Dragan Ivanović[1)]

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

DATA EXCHANGE BETWEEN CRIS UNS, INSTITUTIONAL REPOSITORIES AND LIBRARY INFORMATION SYSTEMS

Abstract: *This paper presents data exchange between research management system of the University of Novi Sad (CRIS UNS), institutional repositories and library information systems. The data model of CRIS UNS enables exchange of data about scientific-research results between those systems in accordance with MARC 21 format and Dublin Core format. Also, the data model enables exchange of data with other CERIF-compatible research management systems.*
Keywords: *scientific-research results, CERIF, MARC 21, Dublin Core*

1. INTRODUCTION

Main output of scientific-research activity is scientific-research result: paper published in journal, paper published in conference proceedings, monograph, etc. By 2011, scientific-research results in digital form along with associated metadata are available via various applications such as library information systems, research management systems, institutional repositories, information systems of publishing activity (Springer - http://www.springer. com/, Emerald - http://www.emeraldinsight.com/), etc. Those systems are built on different protocols and standards.

On one hand, metadata about scientific-research results can be entered in a system by researcher or by librarian. On the other hand, metadata about scientific-research results can be imported from some of previously listed systems. Therefore, creation of a data model for preserving scientific-research results which enables interoperability with all listed systems is necessary.

CRIS UNS is a research management system that has been being developed since 2008 at the University of Novi Sad. The data model of CRIS UNS enables interoperability of the system with CERIF-compatible research management systems, institutional repositories, library information systems.

The rest of the paper is structured as follows. Firstly, the various systems which contain scientific-research results and adopted standards in those systems are described. After that, the data model of CRIS UNS that enables data exchange between CRIS UNS, institutional repositories and library information systems is presented. The last section concludes the paper and outlines some directions for further investigations.

2. SYSTEMS OF SCIENTIFIC-RESEARCH RESULTS

Research management system is a software system for collecting, preserving and disseminating data about institutions, researchers, research projects, equipment, published results and other relevant data for scientific-research activity. European Union encourages the development of

national research management systems in accordance with the CERIF standard (*the Common European Research Information Format* - http://www.eurocris.org/cerif/introduction/) because the European Union wants to achieve maximum competitiveness of Europe at all levels of research activity. CERIF compatible research management systems are called CRIS (*Current Research Information System*). The first version of a CERIF was published in 1991 and current version is the CERIF2008 v1.2 that was published in 2010. CERIF describes a physical data model [12] and XML data exchange format [13].

Due to specific local or national requirements CRIS systems are built on different modifications (or extensions) of CERIF data model. There are different aspects of the CERIF data model extensions. The papers [7], [8] describe an extension that uses the formalized Dublin Core for description of published results of scientific research. Furthermore, the papers [1], [7] describe a data model extension that was created in order to satisfy requirements of research information system that was developed at the University of Bergen. Moreover, the paper [1] describes a data model extension that was created in order to satisfy requirements of scientific research information system that was developed on "Science and Technology Facilities Council". A data model extension that was created in order to satisfy requirements of the IST World portal is described in the papers [10], [14], [3], [11]).

By 2011, there are many CRIS systems: IST World (http://www.ist-world.org/), Frida (http://www.uio.no/english/research/frida/) , Pure (http://www.atira.dk/en/pure/), HunCRIS (http://nkr.info.omikk.bme.hu/HunCRIS_eng.htm), SICRIS (http://sicris.izum.si/default.aspx?lang=en) , CRIS UNS (http://www.cris.uns.ac.rs/).

CRIS UNS is a CERIF compatible research management system which has been being developed since 2008 at the University of Novi Sad in the Republic of Serbia. The first phase covered developing a system for entering metadata about published scientific-research outputs in the following forms: papers published in journals, papers published on scientific conferences, monographs, papers published in monographs. The system is built on the CERIF compatible data model based on the MARC21 format which is presented in the paper [5]. The system implementation is described in the papers [4], [15]. Published results from the system are available to anonymous user via Internet. 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 MARC21 format. Moreover, the system is in accordance with the CERIF standard and meets requirements prescribed by Ministry of Science and Technological Development of the Republic of Serbia in the field of scientific-research results evaluation. Therefore, the system data model is extended with necessary entities [6]. The system data model and architecture allow easy integration of the system with library information systems and interoperability with other CERIF compatible national CRIS systems of European countries.

Institutional repository is a software system for collecting, preserving and disseminating the scientific institution intellectual results in digital form. The intellectual results are papers published in journals, papers presented at scientific conferences, theses, dissertations, etc. Those systems increase availability of scientific-research results because most of those systems besides possibility of downloading digital archives offer gathering metadata about it in Dublin Core format using OAI-PMH protocol. The most used institutional repositories

D. Ivanović

software platforms are Eprints (http://www.eprints.org/) and DSpace (http://www.dspace.org/).

Library information system is a software system for cataloguing and circulation of bibliographic records. Those systems are built on various bibliographic standards, and a lot of library information systems are based on MARC 21 formats that are rich in metadata. By 2011, there are many library information systems: ALEPH 500, Voyager, Atriuum, Concourse, BISIS, COBISS, etc.

The relation between various systems which contains scientific-research results has been discussed in recent years [9]. The paper [2] 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). Furthermore, *Olivier* [16] describes collaboration between research management system and digital library at Pretoria University. Also, there are similar collaborations between a research management system and a digital library at Glasgow University, Southampton University and Kingston University. General objective of the CRIS-IR group (http://www.eurocris.org/Index.php?page= CRIS-IR_workplan&t=1) 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. The aim of the CRIS-OAR Interoperability project (http://www.knowledge-exchange.info/Default.aspx?ID=340) is to increase the interoperability between research management systems and Open Access Repositories by defining and proposing a metadata exchange format.

The aim of integration of systems that contain scientific-research results is

- avoiding or reducing duplicated inputs of the same metadata in various systems and
- increasing quality level of services based on these metadata.

3. DATA EXCHANGE

The data model of CRIS UNS is CERIF-compatible and it is proved in the paper [5]. Scientific-research results are stored using the MARC 21 format in accordance with ISO2709 standard. MARC21 (Machine-Readable Cataloguing for 21st century) is a standard that prescribes format for storing data relevant for library systems. An example of the MARC 21 record is shown in **Listing 1**. MARC 21 format is rich in metadata and allows detailed description of scientific-research results. The records in this form (without any changes) can be exchanged with library information systems using the Z39.50 protocol.

01142cam 2200301 a
4500001001300000000300040001300500170
0017008004100034010001700075020002500
0920400001800117042000900135050002600
0144082001600170100003200186245008600
0218250001200304260005200316300000490
0368500004000417520022800045765000330
0685650003300718650002400075165000210
0775650002300796700000210081900819_
92005291
_DLC_19930521155141.9_920219s1993
caua j 000 0 eng _ a 92005291 _
_a0152038655 :_c$15.95_
_aDLC_cDLC_dDLC_
_alcac_00_aPS3537.A618_bA88
1993_00_a811/.52 220 1 _aSandburg,
Carl, d1878-1967._10_aArithmetic /_cCarl
Sandburg ; illustrated as an anamorphic
adventure by Ted Rand._ _a1st ed._ _aSan
Diego :_bHarcourt Brace
Jovanovich,_cc1993._ _a1 v. (unpaged)
:_bill. (some col.) ;_c26 cm._ _aOne Mylar
sheet included in pocket._ _aA poem about
numbers and their characteristics. Features

anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone._ 0_aArithmetic_xJuvenile poetry._ 0_aChildren's poetry, American._ 1_aArithmetic_xPoetry._ 1_aAmerican poetry._ 1_aVisual perception._1 _aRand, Ted,_ eill.__

Listing 1. An example of MARC21 record (ISO2709)

With the development of new technologies such as XML there are also new standards for data exchange such as SRU and SRW. Those standards exchange data in the form of XML documents in accordance with XML schema for MARC records, XML schema for UNIMARC records, XML schema for Dublin Core records, and other XML schemas. An example of XML document that contains information about a MARC 21 record is shown in **Listing 2**. The records in this form (without any changes) can be exchanged with library information systems using the SRU/SRW protocol. The same MARC 21 record in accordance with ISO 2709 standard was previously shown in **Listing 1**.

```xml
<?xml version="1.0" encoding="UTF-8" ?>
<collection
    xmlns="http://www.loc.gov/MARC21/slim">
  <record>
    <leader>01142cam 2200301 a 4500</leader>
    <controlfield tag="001">
        92005291
    </controlfield>
    <controlfield
        tag="003">
        DLC
    </controlfield>
    <controlfield
        tag="005">
        19930521155141.9
    </controlfield>
    <controlfield tag="008">
        920219s1993 caua j 000 0 eng
```

```xml
  </controlfield>
  <datafield tag="010" ind1="" ind2="">
    <subfield
        code="a">
        92005291
    </subfield>
  </datafield>
  <datafield tag="020" ind1=""
      ind2="">
    <subfield
    code="a">
        0152038655 :
    </subfield>
    <subfield
        code="c">
        $15.95
    </subfield>
  </datafield>
  <datafield tag="040" ind1=""
      ind2="">
    <subfield
        code="a">
        DLC
    </subfield>
    <subfield
        code="c">
        DLC
    </subfield>
    <subfield
        code="d">
        DLC
    </subfield>
  </datafield>
  <datafield tag="042" ind1=""
      ind2="">
    <subfield
        code="a">
        lcac
    </subfield>
  </datafield>
  <datafield tag="050" ind1="0"
      ind2="0">
    <subfield
        code="a">
        PS3537.A6
        18
    </subfield>
    <subfield
        code="b">
        A88 1993
    </subfield>
  </datafield>
  <datafield tag="082" ind1="0"
      ind2="0">
    <subfield
```

D. Ivanović

```
            code="a">
                811/.52
        </subfield>
        <subfield code="2">20</subfield>
    </datafield>
    <datafield tag="100" ind1="1"
        ind2="">
        <subfield
            code="a">
                Sandburg,
                Carl,
        </subfield>
        <subfield code="d">1878-
            1967.</subfield>
    </datafield>
    <datafield tag="245" ind1="1"
        ind2="0">
        <subfield code="a">
                Arithmetic
        </subfield>
        <subfield code="c">Carl
            Sandburg ; illustrated as an
            anamorphic adventure by
            Ted Rand.
        </subfield>
    </datafield>
    <datafield tag="250" ind1="" ind2="">
        <subfield code="a">1st
            ed.</subfield>
    </datafield>
    <datafield tag="260" ind1="" ind2="">
        <subfield code="a">San Diego
            :</subfield>
        <subfield code="b">
            Harcourt Brace Jovanovich,
        </subfield>
        <subfield
            code="c">
                c1993.
        </subfield>
    </datafield>
    <datafield tag="300" ind1="" ind2="">
        <subfield
            code="a">
                1 v. (unpaged) :  </subfield>
        <subfield code="b">ill. (some
            col.) ;
        </subfield>
        <subfield code="c">26
            cm.</subfield>
    </datafield>
    <datafield tag="500" ind1="" ind2="">
        <subfield
            code="a">
                One Mylar sheet included in
                    pocket.
        </subfield>
    </datafield>
    <datafield tag="520" ind1="" ind2="">
        <subfield
            code="a">
                A poem about numbers and
                their characteristics.
                Features anamorphic, or
                distorted, drawings which
                can be restored to normal by
                viewing from a particular
                angle or by viewing the
                image's reflection in the
                provided Mylar cone.
        </subfield>
    </datafield>
    <datafield tag="650" ind1=""
        ind2="0">
        <subfield code="a">Arithmetic
            </subfield>
        <subfield
            code="x">
                Juvenile
                poetry.
        </subfield>
    </datafield>
    <datafield tag="650" ind1=""
        ind2="0">
        <subfield
            code="a">
                Children's poetry,
                American.
        </subfield>
    </datafield>
    <datafield tag="650" ind1=""
        ind2="1">
        <subfield
            code="a">Arithmetic</subfie
            ld>
        <subfield
            code="x">Poetry.</subfield>
    </datafield>
    <datafield tag="650" ind1=""
        ind2="1">
        <subfield code="a">
                American poetry.
        </subfield>
    </datafield>
    <datafield tag="650" ind1=""
        ind2="1">
        <subfield code="a">
                Visual perception.
        </subfield>
    </datafield>
```

```
    <datafield tag="700" ind1="1"
         ind2="">
      <subfield code="a">Rand,
          Ted,</subfield>
      <subfield code="e">ill.</subfield>
    </datafield>
  </record>
</collection>
```

Listing 2. An example of MARC21 record (XML document)

Conversion between MARC 21 records and Dublin Core records is defined in both directions, but during the conversion some of the data can be lost due to the differences in format metadata richness. There are software packages that implement that conversion: MarcEdit, MARConvert[tm] and others. It means that CRIS UNS can exchange data with institutional repositories based on Dublin Core. An example of Dublin Core record as XML document is shown in **Listing 3**. This record in MARC 21 format was previously shown in **Listing 1** (in accordance with ISO2709 standard) and **Listing 2** (as XML document).

```
<?xml version="1.0" ?>
<dc
    xmlns="http://purl.org/dc/elements/1.1
    /">
  <title>Arithmetic /</title>
  <creator>Sandburg, Carl, 1878-
      1967.</creator>
  <creator>Rand, Ted,
      ill.</creator>
  <type />
  <publisher>
      San Diego :Harcourt Brace
      Jovanovich,
  </publisher>
  <date>c1993.</date>
  <language>eng</language>
  <description>
      A poem about numbers and their
      characteristics. Features
      anamorphic, or distorted, drawings
      which can be restored to normal by
      viewing from a particular angle or
      by viewing the image's reflection in
      the provided Mylar cone.
  </description>
```

```
  <description>
      One Mylar sheet included in
      pocket.
  </description>

      <subject>Arithmetic</su
      bject>
  <subject>
      Children's poetry, American.
  </subject>

      <subject>Arithmetic</su
      bject>
  <subject>American
      poetry.</subject>
  <subject>Visual
      perception.</subject>
</dc>
```

Listing 3. An example of Dublin Core record

4. CONCLUSSION

The data model of CRIS UNS is CERIF-compatible and stores metadata about scientific-research results in the MARC 21 format. Scientific-research results can be also stored in various library information systems (LIS) or institutional repositories (IR). Therefore, data exchange between CRIS UNS, LIS and IR can significantly increase availability of scientific-research results. MARC 21 format is rich in metadata and allows detailed description of scientific-research results. A lot of library information systems are based on MARC 21 format. CRIS UNS can exchange data with library information systems in accordance with MARC 21 format. On the other hand, a lot of institutional repositories are based on Dublin Core format. Conversion between MARC 21 records and Dublin Core records is defined in both directions. There are software packages that implement that conversion: MarcEdit, MARConvert[tm] and others. Therefore, CRIS UNS can exchange data with institutional repositories in accordance with Dublin Core format.

.

REFERENCES:

[1] Asserson, A., Jeffery, K. and Lopatenko, A. (2002), "CERIF: Past, Present and Future: An Overview", Proceedings of the 6th International Conference on Current Research Information Systems, University of Kassel, August 29 - 31, 2002, pp. 33-40.

[2] Dijk, E., Baars, Chr., Hogenaar, A. and Meel, M. van (2006), "NARCIS: The Gateway to Dutch Scientific Information", Elpub 2006 Conference, Bansko, June 14-16, 2006, pp. 49-57

[3] Ferlež, J. (2005), "Public IST World Deliverable 1.3 – Data Model for Representation of Expertise", 12 p., available at: http://ist-world.dfki.de/downloads/deliverables/ISTWorld_D1.3_DataModelForRepresentationOfExpertise.pdf (accessed 13 October 2009)

[4] Ivanović, D., Milosavljević, G., Milosavljević, B. and Surla, D. (2010a), "A CERIF-compatible research management system based on the MARC21 format", Program: Electronic libarary and information systems, Vol. 44, No. 3, pp. 229-251, DOI: 10.1108/00330331011064249

[5] Ivanović, D., Surla, D. and Konjović, Z. (2011a), "CERIF compatible data model based on MARC21 format", The Electronic Library, Vol. 29, No. 1, pp. 52-70, DOI: 10.1108/02640471111111433

[6] Ivanović, D., Surla, D. and Racković, M. (2011b), "A CERIF data model extension for evaluation and quantitative expression of scientific research results", Scientometrics, Vol. 86, No. 1, pp. 155-172, DOI: 10.1007/s11192-010-0228-2

[7] Jeffery, K., Asserson, A. and Revheim, J. (2000), "CRIS, Grey Literature and the Knowledge Society", Proceedings CRIS-2000, Helsinki, 22 p., available at: ftp://ftp.cordis.lu/pub/cris2000/docs/jeffery_fulltext.pdf (accessed 13 October 2009)

[8] Jeffery, K. (2000), "An architecture for grey literature in a R&D context", The International Journal on Grey Literature, Vol. 1, No 2, pp. 64-72.

[9] Joint, N. (2008), "Current research information systems, open access repositories and libraries ", Library Review, Vol. 57, No 8, pp. 570-575

[10] Jörg, B., Ferlež, J. and Grabczewski, E. (2005), "Public IST World Deliverable 1.2 – Data Model for Knowledge Organisation", 18 p., available at: http://ist-world.dfki.de/downloads/deliverables/ISTWorld_D1.2_DataModelForKnowledgeOrganisation.pdf (accessed 13 October 2009)

[11] Jörg, B., Ferlež, J., Grabczewski, E. and Jermol, M. (2006), "IST World: European RTD Information and Service Portal", 8th International Conference on Current Research Inforation Systems: Enabling Interaction and Quality: Beyond the Hanseatic League (CRIS 2006), Bergen, Norway, 10 p., available at: http://epubs.cclrc.ac.uk/bitstream/905/ISTWorld01.pdf (accessed 13 October 2009)

[12] Jörg, B., Krast, O., Jeffery, K. and Grootel, G. (2009a) "CERIF 2008 - 1.0 Full Data Model (FDM) Introduction and Specification", 43 p., available at: http://www.eurocris.org/fileadmin/cerif-2008/CERIF2008_1.0_FDM.pdf (accessed 13 October 2009)

[13] Jörg, B., Krast, O., Jeffery, K. and Grootel, G. (2009b), "CERIF 2008 – 1.0 XML Data Exchange Format Specification", 33 p., available at: http://www.eurocris.org/fileadmin/cerif-2008/CERIF2008_1.0_XML.pdf (accessed 13 October 2009)

[14] Kiryakov, A., Grabczewski, E., Ferlež, J., Uszkoreit, H. and Jörg, B. (2005), "Public IST World Deliverable 1.1 – Definition of the Central Data Structure", 23 p., available at: http://ist-world.dfki.de/downloads/deliverables/ISTWorld_D1.1_CentralDataStructure.pdf (accessed 13 October 2009)

[15] Milosavljević, G, Ivanović, D., Surla, D. and Milosavljević, B. (2011), "Automated Construction of the User Interface for a CERIF-Compliant Research Management System", The Electronic Library, Vol. 29, No. 5 (in press)

[16] Olivier, E., (2009), "Open Scholarship & research reporting in tandem: creating more value", The African Digital Scholarship & Curation Conference, May 12-14, 2009, available at: http://www.ais.up.ac.za/digi/docs/olivier_paper.pdf (accessed 4 September 2010)

D. Ivanović