Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2014) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

CRIS 2014

Multi-interoperable CRIS repository

Ivanović Dragan[a]*, Ivanović Lidija[b], Dimić Surla Bojana[c]

[a]*University of Novi Sad, Faculty of Technical Sciences, Novi Sad*
[b]*University of Novi Sad, Faculty of Education, Sombor*
[c]*University of Novi Sad, Faculty of Sciences, Novi Sad*

Abstract

The paper presents the Exporter component which facilitates export of scientific-research results metadata from CRIS systems and enables creation of multi-interoperable CRIS repositories. The Exporter architecture is extensible with plugins for export of scientific-research results by various protocols and metadata formats. The Exporter is implemented as part of the CRIS UNS system including three plugins for export data throw three protocols: OAI-PMH and SRU standardized protocols, as well as an XML non-standardized protocol for exporting data to a repository of published scientific-research results of Autonomous province of Vojvodina.

Peer-review under responsibility of euroCRIS.

Keywords: CRIS; interoperability; export data; OAI-PMH; SRU

1. Introduction

CRIS (Current Research Information System) based on CERIF data model (Common European Research Information Format) is meant for processing all relevant entities of research domain. In the last couple of years, the most scientific institutions already have or are in the process of implementation of the CRIS systems and these systems already contain significant amount of data about scientific activity. Main output of scientific-research activity is scientific-research result published in journal, conference proceedings or monograph. Metadata about published scientific-research results can be stored in a CRIS system database using the semantically rich CERIF data model. However, metadata about scientific-research results can be also visible via other Internet based applications

* Corresponding author. Tel.: +381-21-485-2447.
E-mail address: dragan.ivanovic@uns.ac.rs

1877-0509
Peer-review under responsibility of euroCRIS.

such as library information systems, institutional repositories, digital libraries, information systems of publishing activity (Springer - www.springer.com, Emerald - www.emeraldinsight.com), etc. The importance of scientific-research results visibility for further development of science is discussed in many manuscripts[1,2,3,4,5,6,7,8]. On one hand, metadata about scientific-research results can be separately entered in all those Internet based systems by researchers or by librarians. This is hard and error-prone job. On the other hand, metadata about scientific-research results can be entered in one system and exported to other systems. This approach contributing to:

- Avoiding duplicated inputs on the two platforms,
- Increasing metadata quality, reliability and reusability.

It is evident that migration of data about scientific-research results from CRISs to various systems can increase visibility of those results. Because of diversity of systems to which data should be exported, a module for data export from CRISs should be able to export data to other systems by various protocols and metadata standards. Furthermore, the module should have open architecture enabling easy module extension with support for some specific protocol or metadata format.

The paper presents Exporter - the software module which facilitates export of scientific-research results metadata from CRISs. The Exporter architecture is extensible with plugins for export of scientific-research results by various protocols and metadata formats. Exporter is implemented as part of the CRIS of the University of Novi Sad (CRIS UNS). Plugins for export data throw three protocols are presented in this paper: OAI-PMH and SRU standardized protocols, as well as an XML non-standardized protocol for exporting data to a repository of published scientific-research results of Autonomous province of Vojvodina.

2. CRIS UNS

The system CRIS UNS has been developed for the needs of the University of Novi Sad according to the recommendations of non-profit organisation euroCRIS (www.eurocris.org). The implementation of the system started 2009 and it is publically available at www.cris.uns.ac.rs. Two main requirements for the specification and implementation of the system were the compliance with the international standards in the field of representing scientific-research data on one side and fulfilling the specific local requirements defined by the University and Republic of Serbia within which the University was established.

Ivanović and colleagues proposed the data model compatible with CERIF and MARC 21 library format[9]. It is shown that MARC 21 format supports all metadata proposed by Dublin Core and EDT-MS format[10]. The previously mentioned CERIF-compatible data model was the basis for developing the information system of scientific-research activity for the needs of University of Novi Sad. The architecture and the implementation of the system are presented in the paper *A CERIF-compatible research management system based on the MARC 21 format* [11]. The digital library of PhD dissertations defended at the University of Novi Sad is integrated with the CRIS UNS system[12]. Public service for searching repository of dissertations defended at the University of Novi Sad is available at: www.cris.uns.ac.rs/searchDissertations.jsf.

The CRIS UNS system contains significant amount of data about scientific-research activity at the University of Novi Sad including:

- Over 3000 researchers
- 14 faculties that belong to the University of Novi Sad
- Over 20000 published results, 3500 of which are PhD dissertations
- Over 7000 scientific conferences
- etc.

To sum up, CRIS UNS has significant amount of scientific-research data that can exported to various Internet based applications using the Exporter component which is described in the following sections.

3. The Exporter architecture

The Exporter module is implemented using Java platform and open source libraries written in Java. The module (Figure 1) contains the four components:

- *Request validator* – This component contains classes meant for checking syntax of request and for validation of provided parameters. The syntax can be expressed using XML Scheme, some ontology, CQL profile, etc.
- *Request processor* – This component contains classes meant for processing request by creation and execution queries expressed in some notation (SQL, SPARQL, Lucene query, etc.) or by invocation appropriate services of other layers of a CRIS system. Thus, this component communicates with an *Information retrieval (IR) component* or a *Database (DB) component* of a CRIS system and retrieves a set of CERIF entities which should be exported.
- *Response creator* – This component contains classes which implement or invoke appropriate services of a *Formats Converter Component* of a CRIS system for transformation of CERIF entities to a requested format: Dublin Core, ETD-MS, MARC 21, CERIF xml, etc.
- *Server side protocol manager* – This component coordinate whole export data process by some protocol. The classes of this component:

    - Invoke an appropriate class method belonging to *Request validator* component.
    - If validation of request is successfully passed, invoke an appropriate class method from *Request processor* component which retrieves a set of CERIF entities which should be exported.
    - Invoke an appropriate class method which transforms the set of CERIF entities to a requested format.
    - Create response and send it back to a client system which has sent a request.

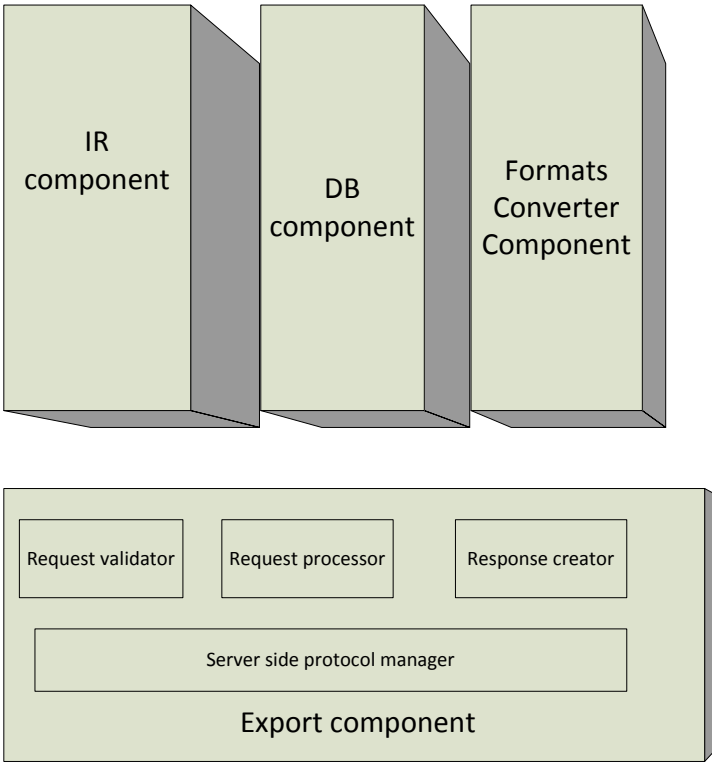

Fig. 1. The Exporter architecture

3.1. Export via OAI-PMH protocol

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a client-server protocol for exchanging the data in numerous formats (www.openarchives.org/OAI/openarchivesprotocol.html).
For the purpose of exporting data from CRIS UNS system using the Exporter component we used the OAICat library (www.oclc.org/research/activities/oaicat.html). This library primarily meant for collecting metadata about theses and dissertation for NDLTD networks provides open access and includes collection of abstractions that enable adjustments and customization for different data sources. The basic URL for OAI-PMH protocol for CRIS UNS is http://cris.uns.ac.rs/OAIHandler. Data exported in this way can be downloaded by using existing and open-access web application OAI PMH *Validator & Data Extractor* (http://validator.oaipmh.com/) created by *Vangelis Banos* (http://vbanos.gr/), a PhD student from Greece.

CRIS UNS supports exporting data in OAI-PMH protocol in the following formats: Dublin Core, MARC 21 and ETD-MS, with the plans of introducing export in CERIF format which is supported by the data model used in CRIS UNS. The CRIS UNS system is a member of the DART-Europe network (www.dart-europe.eu/) and the OpenAIRE+ network (https://beta.openaire.eu/). Interoperability of nodes of these networks is implemented using the OAI-PMH protocol and the Dublin Core format.

3.2. Export via SRU protocol

Search/Retrieve via URL (SRU) is a standard XML-based search protocol for Internet search queries, utilizing Contextual Query Language (CQL). CQL is a standard syntax for representing queries. SRU is successor to Z39.50 binary protocol, whose purpose is to simplify the Z39.50, but to save its main functionalities. The main characteristics of SRU protocol are:

- enables search of the remote systems via the Internet,
- search is based on CQL meaning it is independent of internal data representation,
- client side of the protocol can choose the format for the search results

For the purpose of implementing search of scientific-research data the new profile named CRIS was defined[13]. The implementation of the server side of the SRU protocol within CRIS UNS is in progress and is planned to be put into operation in the summer of this year. Details of this implementation are described in the paper *SRU/W service for CRIS UNS system*[14]. The component is integrated in the Exporter architecture and includes *SRU validator* for validating a CQL query, *SRU mediator* for processing a CQL query and *SRU response creator* for formatting response to a format requested by a SRU client.

3.3. Export via an XML non-standardized protocol

For the specific needs of exchanging data with the system of published scientific-research results of Autonomous province of Vojvodina (www.knr.uns.ac.rs) we defined and implemented client-server protocol that includes data about evaluation of scientific results according to the rules proposed by the Serbian Government. The protocol is based on XML and enables retrieving of data about researchers and published results. The component is integrated into the Exporter component architecture and the main URL for retrieving data is http://cris.uns.ac.rs/ReportsServlet/knr?reportType=resultsKnrXML. The parameters that can be specified in the request are *researcherId, type, resultId*. The example of the request that gives all results for the researcher Bojana Dimić Surla is http://cris.uns.ac.rs/ReportsServlet/knr?reportType=resultsKnrXML&researcherId=5514. XML documents with data obtained by using this protocol can be downloaded in a web browser without implementation of the client side. The response is given as an XML document created by the XML schema available at http://cris.uns.ac.rs/interoperability/sema-cris.xsd.

4. Conclusion

It is obvious that migration of data about scientific-research results from CRISs to various systems is very useful because it increases visibility of scientific-research results and avoids duplicated inputs on the two platforms.

The paper presents the Exporter software module meant for exporting scientific-research results metadata. The Exporter architecture is extensible with plugins for export of scientific-research results to various systems by various standardized and non-standardized protocols and metadata formats. Exporter is implemented as part of the CRIS of the University of Novi Sad (CRIS UNS). Details about implementation of plugins for export data by three protocols are presented in this paper: OAI-PMH and SRU standardized protocols, as well as an XML non-standardized protocol for exporting data to a repository of published scientific-research results of Autonomous province of Vojvodina. All three protocols retrieve the results in XML-based formats and the provided XML schemas entirely describe the semantics of the data structure. Namely, every record according to these schemas is defined by its ID number from CRIS UNS. In this way, we achieved that the exported data by this protocol can be imported in other systems without loss of information. For instance, all papers of the same researcher can be identified despite the difference in form of author's name or the same journal with the different names can be uniquely identified. Of course, the system which should import the data has to implement importing data through a user interactive process by which consolidation of data should be achieved[15]. Also, we are going to implement exporting of the CRIS UNS data according to the ontologies we already published in two papers[16, 17].

The future work on development of this module should enable definition of plugins for export data by protocols through XML files or using some formal language for describing protocols. Thus, implementation of classes for new protocols should be avoided as much as possible, i.e., customization of this module for different protocols and formats should be as simple as possible. In this way, the presented module could be used by many CRIS systems for export data by various protocols and metadata formats.

Acknowledgements

Results presented in this paper are part of the research conducted within the Grant No. III-47003 financed by the Ministry of Education and Science of the Republic of Serbia.

References

1. Lawrence, S. (2001) 'Free online availability substantially increases a paper's impact', *Nature*, vol. 411, no. 6837, paper 521.
2. Harnad, S. and Brody, T. (2004) 'Comparing the impact of open access (OA) vs. non-OA articles in the same journals', *D-Lib Magazine*, vol. 10, no. 6, retrieved from http://www.dlib.org/dlib/june04/harnad/06harnad.html.
3. Antelman, K. (2004) 'Do open-access articles have a greater research impact? ', *College & Research Libraries*, vol. 65, no. 5, pp. 372–382.
4. Anderson, K., Sack, J., Krauss, L. and O'Keefe, L. (2001) 'Publishing online-only peer-reviewed biomedical literature: Three years of citation, author perception, and usage experience', *Journal of Electronic Publishing*, Vol. 6, No. 3, DOI: 10.3998/3336451.0006.303
5. Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., & Murray, S. S. (2005a) 'The effect of use and access on citations', *Information Processing and Management*, vol. 41, no. 6, pp. 1395–1402.
6. Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M. and Murray, S. S. (2005b) 'Worldwide use and impact of the NASA astrophysics data system digital library', *Journal of the American Society for Information Science and Technology*, vol. 56, no. 1, pp. 36–45.
7. Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Murray, S. S., & Elwell, B. (2005c) 'The bibliometric properties of article readership information', *Journal of the American Society for Information Science and Technology*, vol. 56, no. 2, pp. 111–128.
8. Eysenbach, G. (2005) 'Citation advantage of open access articles', *PLoS Biology,* vol. 4, no. 5, pp. 692–698.
9. Ivanović, D., Surla, D. and Konjović, Z. (2011), "CERIF compatible data model based on MARC 21 format", *The Electronic Library*, Vol. 29, No. 1, pp. 52-70, DOI: 10.1108/02640471111111433
10. Ivanović, L., Ivanović, D. and Surla, D. (2012a), "A data model of theses and dissertations compatible with CERIF, Dublin Core and EDT-MS", *Online Information Review*, Vol. 36, No. 4

11. Ivanović, D., Milosavljević, G., Milosavljević, B. & Surla, D. (2010), "A CERIF-compatible research management system based on the MARC 21 format", *Program: Electronic libarary and information systems*, Vol. 44, No. 1, pp. 229-251, DOI: 10.1108/00330331011064249
12. Ivanovic, L., Ivanovic, D., Surla, D. (2012b), "Integration of a Research Management System and an OAI-PMH Compatible ETDs Repository at the University of Novi Sad, Republic of Serbia", *Library Resources & Technical Services*, Vol. 56, No. 2, pp. 104-112
13. Penca, V., Nikolić, S., Ivanović, D., Surla, D. & Konjović, Z. (2014a), "SRU/W Based CRIS Systems Search Profile", *Program: Electronic libarary and information systems*, Vol. 48, No. 2, in print
14. Penca, V., Nikolić, S., Ivanović, D. (2014b), " SRU/W service for CRIS UNS system ", *Proceedings of the ICIST 2014 Conference (CD), Kopaonik, March 9-13. 2014*, in print
15. Ivanovića, 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. 2012*, pp. 313-322
16. Dimić Surla, B., Segedinac, M. & Ivanović, D. (2012), "A BIBO ontology extension for evaluation of scientific research results", *Proceedings of the Fifth Balkan Conference in Informatics, Novi Sad*, Serbia, September 16 - 20, 2012, pp. 275-278, DOI: 10.1145/2371316.2371376
17. Ivanović, L., Dimić-Surla, B., Segedinac, M. & Ivanović, D. (2012c), "CRISUNS ontology for theses and dissertations", *Proceedings of the ICIST 2012 Conference (CD), Kopaonik, February 29 – March 3. 2012*, pp. 164-169, http://e-drustvo.org/icist/2012/html/pdf/509.pdf