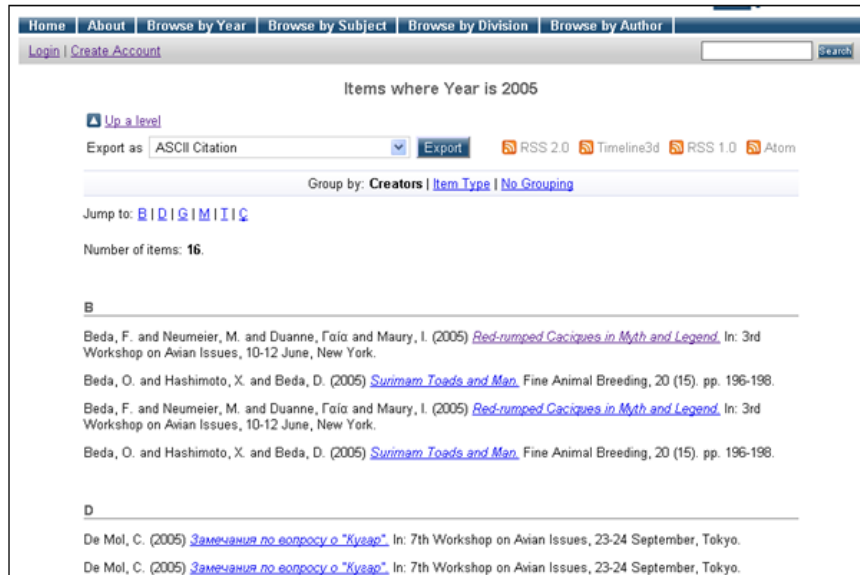


Figure 1: EPrints

3. NDLTD

The NDLTD (Networked Digital Library of Thesis and Dissertations) [36] is an international organization whose aim is to create a worldwide network of theses and dissertations in digital form. This organization promotes open-access to scientific knowledge which exists in theses and dissertations of researchers from all over the world. The network has one central node and hundreds of peripheral nodes that contain theses and dissertations of institutions that are members of the network. The server of the central node periodically downloads the metadata about theses and dissertations located in the network peripheral nodes. The metadata are title, author, other basic data and URL for downloading of the thesis or dissertation in digital form from the peripheral node. The download of metadata is realized according to the OAI-PMH protocol [37] and metadata are structured according to the EDT-MS format (an Interoperability Metadata Standard for Electronic Theses and Dissertations) [14]. In this way, the NDLTD Union Catalog is created and can be searched via certain web applications such as Scirus ETD Search [43], VTLS Visualizer [48], etc. Authors whose institutions are not members of the network can submit their theses and dissertations through a web application of the network central node.

Figure 2 shows the web page for query creation using the web application Scirus ETD Search. A separate web page shows results that match the executed query.

Currently, the NDLTD contains over a million theses and dissertations in digital form. Hundreds of scientific institutions are members of the network.

SCIRUS ETD Search
 This search service is provided courtesy of [EduX](#), based on data harvested from the Union Archive hosted by OCLC.

Basic Search
 Query: powered by **SCIRUS**

Advanced Search

All of the words in
 AND
 All of the words in
 AND
 All of the words in
 Only show results published between and

Only show results in

<input checked="" type="checkbox"/> All subject areas	<input type="checkbox"/> Agriculture and Biological
<input type="checkbox"/> Astronomy	<input type="checkbox"/> Chemistry and Chemical E
<input type="checkbox"/> Earth and Planetary Sciences	<input type="checkbox"/> Economics, Business and
<input type="checkbox"/> Engineering, Energy and Technology	<input type="checkbox"/> Environmental Sciences
<input type="checkbox"/> Languages and Linguistics	<input type="checkbox"/> Law
<input type="checkbox"/> Life Sciences	<input type="checkbox"/> Materials Sciences
<input type="checkbox"/> Mathematics	<input type="checkbox"/> Medicine
<input type="checkbox"/> Neuroscience	<input type="checkbox"/> Pharmacology

The complete document
 Title
 Author
 Date
 Keywords
 Publisher
 Abstract
 Type
 Format
 Language
 Rights
 Committee

Figure 2: Scirus ETD Search

4. Research Management Systems

Scientific institutions are major source of verified knowledge. That knowledge has different forms and research institutions make it available in two basic ways, by publishing results in scientific journals and scientific conference proceedings and by monitoring their own results within their internal research management systems and institutional repositories. This section presents standards relevant for research management systems and some well-known research management systems.

Research management systems are responsible for processing data about scientific research projects, organizations, researchers, research outputs, the necessary equipment for research and other relevant data for scientific research activities. The European Union encourages the development of national research management systems in accordance with the CERIF standard (*the Common European Research Information Format*) [7] because the European Union wants to achieve maximum competitiveness of Europe at all levels of research activity. CERIF-compliant research management systems are called CRIS (*Current Research Information System*).

CERIF data model consists of six entities' groups:

- *Core entities,*
- *Result entities,*
- *2nd level entities,*
- *Link entities,*
- *Semantic layer entities and*
- *Multilingual entities.*

Entities *Project*, *Person* and *Organization* are the CERIF *Core entities*. Entities *ResultPublication*, *ResultPatent* and *ResultProduct* belong to the CERIF *Result entities* and contain metadata about scientific-research outputs. The group *2nd level entities* consists of the entities dedicated for storage of relevant data for scientific-research activities, which do not belong to groups *Result entities* and *Core entities*. There are the following entities: *cfCite*, *cfCountry*, *cfCurrency*, *cfCV*, *cfEAddr*, *cfEquip*, *cfEvent*, *cfExpSkills*, *cfFacil*, *cfFundProg*, *cfLang*, *cfMetrics*, *cfPAddr*, *cfPrize*, *cfQual*, *cfSrv*. Those entities are linked with CERIF *Core* and *Result entities* through *Link entities*. *Link entities* have references to the two linked entities, time period in which relation between entities applied (attributes *startDate* and *endDate*), as well as the classification of relationship. The CERIF data model has a semantic layer which enables classification of entities and relations between entities according to some classification scheme. Example of classification of the relations between the *Person* entity and *ResultPublication* entity: author of the publication, editor of the publication, lecturer. Example of classification of *ResultPublication* entity: thesis, dissertation, paper published in the journal, monograph. The CERIF data model provides storage for certain data, such as title, abstract, keywords, research area in multiple languages. These data are stored using *Multilingual entities* group.

Because of the local and national specific requirements, CRIS systems are developed on different modifications of the CERIF data model.

Papers [2] and [25] describe a data model extension that was created in order to satisfy requirements of a research management system developed at the University of Bergen in Norway. Furthermore, the paper [2] describe a data model extension that was created in order to satisfy requirements of research management system that was developed on “Science and Technology Facilities Council”. Moreover, a data model extension that was created in order to satisfy requirements of the Information Society Technologies (IST) World portal is described in the papers [17, 27, 28, 29]. The papers [24] and [25] describe an extension that uses the formalised Dublin Core metadata [13] for description of scientific research outputs. Furthermore, the paper [21] describes a CERIF-compatible data model based on the MARC 21 format [32]. In that model, a part of the CERIF data model relating to published results is mapped to the MARC 21 data model. MARC 21 is a standard that defines format of bibliographic data.

By 2010, there have been many CRIS systems: IST World [20], Frida [18], Pure [40], HunCRIS [19], SICRIS [44], CRIS UNS [8].

IST World (*Information Society Technology World*) is a portal that provides access to scientific-research outputs from several countries. This site was developed within a FP6 (*Sixth Framework Programme*) project. The data model created for the purpose of this system is an extension of CERIF data model. Frida is an information system used by researchers at the University of Oslo in Norway. Pure is commercial software that can be installed and customized for the needs of scientific institutions. That software system is used by many universities such as the University of Helsinki and the University of Copenhagen. HunCRIS is Hungarian national research management system. That system is based on the CERIF data model. SICRIS (*Slovenian Current Research Information System*) is Slovenian national CRIS system. The SICRIS information system is being developed and maintained by the Institute of Information Science in Maribor and the SRA (*the Slovenian Research Agency*).

5. CRIS UNS

The CRIS UNS has been developing at the University of Novi Sad in the Republic of Serbia since 2008. The main feature of the CRIS UNS is the compatibility with institutional repositories, members of the NDLTD and CERIF-complaint CRIS systems. The development of the CRIS UNS system used the experience gained in developing the library information system BISIS [5]. The BISIS system has been developed since 1993 at the University of Novi Sad in the Republic of Serbia. The current version 4 is based on the XML technology. Within that version, an XML editor for cataloguing in the UNIMARC and MARC 21 format [11, 12] that uses the component for generating catalogue cards [41] and text server [34] is developed. The paper [6] describes an XML editor for retrieving bibliographic records in accordance with Z39.50 standard, the papers [46] and [33] describe a library circulation system, and the paper [47] deals with distributed library catalogues in BISIS. The subject matter of the paper [42] is the application of XML for conversion of bibliographic records into MARC 21 format. The conversion is implemented using XSLT technology. The papers [3] and [4] describe a model and an implementation of a system for entering data about published scientific-research outputs in the UNIMARC format.

The first phase was related to the development of a system for entering metadata about published scientific-research outputs in the form: papers published in journals, papers published in conferences proceedings, monographs, papers published in the monographs. The system is developed on CERIF-compatible data model based on the MARC 21 format, which is described in the paper [23]. The implementation of the system is the topic of papers [21] and [35]. The system is implemented as web application that enables authors to input metadata about their own research outputs without the knowledge of the CERIF standard and the MARC 21 format. Research outputs from the system are available to anonymous user via Internet. This system, in addition to international require-

ments prescribed by CERIF standard meets the requirements prescribed by the Ministry of Science and Technological Development of the Republic of Serbia in the domain of scientific results evaluation. Therefore, the data model of the system is extended with the necessary entities. This extension is described in the paper [22]. The main feature of the system architecture is the independence of the components for interaction with system users and for persisting and retrieving data from the bibliographic records database. This architecture allows an easy transition to other bibliographic standards and easy integration with library systems based on the adopted bibliographic standard. The system is verified and tested on data about published results of researchers employed at the Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad.

The following text describes a part of the CRIS UNS system related to entering data about papers published in journals. First of all, the author has to choose the journal in which the paper is published. If the journal is not yet registered in the system, the author can input the basic metadata about the journal. After choosing the journal, the system shows a list of all papers from that journal authored by the current user (see Figure 3)

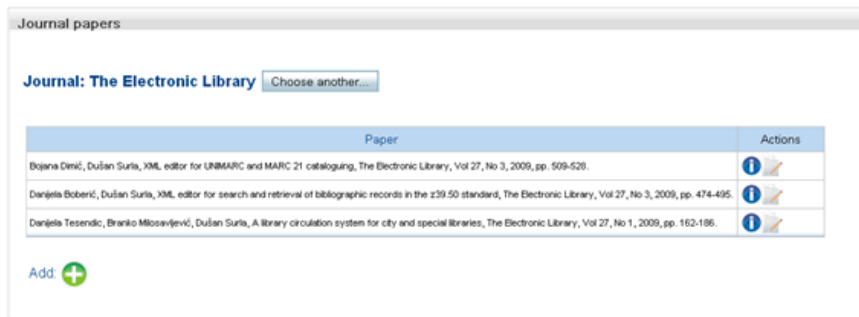


Figure 3: The display of the existing papers in a journal

The user can change the paper metadata (✎). Clicking the button (+) opens the form for adding a paper. Firstly, the user inputs the following metadata: paper title, subtitle, keywords, abstract, note, URI, start page, end page, journal publication volume, number, and year. Next, the user inputs the paper authors. If an author is not yet registered in the system, the logged-in user can input the basic metadata about that author (first name, last name, affiliation). After entering those metadata the system will check whether authors with similar metadata exist in the database. If authors with similar metadata exist, the form with a list of those authors is displayed (see Figure 4).

After specifying paper authors, the user saves paper metadata in the database. During the process of entering metadata about the paper, a preview of the paper reference is always displayed, enabling an easy control of the data entered.

The other parts of the implemented system, related to entering data about

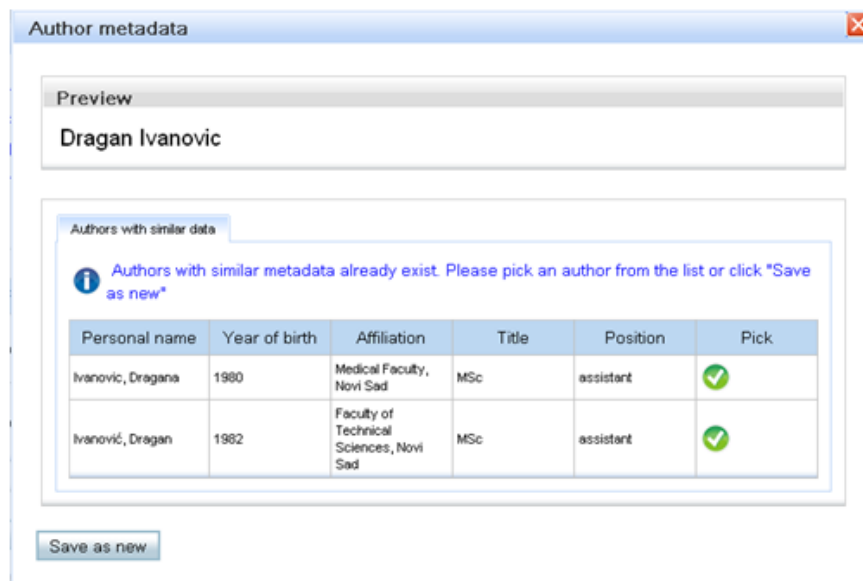


Figure 4: The list of authors with similar data

paper published in the conference proceedings, monographs and papers published in monographs are similar to the part of the system presented in the previous text.

The published results catalogued this way can be used for:

- evaluation of publications according to various criteria,
- generating various reports,
- downloading of bibliographic records formatted according to various standards.

The next phase of the development of the CRIS UNS includes the integration of the institutional repository of theses and dissertations DIGLIB UNS [9] in the CRIS UNS, and implementation of subsystems within the CRIS UNS that enables input of theses and dissertations in digital form as well as their associated metadata. The defined set of metadata includes all metadata specified by the CERIF data model, all metadata necessary for the metadata exchange in the Dublin Core and the EST-MS format in accordance with the OAI-PMH protocol and all metadata specified by the Regulations of the University of Novi Sad, which prescribes key words documentation for theses and dissertations. In this way, the CRIS UNS will be able to exchange metadata about scientific-research outputs with CERIF-compliant systems, exchange metadata with institutional repositories in the Dublin Core format according to OAI-PMH protocol and

become a member of the NDLTD that requires metadata exchanging in EDT - MS format according to the OAI-PMH protocol.

Other parts of the system relating to projects, institutions, researchers, equipment, prizes and other entities of the CERIF data model will be the subject of further investigations.

6. Conclusions

By 2010, a lot of published scientific-research outputs in digital form along with associated metadata have been available via various web systems such as institutional repositories, NDLTD and CRIS systems. The aim of the development of these systems is to increase the availability of published scientific-research outputs. Various protocols for exchanging metadata between those systems have been developed in recent years. These protocols enable the interoperability of these systems. Also, the integration of various web systems that contain scientific-research outputs has been done in the recent years. This paper describes the CRIS UNS that has been developed at the University of Novi Sad since 2008. The system is interoperable with other CRIS systems and can be integrated with digital repository of the published outputs which can exchange metadata with institutional repositories and members of the NDLTD in accordance with the OAI-PMH protocol.

The future expansions of the CRIS UNS will be in accordance with the recommendations of euroCRIS organization [16].

References

- [1] Afshari, F., Jones, R., Developing an integrated institutional repository at Imperial College London. Program: Electronic library and information systems, Vol. 41 No. 4 (2007), 338-352.
- [2] Asserson, A., Jeffery, K., Lopatenko, A., CERIF: Past, Present and Future: An Overview. In: Proceedings of the 6th International Conference on Current Research Information Systems, University of Kassel, August 29 - 31, 2002, pp. 33-40, 2002.
- [3] Belić, K., Surla, D., Model of a User Friendly System for Library Cataloguing. Computer Science and Information Systems, Vol. 5 No. 1 (2008), 61-85.
- [4] Belić, K. and Surla, D., User Friendly Web Application for Bibliographic Material Processing. The Electronic Library, Vol. 26 No. 3 (2008), 400-410.
- [5] BISIS, <http://www.bisis.uns.ac.rs/> (accessed 4 September 2010)
- [6] Boberić, D., Surla, D., XML Editor for Search and Retrieval of Bibliographic Records in the Z39.50 Standard. The Electronic Library, Vol. 27 No. 3 (2009), 474-495.
- [7] CERIF, <http://www.eurocris.org/cerif/introduction/> (accessed 4 September 2010)
- [8] CRIS UNS, <http://www.cris.uns.ac.rs/> (accessed 4 September 2010)
- [9] DIGLIB UNS, <http://diglib.uns.ac.rs/> (accessed 4 September 2010)

- [10] Dijk, E., Baars, Chr., Hogenaar, A. and Meel, M. van, NARCIS: The Gateway to Dutch Scientific Information. .Elpub 2006 Conference, *Bansko, June 14-16, 2006* (2006), 49-57.
- [11] Dimić, B., Surla, D., XML Editor for UNIMARC and MARC21 cataloguing. *The Electronic Library*, Vol. 27 No. 3 (2009), 509-528.
- [12] Dimić, B., Milosavljević, B. and Surla, D., XML schema for UNIMARC and MARC 21 formats. *The Electronic Library*, Vol. 28 No. 2 (2010), 245-262.
- [13] Dublin Core, <http://dublincore.org/> (accessed 4 September 2010)
- [14] EDT-MS, <http://www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.html> (accessed 4 September 2010)
- [15] EPrints, <http://www.eprints.org/> (accessed 4 September 2010)
- [16] euroCRIS, <http://www.eurocris.org/> (accessed 4 September 2010)
- [17] Ferlež, J. (2005), "Public IST World Deliverable 1.3 – Data Model for Representation of Expertise", available at: http://ist-world.dfki.de/downloads/deliverables/ISTWorld_D1.3_DataModelForRepresentationOfExpertise.pdf (accessed 4 September 2010)
- [18] Frida, <http://www.uio.no/english/research/frida/> (accessed 4 September 2010)
- [19] HunCris, <http://nkr.info.omikk.bme.hu/HunCRIS{ }eng.htm> (accessed 4 September 2010)
- [20] IST World, <http://www.ist-world.org/> (accessed 4 September 2010)
- [21] Ivanović, D., Milosavljević, G., Milosavljević, B., Surla, D., A CERIF-compatible research management system based on the MARC 21 format. *Program: Electronic library and information systems*, Vol. 44 No. 3 (2010), 229-251
- [22] Ivanović, D., Surla, D., Racković, M., A CERIF data model extension for evaluation and quantitative expression of scientific research results. *Scientometrics*, DOI: 10.1007/s11192-010-0228-2 (in press)
- [23] Ivanović, D., Surla, D., Konjović, Z., CERIF compatible data model based on MARC 21 format. *The Electronic Library*, Vol. 29, No. 1 (2011) (in press)
- [24] Jeffery, K., An architecture for grey literature in a R&D context. *The International Journal on Grey Literature*, Vol. 1 No. 2 (2000), 64-72.
- [25] Jeffery, K., Asserson, A., Revheim, J., CRIS, Grey Literature and the Knowledge Society. *Proceedings CRIS-2000*, Helsinki, 2000. Available at: https://www.researchgate.net/publication/2376388_CRIS_Grey_Literature_and_the_Knowledge_Society (accessed 4 September 2010)
- [26] Joint, N., Current research information systems, open access repositories and libraries. *Library Review*, Vol. 57 No 8 (2008), 570-575
- [27] Jörg, B., Ferlež, J., Grabczewski, E., Public IST World Deliverable 1.2 – Data Model for Knowledge Organisation. 2005. Available at: http://ist-world.dfki.de/downloads/deliverables/ISTWorld_D1.2_DataModelForKnowledgeOrganisation.pdf (accessed 4 September 2010)
- [28] Jörg, B., Ferlež, J., Grabczewski, E., Jermol, M., IST World: European RTD Information and Service Portal. 8th International Conference on Current Research

- Information Systems: Enabling Interaction and Quality: Beyond the Hanseatic League (CRIS 2006), Bergen, Norway, 2006. Available at: <http://epubs.cclrc.ac.uk/bitstream/905/ISTWorld01.pdf> (accessed 4 September 2010)
- [29] Kiryakov, A., Grabczewski, E., Ferlež, J., Uszkoreit, H., Jörg, B., Public IST World Deliverable 1.1 – Definition of the Central Data Structure. 2005. available at: http://ist-world.dfki.de/downloads/deliverables/ISTWorld_D1.1_CentralDataStructure.pdf (accessed 4 September 2010)
- [30] Lewandowski D., Google Scholar as a tool for discovering journal articles in library and information science. *Online Information Review*, Vol. 34 No. 2 (2010), 250-262
- [31] Luyt, B., Goh D. and Lee C.S., Searching locally: a comparison of Yehey! and Google. *Online Information Review*, Vol. 33 No. 3 (2009), 499 – 510.
- [32] MARC 21, <http://www.loc.gov/marc/> (accessed 4 September 2010)
- [33] Milosavljević, B., Tešendić, D., Software Architecture of Distributed Client/Server Library Circulation. *The Electronic Library*, Vol. 28 No. 2 (2010), 286-299.
- [34] Milosavljević, B., Boberić, D., Surla, D., Retrieval of Bibliographic Records Using Apache Lucene. *The Electronic Library*, Vol. 28 No. 4 (2010), 525-539.
- [35] Milosavljević, G., Ivanović, D., Surla, D., Milosavljević, B., Automated Construction of the User Interface for a CERIF-Compliant Research Management System. *The Electronic Library*, Vol. 29 No. 5 (2011) (in press)
- [36] Networked Digital Library of Thesis and Dissertations, <http://www.ndltd.org/> (accessed 4 September 2010)
- [37] OAI-PMH, <http://www.openarchives.org/OAI/openarchivesprotocol.html> (accessed 4 September 2010)
- [38] Olivier, E., Open Scholarship & research reporting in tandem: creating more value. The African Digital Scholarship & Curation Conference, May 12-14, 2009, 2009. Available at: http://www.ais.up.ac.za/digi/docs/olivier_paper.pdf (accessed 4 September 2010)
- [39] Pembe, F.C., Güngör, T., Structure-preserving and query-based document summarisation for web searching. *Online Information Review*, Vol. 33 No. 4 (2009), 696–719.
- [40] Pure, <http://www.atira.dk/en/pure/> (accessed 4 September 2010)
- [41] Rađenović, J., Milosavljević, M., Surla, D., Modelling and implementation of catalogue cards using FreeMarker. *Program: Electronic library and information systems*, Vol. 43 No.1 (2009), 63-76.
- [42] Rudić, G., Surla, D., Conversion of bibliographic record to MARC 21 format. *The Electronic Library*, Vol. 27 No 6 (2009), 950-967.
- [43] Scirus EDT Search, <http://www.ndltd.org/serviceproviders/scirus-etd-search> (accessed 4 September 2010)
- [44] SICRIS, <http://sicris.izum.si/default.aspx?lang=eng> (accessed 4 September 2010)
- [45] Stijepović, M., Povezivanje sistema SICRIS-COBIS.SI-Web of Science. *Infoteka*, Vol. 8 No. 1-2 (2007) 47-49.

- [46] Tešendić, D., Milosavljević, B., Surla, D., A Library Circulation System for City and Special Libraries. *The Electronic Library*, Vol. 27 No. 1 (2009), 162-186.
- [47] Vidaković, M., Milosavljević, B., Konjović, Z., Sladić, G., Extensible Java EE-based Agent Framework and Its Application on Distributed Library Catalogues. *Computer Science and Information Systems*, Vol. 6 No. 2 (2009), 1-28.
- [48] VTLS Visualizer, <http://www.vtls.com/ndltd/> (accessed 4 September 2010)

Received by the editors October 11, 2010