



CRIS 2014

BoRIS and BIA: CRIS and Institutional Repository integration at the Free University of Bozen-Bolzano

Luigi Siciliano^a, Sabine Schmidt^{a*}, Manuel Kinzler^b

^aLibrary of the Free University of Bozen-Bolzano, Universitätsplatz 1 - piazza Università, 1-1 - 39100 Bozen-Bolzano Italy

^bDepartment Information & Communication Technologies (ICT) of the Free University of Bozen-Bolzano, Universitätsplatz 1 - piazza Università, 1-1 - 39100 Bozen-Bolzano Italy

Abstract

This paper describes the implementation at the Free University of Bozen-Bolzano (FUB) of a Current Research Information System named BoRIS (Bozen-Bolzano Research Information System, an implementation of Avedas Converis 5 software) integrated not only with locally developed Human Resource Information (HRIS) and Financial systems (PIS) but also with an Institutional Repository BIA (Bozen-Bolzano Institutional Archive, an hosted instance of open source software DSpace 3.2) and the publication database of the Italian Ministry for Education, Universities and Research “Ministero dell’Istruzione, dell’Università e della Ricerca, MIUR”.

© 2014 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of euroCRIS.

Keywords: CRIS; Dspace; Institutional Repository; Interoperability

1. Background

The Free University of Bozen-Bolzano (FUB) was founded October 31st, 1997 as a multilingual, internationally-oriented institution. Faculties are Computer Science, Economics and Management, Education, Design and Art, Science and Technology. At present, there are approx. 3,000 students enrolled and 105 tenured professors and researchers.

When first planning the implementation of an Institutional Repository at the FUB, we soon understood that researchers’ level of annoyance due to recurrent requests of data entry, ranging from the MIUR database[1] to

* Corresponding author. Tel.: +39 (0)471012318; fax: +39 (0)471012309.

E-mail address: sabine.schmidt@unibz.it

University or Faculty homepages was already so high, that any attempt to implement an Institutional Repository asking researchers to again submit their publications would have failed miserably, despite our pointing out of the advantages of self-archiving in such a platform. For this reason even though we quickly identified DSpace[2] as the best platform for our purposes, we decided to wait for the launch and, instead, to join our forces with the Research and Innovation Service of the University in order to implement and customize the Current Research and Information System (CRIS)[3] Converis by Avedas. We named this local implementation of Converis BoRIS (Bozen-Bolzano Research Information System).

2. BoRIS as a Converis 5 instance

BoRIS as a CRIS covers of course the whole research life cycle, from the initial idea through the project application workflow, the project management, the management of research outcomes up to reporting.

Key factors for choosing Converis were:

- Entirely web-based application with no need of OS dependent client installation;
- User-friendly and intuitive UI. Thanks to the user-friendly interface of Converis the introduction required by researchers could be confined to a series of two-hour training sessions. Only for particular FAQs it will be necessary to implement a set of help pages with guidelines and screencasts;
- Flexible customization of templates for projects, publications and activities; easy definition of user roles and rights and, even more important, workflow steps - for example in the approval process of project applications. According to the role of a person the right to edit or only see a single field can be adjusted. According to the type of a project – internally or externally funded, EU or ministerial funding etc. – the approval process varies and by choosing the type the related workflow is automatically assigned;
- Possibility of integration with other systems. For us the CRIS should be the only point of data entry for projects, publications and other activities (conference participations, professorial memberships etc.) related to research, and it should be possible to reuse data in other systems or platforms (websites, databases etc.). Integration with in-house developed software solutions for human resource information and financial data was also a must-have;
- Being located in a multilingual region such as South Tyrol, a completely trilingual –Italian, English and German – localization of both front-end and back-end was a crucial point in the project. In Converis each user can choose the language anywhere anytime, not only when logging in. The Converis out of the box localization was English, German and French. For the FUB the French localization has been replaced with an Italian translation provided and customized mainly by internal staff from the research office;
- Converis is fully compliant with the standard Common European Research Information Format Metadata Scheme (Cerif) [4];
- Requirement for an easy way to configure the system is addressed by the Converis Configuration Module, which features a User Interface where all key elements of the system can be customized to individual needs.

3. Integration of BoRIS with in-house systems

Most important for us was the possibility to implement interfaces between our CRIS and other existing applications already used at the FUB. Since both Human Resource Information System (HRIS) and Purchasing Information System (PIS) are based upon relational databases and BoRIS is also locally installed on a FUB server, in our implementation the Converis software can connect directly to the databases via SQL. In both interfaces Converis connector loads the required data, maps the attributes and saves them into the corresponding Converis tables.

All of the following three interfaces are up and running in our production environment.

3.1. Authentication

Key feature not only for an easier maintenance but also for a wider acceptance of the system among researchers was a hassle-free way to log in, without the need of a double registration and yet another user and password to remember. This was achieved by implementing a Single sign-on authentication system, and BoRIS is now integrated

with FUB Shibboleth Identity Provider (IdP) server[5]. This allows users to authenticate using their usual FUB user name and password.

Authentication was also seen as an opportunity to seamlessly provide additional features. Shibboleth attributes are used to connect a user account via an entry point to its data in the system and to automatically assign its BoRIS user roles. As a consequence no further user and password administration is necessary in BoRIS and roles and affiliation are automatically assigned at the login stage.

3.2. Integration with the in-house developed Human Resource Information System

To provide access to all essential master data for managing research projects avoiding multiple data entry and inconsistencies, a one-directional connection with the in-house developed HRIS was developed in BoRIS.

Once a day essential data about staff, their organizational affiliations and positions as per contract are imported from HRIS to BoRIS. The connection also includes master data on PhD programs, PhD students and study plans, thus granting consistency in the system.

3.3. Integration with our in-house developed Financial System

The in-house developed PIS consists of:

- a budget management module covering the whole process of planning, approving and releasing any budget at the FUB;
- a form-based workflow management tool which allows the community to request budget-consuming expenditures such as purchase orders or reimbursement of travel expenses. Availability control performs checks against the released budget.

The purpose of the integration with PIS is to make the projects' current budget data available in BoRIS, including all budget modifications such as supplements, returns, etc. and recorded expenses in details.

This is achieved via a one-directional connection from PIS to BoRIS. Data are updated in BoRIS on a scheduled basis every 15 min.

4. BIA as a DSpace implementation

The counterpart of BoRIS, focusing on the dissemination of research outputs, is the Bozen-Bolzano Institutional Archive (BIA). As a matter of fact, availability of publications in a standard compliant Institutional Repository provides further advantages to the community of researchers:

- Global visibility of the institution's research outputs;
- Open Access, when allowed, to the full text of the research output;
- Interoperability via OAI-PMH[6] with other important assets such as the library catalogue, currently running an Ex Libris Primo discovery tool, making research even more accessible to the academic community;

Nowadays many solutions for Institutional Repository software can deliver these features. Compared to other open source solutions DSpace was preferred because it:

- Enables out of the box full text search of the research outcomes contents and not only in the metadata with results ranked by relevance;
- Provides out of the box persistent identifiers via Handle system[7], thus allowing persistence of the links and more reliable citability of the outcome in papers and scholarly publications;
- We wanted a hosted solution, and by choosing Dspace one can rely on the widest range of registered service providers offering qualified hosted solutions[8]. After a careful evaluation, having an eye especially on value for money, we decided for the service provider LongSight[9].

5. BoRIS as a unique data entry point

As one of the partners involved in the project the University Library paid particular attention to the descriptive metadata used in the system, and worked actively to make BoRIS the one and only data entry point for publication

data for our researchers. To achieve this goal we also collected, documented and exchanged required documentation about services and interfaces between our partners and counterparts (Avedas, Cineca and Longsight).

Eventually BoRIS has been officially presented on November 13th, 2013 and as of today we are fine tuning the implementation of its interfaces for reuse of publication data. Authors can thus decide, in a single interface, for every single publication, if they want to send the bibliographic data to the University's research website, to the Ministerial Database (for national validation not only of the person, but also of the universities they are affiliated with) and to the institutional archive BIA. Eventually, in the case of BIA, users can also agree to the transmission of the full text of the publication.

5.1. *Dspace*

Out of the box Converis had the possibility of an integration with DSpace or Eprints repositories using a Sword2[10] compliant interface. We used the default DSpace integration as a starting point for our customization.

The preliminary work for this integration was to change the existing Converis-DSpace metadata mapping, therefore a lot of configuration work to align the publication types and templates to the Cineca database was required. On the side of DSpace the sword2-server.cfg file had to be updated and installed by our service provider Longsight.

DSpace is hierarchically organized in communities and collections. These are the predefined ways of browsing content, and an item cannot be outside these logical containers but is required to depend on one of them. It was therefore important to create the collections in a way allowing a consistent mapping in BoRIS. Since BoRIS was the unique point for data entry, it was important for any published item to end up in a proper collection in BIA. It is worth mentioning that even though in DSpace it is theoretically possible to place an item in more than one collection using the so called "item mapper tool", this is not possible using an automatic export procedure. In addition, each collection must have not only a unique ID but also a unique name. These additional technical constraints led us to keep our structure a bit redundant. We created a Community for PhD Thesis divided in three collections reflecting the three Faculties currently offering a PhD-programme, and six more communities, again one for each Faculty plus the Centre for Regional History, each divided in different broad resource types (articles, books related publications, conference related publications, grey literature, teaching materials and of course other). As a result we ended up with 39 uniquely identified collections.

On the other hand, in BoRIS a new section for the transmission to BIA, named "DSpace attributes" has been created. The logical container corresponding to a DSpace collection in BoRIS is the so called choice group "IR Collection". Each value for choice group "IR Collection" is mapped 1:1 to a collection, thus providing for each submission a safe and consistent place to be stored in.

The section for the transmission of DSpace attributes also contains two fields updated automatically once the data have been sent to BIA (irHandle and transmission date); another date field to enter an embargo for the transmission of the full text if requested by publisher's policy; a flag field for "full text deposited" and, last but not least, the yes-or-no flag to start the transmission.

The validation procedure, carried out by the library staff, not only includes a check of the bibliographic data, but also the verification that the uploaded full text version can be archived in the Institutional Archive according to the publishers' policies. In the case of journal article this check is made easier by an embedded Sherpa/Romeo lookup[11]. Then each publication needs to be categorized considering resource type and affiliation of the author in the appropriate DSpace collection before the transmission is started. Validation is performed by a librarian and once is done, the full text of the publication together with its bibliographic description is loaded onto our DSpace instance. Dedicated personnel is currently performing this validation task for publications inserted in BoRIS so far, and BIA will be officially launched as soon as we have reached a critical mass of content.

5.2. *MIUR database*

The publication data in BoRIS must be automatically published to the official MIUR database maintained by CINECA via a dedicated web service. The workflow for the implementation of this web service started with the translation of the documentation provided by Cineca in Italian language into a German version, to make it easier

understandable for the Converis team at the Germany based company Avedas. The next step was the mapping between the metadata. Both systems use standard XML formats (Converis Cerif and Cineca MODS[12]) and we managed to map each element and also each value from the different controlled vocabularies.

As we knew that BoRIS should have been the only data entry point and then all data should have been reused in other systems, we changed the publication type specification in BoRIS from the beginning in order to align them with the publication types in the MIUR database, thus achieving a 1:1 concordance that simplified the mapping of the typologies. In the same way we identified the mandatory fields in the publication templates of BoRIS: we figured out the required attributes in the Cineca database and then marked as mandatory the corresponding attributes in BoRIS.

The validation of the authors in BoRIS is guaranteed thanks to the interface with the HRIS: only authors with an institutional affiliation can be linked to publications. To avoid duplicates in the system, one publication can be linked to more internal authors – so this publication can be edited by all of them and once validated it appears in the personal publications lists of every author.

As mentioned before, the authors have the choice whether they want to transmit their publication data to one of the available gated systems or not. To this purpose we implemented 4 simple yes-or-no radio buttons: Data publication to the FUB website, data transmission to Cineca (i.e. the MIUR database), bibliographic data transmission to BIA and full text transmission to BIA.

If the publication was a collaborative work of several internal authors, the first who inserted the data is held responsible for the correctness of the data, for the permission to transmit them to BIA and to publish them in the FUB Website. Unfortunately this kind of workflow based on the first submitter cannot work for the transmission to the Cineca/MIUR database, where it is absolutely mandatory that every single author decides autonomously if he/she wants to list a publication in his/her account or not. In this case even though a publication entry can still be linked to different authors, it's each internal author of the publication, and not just the first who inserted it, responsible to authorize its transmission to his/her account in the Cineca/MIUR database.

6. Lessons learnt and open issues

The path leading to integration of so many different systems was all but easy and flowing. The official beginning of the BoRIS project was January 3rd, 2012. The implementation of the CRIS started with an older version of the software, Converis 4, knowing that a much improved version was about to be delivered. The whole year was therefore spent dealing with workpackages such as templates, roles and rights, workflows, label translation into Italian, being aware that all of this adjustments would later be imported in the newer version. Only in January 2013 the new Converis 5 arrived and it took us up to September 2013 to perform a deep customization and run an import of research projects and legacy Cineca publication data.

A problem we are actually still dealing with is the finalization of the integration of the Cineca web service. Above mentioned translation into German of the documentation for the web service was an additional work to do. But even with a translated documentation, the proper configuration of the XML transaction for the Web service is taking more time as expected.

We took also in consideration that in every community any change and innovation in consolidated procedures and workflows is considered with suspicion. To achieve a faster and easier acceptance of the new system we decided therefore to feed it with project and publication data produced by researchers of the FUB since its existence (i.e. in the last 17 years) before committing it to the research community. Publication data were imported from Cineca database and we are confident that this valuable critical mass of research outcomes easily searchable and reusable in the system would improve its acceptance. However this effort also meant a lot of additional work for the implementation team.

As for the institutional archive, we underestimated the issues related to privacy and copyright and their impact on the effort required for the validation step. Definition of more comprehensive and persuasive policies for submission to the Institutional Archive going beyond statements of principle is also something we have still some delay in dealing with.

7. Further developments

Achieved integration between BoRIS and BIA make further developments possible.

First of all, in DSpace the OAI-PMH Data provider module features out of the box support for the MPEG-21 Digital Item Declaration Language (DIDL) metadata format[13] which is a technical requirement for the automatic deposit of PhD thesis to Legal Deposit / Deposito Legale project at the Italian National Library of Florence[14]. Since PhD theses are completely integrated as research outputs in BoRIS and they will be seamlessly published in DSpace, we plan to rely on DSpace to deliver them automatically to Deposito Legale, thus dismissing the cumbersome and time consuming manual delivery of the digital version via cd-rom sent through the postal service.

Another possible improvement could be a tighter integration with PIS. Since research project budgets are planned and approved in BoRIS, as a next step an automatic transfer to PIS budget management module should be implemented, thus making the connection between BoRIS and BIA bidirectional.

Avedas developed another module for the Converis CRIS, the “Research Analytics Tool”. For the second half of 2014 we plan to implement this tool for our data model and use it to generate specific templates for the different institutional and personal reports and analytical requests.

Last but not least, we are fine tuning the functionality for the automatic generation of Curriculum Vitae. Having already imported in the system a valuable amount of bibliographic references, we are confident that this feature will be greatly appreciated.

8. Conclusions

Even if quite time consuming, and even at the cost of some delay, we believe that the level of interoperability achieved and described so far is worth the effort. Our CRIS is integrated with the main Single sign-on authentication System, with the Human Resources and the Financial Systems, with the Institutional Repository and also the interface with the national mandatory Cineca database is very close to switch to production. We are confident that having made interoperability and heavy reuse of existing data the main requirements right from the beginning will make BoRIS and BIA valuable tools not only in theory, but also real good friends in the researchers’ everyday activities.

Acknowledgements

We would like to thank Monika Stufferin (Head of Research Office of the Free University of Bozen-Bolzano) and Paolo Buoso (Deputy Head of the Library of the Free University of Bozen-Bolzano) for their support and helpful suggestions.

References

1. MIUR CINECA Database, <login.miur.it>.
2. DSpace, <http://www.dspace.org/>.
3. Avedas Converis 5, <http://converis5.com/>.
4. Common European Research Information Format, <http://www.eurocris.org/Index.php?page=CERIFreleases&t=1>.
5. Shibboleth < <https://shibboleth.net/> >.
6. The Open Archives Initiative - Protocol for Metadata Harvesting, <http://www.openarchives.org/OAI/openarchivesprotocol.html>.
7. Corporation for National Research Initiatives (CNRI) Handle System, <http://www.handle.net/factsheet.html>.
8. Registered DSpace Service Providers (RSP). <http://www.dspace.org/service-providers>.
9. Longsight, <http://www.longsight.com/>.
10. Simple Web-service Offering Repository Deposit v.2, <http://swordapp.org/sword-v2/>.
11. Jisc Sherpa Romeo, Publisher copyright policies & self-archiving , <http://www.sherpa.ac.uk/romeo/>.
12. Metadata Object Description Schema (MODS) <http://www.loc.gov/standards/mods/>.
13. MPEG21-DIDL <http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm>.
14. Depositolegale.it, Servizio nazionale coordinato di conservazione e di accesso a lungo termine per le risorse digitali, <http://www.depositolegale.it/deposito-legale-digitale-delle-tesi-di-dottorato/>.