Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 106 (2017) 82 – 86

Procedia Computer Science

13th International Conference on Current Research Information Systems, CRIS2016, 9-11 June 2016, Scotland, UK

Regional Portal FVG: effective interoperability through DSpace-CRIS and open standards

Jordan Piščanc[a*], Romano Trampus[a], Luisa Balbi[a], Michele Mennielli[b], Susanna Mornati[c] , Luigi Andrea Pascarelli[d], Andrea Bollini[d]

[a]*University of Trieste, Via E.Weiss 21, Trieste 34127, Italy*
[b]*Cineca, Via Magnanelli 6/3, Casalecchio di Reno 40033, Italy*
[c]*Cineca, Via R. Sanzio 4, Segrate 20090, Italy*
[d]*Cineca, Via dei Tizi 6/B, Roma 00185, Italy*

Abstract

Friuli-VeneziaGiulia (FVG) "Regional Scientific System" includes three Public Research Institutions, three Universities, four International Institutions, four Technological Parks in FVG region in North-East Italy.

In 2014 the three Universities started to cooperate for a common research output inside a project named UnityFVG: United Universities-FVG. They already have 10 years experience in OA with four Institutional Repositories (DSpace/DSpace-CRIS based) and more than 100,000 Research Publications. So they decided to provide a single point of access under a new Regional Research Portal based on DSpace-CRIS.

The project, with the technical support of Cineca consortium, offers a great opportunity to improve the interoperability of DSpace-CRIS based solutions. The European standard for the research domain, CERIF, is the best option to drive rich information to the portal in a standard and reusable way. A plugin/patch for DSpace will be freely available to enable data export using CERIF-XML over OAI-PMH. CERIF-XML will be available for all the main entities (People, Projects, Organizations, Journals, Conferences, Dataset, Publications and metrics). The DSpace OAI-PMH harvester will be extended to support ingestion of complex, interconnected information as provided by the CERIF-XML format. This will enable content replication between DSpace-CRIS instances and easy setup of public OA oriented portals.

* Corresponding author. Tel.:+39-040-558-6102 fax: +39-040-558-6133.
E-mail address: piscanc@units.it

Peer-review under responsibility of the Organizing Committee of CRIS2016
doi:10.1016/j.procs.2017.03.038

Peer-review under responsibility of the Organizing Committee of CRIS2016

Keywords: DSpace-CRIS, CERIF-XML, OpenAIRE, OAI-PMH

1. The Research System in Friuli Venezia Giulia

Friuli Venezia Giulia Autonomous Region is among the most advanced areas in the European Union thanks to its economic activities, focused on innovation and internationalisation, and its central position in the enlarged Europe. Friuli Venezia Giulia is also one of the Lonely Planet Best In Travel Region † to visit in 2016.

The Regional Scientific System currently includes many organizations: three Public Research Institutions under the aegis of the Italian Ministry of Education, University and Research; three Universities; four International Institutions; four Technology Parks.

The high concentration of public and private research institutes in Friuli Venezia Giulia led to the creation of a regional "Research System", which was formally set up in 2004 under the name "Coordination Network of the National and International Research Centres located in Friuli Venezia Giulia" (http://cer.areasciencepark.it/en/ - the Coordination Network agreement was signed on January 27, 2004). This organization is supported by the Friuli Venezia Giulia Regional Authority, the Italian Ministry of Foreign Affairs and the Italian Ministry of Education, University and Research. It aims at developing technological innovations, increasing business competitiveness and creating the most favourable conditions for researchers coming from all over the world.

Nowadays in Friuli Venezia Giulia there are more than 8 researchers for every 1.000 inhabitants, the double of the national average 4/1.000. Moreover, every year more than 7.500 researchers from all over the world come to study and carry out research activities in one of the scientific and academic institutions located in the Region.

2. The Research-UnityFVG Project

In this context, in 2014 the University of Trieste (http://www.units.it/), the University of Udine (http://www.uniud.it) and the International School for Advanced Studies (SISSA - http://www.sissa.it/) started to cooperate in some common sectors like the research output, library services, technological transfer and integration of services for students. This cooperation resulted into what is now called UnityFVG, United Universities of Friuli Venezia Giulia.

The three universities already have a 10 years experience in Open Access with their Institutional Repositories. The University of Trieste started in 2006 with the first DSpace Institutional Repository “OpenstarTS” (http://www.openstarts.units.it/). This repository is also well ranked by webometrics: it has reached in 2015 the 2nd place in the Italian national ranking and the 184th place in the international one. Over 10.000 items of Doctoral Thesis, EUT University Press publications and Conferences are archived. Over 5 millions downloads achieved in 2014.

The OpenstarTs started as a pilot project in Bergen during the DSUG meeting[2]. DSpace 1.4 was selected as the Institutional Repository software. In 2010 the repository was one of the first pilot Italian Repositories adopting the OAI-PMH for harvesting Doctoral Thesis by the National Libraries for “Deposito Legale” **(**http://www.depositolegale.it/deposito-legale-digitale-delle-tesi-di-dottorato/**)**.

In 2012 OpenstarTs was involved also in the development of The National Bibliography Number Italia (NBN:IT)

† Lonely Planet Best In Travel 2016: https://www.lonelyplanet.com/italy/friuli-venezia-giulia

Project: a persistent identifier supporting national legal deposit for digital resources. The DSpace plugin for NBN-IT was developed (http://www.depositolegale.it/national-bibliography-number/). Finally in 2016 the migration from DSpace1.8 to the new DSpace-CRIS 5.5 (https://wiki.duraspace.org/display/DSPACECRIS/DSpace-CRIS+Home) started to take advantage of the new potentials of the CRIS Entities: Researcher Profiles, Journals, Events. The EUT University Press could get advantages from the CRIS: Authors Profiles with their Curricula, Journals information with Editorial Boards and instructions for Authors, definition of Conference Events.

SISSA, the International School for Advanced Studies, started its Digital Library Repository (DSpace based) in 2005 (http://urania.sissa.it/). The SISSA Digital Library has digitized all PhD and magister theses of the School and have archived all scientific papers produced by SISSA authors from 2005 to 2014. The collection of more than 8.000 items is composed of magister theses, master theses in HPC, papers, PhD theses, tesi di Master in comunicazione della scienza (in Italian). The repository hold items from 1981 to nowadays.

Also the University of Udine supports the open access movement since 2004 with various initiatives (conferences, seminars, archive of doctoral theses, etc.).

In 2015 the University of Trieste and the University of Udine have adopted the Cineca's IRIS platform[3], based on DSpace-CRIS, which collects the scientific output of professors and researchers. All this IRIS repositories together lead to more than 100,000 research Items: mainly scientific articles, chapters of books, monographs, conference proceedings.

In this context the three universities decided to provide a single point of access under a new Regional Research Portal called: UnityFVG - Ricerca Friuli Venezia Giulia (http://ricerca-fvg.cineca.it). The Portal will be a web gateway for searching and spreading the research data created by the SISSA, University of Trieste and University of Udine. The portal is also an open access repository that allows access without restrictions to the results of the activities financed by public funds. DSpace-CRIS has been selected to implement such portal. Technical support will be provided by Cineca (http://www.cineca.it/en), the Consortium of Italian Universities that developed DSpace-CRIS as the result of a project in collaboration with The University of Hong Kong[3].

3. Interoperability: the use of open standards and open source technologies

The project offers a great opportunity to improve the interoperability of DSpace and DSpace-CRIS based solutions. CERIF[4], the Common European Research Information Format developed and maintained by euroCRIS (http://eurocris.org/), has been chosen as the best option to send rich information to the regional portal in a standard and reusable way. A plugin/patch for DSpace will be freely available to enable data export using CERIF-XML over OAI-PMH[5]. When the ORCID integration is enabled, all the information related to authors present in the DSpace repository will be also exposed thanks to the use of Federated Identifier for the cfPers entity.

The CERIF-XML Data Exchange Format will be available and configured out-of-box in DSpace-CRIS for all the main entities under the proposed sample configuration (People, Projects, Organizations, Journals, Conferences, Dataset, Publications and metrics). More specifically the OpenAIRE Guidelines for CRIS Manager[6] will be implemented in the out-of-box configuration.

DSpace-CRIS administrators will be also able to configure the mapping from their custom data model to the CERIF-XML Data Exchange Format to maximize the interoperability. Different scenarios will be supported to allow the mapping between the internal configurable DSpace-CRIS data model and the CERIF data model.

The following scenarios were considered:

A. A single DSpace-CRIS entity is used to track information managed in a more granular way in CERIF. For example funding information could be stored directly in the project record. In this case the DSpace-CRIS project entity needs to lead to two different CERIF entities, cfProj and cfFund. The system will assume that one specific attribute of the project, such as the funding name or the funding code, univocally defines the linked cfFund entity. The hash of the string value of such attribute will be used to create the identifier for

such entity.

B. Additional and multilingual entities are preferably represented in DSpace-CRIS using simple fields or nested objects, both fully supported. These two options both provide the opportunity to specify in the mapping the default values for each single field in case the information are not enough to completely describe the CERIF entity (e.g.: cfTrans and cfLangCode for multilingual entities).

C. Linking entities can come from DSpace-CRIS nested object or can be implicitly defined by the DSpace-CRIS field definition. For example a preferedName field in the ResearcherPage DSpace-CRIS entity can lead to cfPers_PersName link entity where the cfClass and cfSchema are 55f90543-d631-42eb-8d47-d8d9266cbb26 and 7375609d-cfa6-45ce-a803-75de69abe21f the "Presented name" classification of the "Person Names" schema.

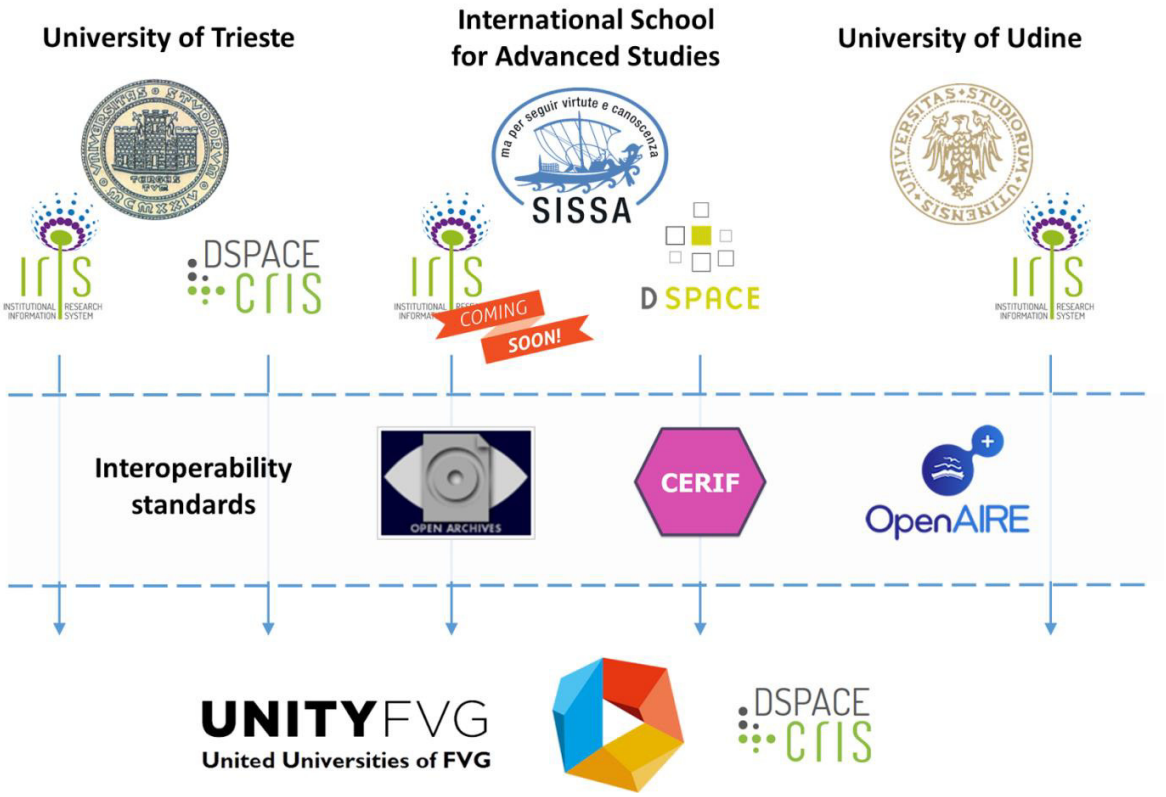

Fig. 1. Research-UnityFVG Ecosystem

A difficult process to manage is the generation of the ID for CERIF entities that are produced with the mapping. In this scenario we cannot use any of the DSpace-CRIS identifiers to generate such an ID. Indeed there are not easy ways to avoid conflicts when more DSpace-CRIS entities are mapped to the same CERIF entity (e.g.: Journals and Articles), or a CERIF entity is derived from a simple string value embedded in a "bigger" DSpace-CRIS entity (e.g.: funding in a project record).

The ID is generated combining: the DSpace-CRIS entity name; the DSpace-CRIS persistent identifier of the object (when available), or the md5 hash of the identifying attribute for embedded entities (cfFund in project); a positional index to represent the target entity and instance. In this way we can guarantee the stability of the identifier throughout multiple generations of the export.

By now, only publicly available information will be exposed in CERIF. The reason for this it is that the information are exposed over OAI-PMH which doesn't support any authentication/authorization mechanism.
To produce the CERIF XML we have successfully used the ETK CERIF-Tools library (https://github.com/EKT/CERIF-Tools). A little hack was required to be able to use our generated ID since the CERIF-Tools library has limited the IDs to a Long while the CERIF format allows also the char(128).

Once created, the CERIF XML representation is stored in the SOLR OAI document and exposed to the XOAI so to be immediately available to the CERIF format crosswalk. The built-in XOAI library support for definition of OAI Sets using filtering condition allows us to easily meet the requirement expressed in the OpenAIRE guidelines on the availability of Sets by entity type.

The current solution is bound to a single configurable CERIF mapping and profile (in our case the OpenAIRE profile). To include the support for additional profiles, it is possible to proceed with a small hack and store the resulting CERIF XML for any profiles. A future development would explore the feasibility of creating a general purpose mapping that will be as wide as possible and automatically produce the specific profile from the CERIF XML once a standard way to describe CERIF profile is implemented.

As part of the project we are also extending the DSpace OAI-PMH harvester to support ingestion of complex, interconnected information as provided by the CERIF-XML format. At the beginning such functionality will be tailored on the OpenAIRE Guideline for CRIS Managers and on the out-of-box configuration of DSpace-CRIS mapping from CERIF to the DSpace-CRIS data model. It is also limited to the full trusted collaboration scenario of the FVG regional portal. A configurable system to remap the CERIF data model to a generic DSpace-CRIS data model is out-of-scope for the project but this first implementation is obviously a first step towards the right direction and a base implementation to hack for other specific projects. In other alternative scenarios it could be necessary to introduce workflows to quality check, approve and merge information coming from external sources.

Finally the implemented functionalities to expose and harvest CERIF OpenAIRE XML records over OAI-PMH will enable content replication between DSpace-CRIS instances and an easy setup of public, free and Open Access oriented portals for Institutions that rely on other CERIF compliant products to manage their research.
In the long run, we expect to support more interactive functionalities for the researchers in the FVG portal. As an example, the ability to login using their ORCID credentials to improve their Publications, Affiliations and biography details.

Another functionality that might be taken into consideration for future developments it is to make it possible for the system to combine the usage Metrics of the original CRIS system with the one gather by the regional portal. This data will then be available to others to be used and used again.

References

1. S. Arabito, F. Asnicar, *OpenstarTs: the institutional repository of the University of Trieste*, Poster, DSUG Meeting, Bergen, 2006.
2. D. Palmer, A. Bollini, S. Mornati, M. Mennielli, *IRIS: Supporting & Managing the Research Life-cycle* in Universal Journal of Educational Research 4(4): 738-743, 2016; DOI: 10.13189/ujer.2016.040410
3. D. Palmer, A. Bollini, S. Mornati, M. Mennielli. *DSpace-CRIS@HKU: Achieving visibility with a CERIF compliant Open Source System* in "Managing Data-Intensive Science: the Role of Research Information Systems in Realising the Digital Agenda"; Proceedings of the 12th International Conference on Current Research Information Systems (2014); Volume 33, 2014, Pages 118–123; doi:10.1016/j.procs.2014.06.019
4. euroCRIS CERIF Task Group, *Main Features of CERIF*, http://eurocris.org/cerif/main-features-cerif
5. The Open Archive Initiative, *Protocol for Metadata Harvesting*, https://www.openarchives.org/OAI/openarchivesprotocol.html ; 2002
6. N. Houssos, B. Joerg, J. Dvořák, *OpenAIRE Guidelines for CRIS Managers 1.0,* Zenodo, 10.5281/zenodo.17065, 2015