IRIS: Supporting & Managing the Research Life-cycle

Andrea Bollini, Michele Mennielli, Susanna Mornati, David T Palmer*

International Business Development Unit, CINECA, Italy

Abstract IRIS is a new Current Research Information System (CRIS) developed by Cineca to upgrade and replace two previous solutions that have been used by Italian universities in the last 10 years. At the end of 2015, sixty-three Italian institutions are using IRIS. One of the main components of IRIS is DSpace-CRIS, an open source solution that can also be used as a standalone system, and as such is now installed in seven institutions outside Italy. The many solutions that IRIS provides include support for national research assessment exercises, institutional repositories (IR) and an expert finder. Both IRIS and DSpace-CRIS are compliant with the Common European Research Information Format (CERIF), allowing easy interchange and benchmarking with other CERIF compliant systems. New IRIS functionality includes Open Researcher and Contributor ID (ORCID) push / pull interoperability. IRIS has five different modules integrated together through standard protocols and interfaces.

Keywords CRIS, DSpace-CRIS, CERIF, Open Source, Reporting, ORCID, Research Assessment Exercise

1. Introduction

There are many players working within the research domain, at various points in the research life-cycle, each with different needs and purposes: researchers, librarians, finance officers, governance, public relations officers, etc. A Current Research Information System (CRIS) system can answer some or all of these needs and purposes by enabling one-time data input from all players, which along with data interchange amongst the institution's existing systems, can be re-used multiple times for many different purposes. In this manner, projects and budgets can be tracked, information and publications can be promoted and disseminated, and regular and ad-hoc reports can be generated and supplied just-in-time to decision makers, research evaluators, potential research collaborators, and desired funders.

IRIS (Institutional Research Information System) is a CRIS developed by Cineca, to replace two earlier CRIS systems also developed by Cineca: U-GOV Research and SURplus. A not-for-profit consortium, Cineca was founded in 1969 as the supercomputing centre of Italy. Now with 70 universities, four research institutes, and the Italian Ministry of Education (MIUR) in the consortium, Cineca is the premier developer and provider of academic IT systems in Italy. With 900+ staff, Cineca has developed and now maintains 500+ applications for its consortia members; library systems, HR, Financial systems, business intelligence, student information systems, and research information systems.

Cineca is an active member of many international organisations for sharing knowledge and developing technology standards. Within the research ecosystem, Cineca is a board member of euroCRIS (developer of the CERIF standard, described below), a member of the DSpace Steering Group, an active player within the DSpace community with two Cineca DSpace committers, and an advisory board member of Pentaho (described below). It is also the OpenAIRE National Open Access Desk for Italy and South Europe.

Cineca began a massive migration for all its managed CRIS systems to IRIS in 2014, which concluded in November 2015. Forty-six Italian universities previously using U-GOV Research, eight universities using SURplus, and nine new ones are now installed with IRIS, for a total of 63 Italian universities [1]. By the first semester of 2016, five additional Italian Universities will adopt IRIS, for a total of 68. The different environments and migration procedures for each site necessitated a large team of system analysts, developers, project managers, consultants, coordinators, and trainers. They held 100s of meetings with the institutions, video tutorials, and webinars, and created 100s of pages of documentation.

1.1. DSpace-CRIS

Parallel to IRIS development was the development of DSpace-CRIS. This was the result of six year collaboration between Cineca and the University of Hong Kong (HKU) [2], and later with the Hong Kong Polytechnic University (PolyU). In response to a 2009 HKU initiative of Knowledge Exchange, the HKU Library designed specifications to enable its institutional repository (IR) then only holding

publications in DSpace, to also provide visibility on HKU research, researchers, and their expertise. Cineca then architected a solution to extend DSpace, allowing the addition of CRIS research entities besides publications, such as people, projects, patents, organizations, etc. All of these entities are integrated, and thus make it easy for local operators to create new displays and reports. Common European Research Information Format (CERIF) compliance and Open Researcher and Contributor ID (ORCID) push / pull (described more fully below) were also added. The resulting product was named DSpace-CRIS, and is available in open source [3]. It is installed in several institutions around the world [Appendix A]. All of the functionality of DSpace-CRIS is now embedded in IRIS.

2. IRIS

IRIS is a best-of-breed CRIS solution based on JAVA technology. It has 5 different modules that interoperate through standard protocols and interfaces such as REST/SOAP. Single Sign-on (SSO) amongst the different modules, as well as with other institutional systems, is assured via an embedded Central Authentication Service (CAS). The SSO can use the regular institutional login whether based upon LDAP, Shibboleth or other scheme. It can provide local accounts as well as allow external players, such as referees, funders, etc.The platform architecture is divided into three different levels following the Model-View-Controller (MVC) paradigm.

The modular nature of the system and the flexibility of its data model facilitate processing, organising and transmitting information in accordance with the international CERIF standard. The compliance with CERIF strengthens relationships amongst the research entities (people, resources, activities, publications, etc.), by providing role-based, and temporally-bound semantic characterization of the various relationships. Locally, IRIS provides the opportunity to easily apply institutional and national requirements.

For reporting and statistical analysis, IRIS uses the following modules:
- Mondrian, an OnLine Analytical Processing (OLAP) engine,
- Saiku, an interface for OLAP cubes analysis,
- Pentaho Data Integration (PDI), an Extract Transform Load (ETL) engine, and,
- Jasper Report + Ireport, to create and use reports in XLS, PDF, Word, etc.

Cineca is committed to updating IRIS as new desired functionalities are identified. Building on open source lessons learned with DSpace and DSpace-CRIS, Cineca is seeking a proper method and time to release IRIS source code in open source. Open source will allow developers at each site to create new solutions with which every other IRIS site can then benefit.

2.1. IRIS Modules

2.1.1. Resource Management Module (RM)

The RM collects all the data needed to populate the "world" of research resources, such as researcher details, research teams, organizational structures, internal and external laboratories, instrumentation, scientific committees, grant funding, etc. Interfaces can be individually constructed to extract and load this information from each internal and external database, depending on the nature and protocols of each database, and data to be extracted.

2.1.2. Activities & Projects Module (AP)

The AP collects information on projects, contracts, collaborations, scientific initiatives, and all activities related to the business aspects of research, from proposals to final results. After this data is collected, AP allows the entering of new data to highlight the scientific value, partnerships and collaborations related to a research project. It also allows connections to be made, automatically or manually, between these entities and other entities such as people, groups, equipment, etc. It allows the customisation of workflows by assigning authorizations to each user depending on the project and the role of each user.

2.1.3. Institutional Repository / Open Archive Module (IR/OA)

The IR/OA allows for the storage, consultation and enhancement of publication outputs reflecting the various activities of a university. With this tool, the university has a unique and interoperable system, able to communicate with, and retrieve information from, national and international databases for the management and dissemination of publications. The module complies with the requirements of MIUR and the European Commission for Open Access (OpenAIRE).

2.1.4. Evaluation & Review Module (ER)

The ER supports the process of research and scientific evaluation based on data analysis and performance indicators. It allows the management of internal evaluation cycles and supports evaluation initiatives promoted by national bodies, such as MIUR, controlling research assessment exercises (RAE).

2.1.5. Expertise & Skills (ES)

The ES module supports the enhancement and promotion of the University's competences and characteristics. It is open to the public for consultation and it allows institutions to promote cooperation and exchange with industry and other partners.

3. Discussion of Functionalities

The various functionalities of, and responsibilities for research management have been traditionally undertaken by

several small autonomous units distributed throughout the university, which most times did not communicate or share data one with another. These different units entered and tracked the research information necessary for only those particular reports for which they were responsible. This meant that the same information was being re-entered by multiple offices in varying forms of completeness, with some information not entered at all. This produced a research environment of much overlap, redundancy, and sometimes costly repetition as the institution commissioned again in ignorance, research projects that had been done years ago by researchers now long gone. In general a CRIS is meant to provide an enterprise-wide solution for this situation. It allows each office to enter new data, or add data to existing records, integrating legacy systems to reuse and enhance existing information. It then can produce -- amongst many other products -- reports specialized for each office, and decision support for university governance. In this manner, a CRIS can preserve IT investments, act as institutional memory, and maximise return on investments (ROI).

With IRIS, Cineca has now considerably extended the CRIS model and its functionality. The experience of creating and maintaining two previous versions of CRIS, collaborating and architecting DSpace-CRIS with international partners, sitting as advisors on several international standards boards, and taking feedback from, and giving user education to all players in the research ecosystem (researchers, finance officers, project managers, librarians, funders, etc.) has uniquely enabled Cineca to do this. This extended functionality allows full description of all research entities (people, publications, patents, projects, funding, etc) and is described below.

3.1. Institutional Repository (IR)

IRs are traditionally created and maintained by the university library, for the purpose of promoting open access on the university's publications. While most other CRISes leave the IR as an autonomous standalone and simply extract some data from it to re-use in the CRIS, Cineca's IRIS thoroughly incorporates the IR for a much tighter and complete integration. This creates a win-win situation for librarians – who may still manage the IRIS IR/OA to promote open access – and other research players such as HR, finance office, governance and public relations, who rely on publications and other research data to inform their own reports and decisions.

The IRIS IR/OA module, integrated with other IRIS modules, can show richer information than traditional IRs. Authors can be shown, and searched for by, academic qualifications, languages spoken, committee appointments, professional society membership, etc. Publications can be shown with the grants that sponsored the research, associated patents, etc. IRIS further augments the publication record by automatically linking to external databases such as Scopus, PubMed, Web of Science and CrossRef, to retrieve and show corresponding bibliometric data (citation count, Impact

Factor, etc) in the local record. These bibliometrics can then also be included in decision support for HR and governance. Publications shown and interlinked within this wider context of the total research ecosystem, receive more search hits, higher readership, and most likely more citations. Cineca engineers regularly update algorithms of search engine optimisation (SEO) for publication metadata to ensure these publications appear at the top of the result list of search engines such as Google.

3.2. Research Assessment Exercise (RAE)

Countries and institutions in mainly Europe and Asia regularly carry out RAEs, with many other various names (e.g. Italy's Evaluation of Research Quality - VQR), to determine past ROI, and future levels of funding support and resource provision to each faculty, department, and research team. In general an RAE asks that a researcher, her research team, and department describe their best research outputs and the impact these outputs have made. As such they are of premier importance to researchers, and all those in the research ecosystem. Unlike most other CRISes, Cineca's IRIS provides a module, the ER, designed to support an institution's, and its researchers' need to respond to their country's call for an RAE.

As a measure of impact on the research output, the IRIS ER automatically matches publication Digital Object Identifier (DOI) or International Standard Serial Number (ISSN) to find the correct Impact Factor, SCImago Journal Rank, or other bibliometric. Other measures of impact, as well as the classification system used by the RAE for each field of discipline of each research output, can be added by the researcher or her research team in the ER. The ER can also be configured to support the different workflows prescribed by an RAE, differing by format of research output, discipline, external co-authors, co-authors in different faculties, etc.

The ER enables drag and drop to form columns and rows on most categories of data collected in the other IRIS modules, to instantly configure a custom view or report in the ER, e.g.,
- table showing number of publications in each department by year,
- line chart showing increase/decrease of citations by each researcher in a department over a certain range of years, and,
- pie charts showing percentage of each type of publication (journal, book, book chapter, etc.) for each department.

Although the ER is now optimised for reporting to Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca (ANVUR) --an agency within MIUR responsible for research assessment -- because of the ER's report generating flexibility, the ER can be used for other research assessment in other countries, and for many other purposes besides RAEs, such as preparing supporting

material for grant applications, benchmarking one department to another, one institution to another, etc.

3.3. Find an Expert

The need for and conceptualization of a CRIS began initially in offices of pro-vice chancellors for research, which needed to quickly make sense of data disparately reported to them by their various departments and centres, for the purpose of providing confidential decision support to the senior management teams of their respective universities on funding, hiring & firing, peer competitiveness, etc. In this regard, providing a public -- i.e., non-confidential -- portal on this same data was just not on the table. Therefore the academy's public relations office, and media relations officers around the campus would frequently scramble to make their own databases, sometimes called, "Find an Expert", to search for, and allow public searches for, an expert in the academy that could speak or write "ex cathedra" on a given topic to newspapers, television, etc. Although this content was entered into these databases by the PR office usually by asking each researcher to contribute, it was a subset of information already held by the HR office, PVC Research, IR, Finance Office, etc. After initial success, ensuring continual completeness, accuracy, and update were extremely problematic.

Unlike traditional CRISes, Cineca's IRIS provides a public portal with the completely integrated module of ES. This allows information to be entered in one go by local staff in the other IRIS modules, and re-used again and again for many purposes, some as yet unknown. Each site with IRIS can determine which information is confidential and which can display in ES.

ES allows the institution's PR office and the public to search on any attribute of a scholar, such as degrees held, languages spoken, committee positions, as well as keywords in publications, grants applications, patents, etc. Because ES metadata is optimised for every search engine, the ES search function, or popular search engines such as Google can also be used to search on this ES metadata to find, besides media spokespersons,

- candidates for contract research in government or industry,
- potential collaborators in new research papers and projects,
- potential supervisors for new PhD students in a given subject, etc.

The fortunes of academic institutions rise and fall depending on their reputations. A winning reputation can attract the best staff, larger funding, and more offers of research collaboration and contracted research. Parents and students often choose institutions with high reputations, and yes, high ranking in league tables. It is therefore imperative for an institution to take charge and manage its own reputation, rather than leave it to the vagaries of the random, and many times, incomplete and inaccurate reporting of third parties. The reputation of an academy chiefly rests in the reputation of each scholar in that academy. Managing the academy's reputation then means for the most part, ensuring that the digital footprint of these scholars is accurate and as large as possible. The IRIS ES module enables digital reputation management and enhances impact with full and rich granular data on each scholar, and metadata optimised and exposed for each search engine.

3.4. ORCID

Research now happens in a global context, with many distributed players: research institutions, publishers, funders, patent offices, etc. For exactness of identity, retrieval and attribution, publications have digital object identifiers (DOI), and patents have unique identifiers assigned them by their respective patent offices. A large missing piece in this paradigm has been unique identifiers for authors. The problem is manifest in lost or mis-attribution of credit because of two people having the same name, variant spellings, legally changed names, variant Romanizations of non-Roman scripts such as Chinese (漢字), etc. ORCID (Open Researcher and Contributor ID) [4]is an international non-profit initiative to provide unique identifiers to authors, and therefore disambiguate their data one from another. It is increasingly being used by publishers, funders, patent offices, and universities to correctly identify authors and thus give proper attribution of credit for research outputs. Many of these agents now require the researcher to have an ORCID before her paper can be published, her application for grant be funded, etc. The ORCID registry can also hold information on research outputs, grants funded, education, and employment for each researcher, after it is entered by the researcher or, received from publishers, funders, or employing institution.

ORCID push / pull is the latest functionality to be added to IRIS. When complete, this new functionality will enable,

- a researcher with one click to give permission to her institution and ORCID, to create her ORCID record, and populate it with data from the local IRIS. IRIS will then do so automatically using the ORCID Members' API. The scripts could write all data in one go, or give the researcher choice on which items to write or not. (Push).
- an institution to retrieve new data from ORCID records, and write to corresponding researcher records in the local IRIS. Scripts will automatically run for this, and then display results to individual researchers for them to accept or reject. This would require the institution to have an ORCID membership allowing usage of the ORCID Members' and Premium APIs. (Pull).

In Italy, ANVUR, with the support of the Conference of Italian University Rectors (CRUI) and Cineca, launched the Italian Research IDentifier for Evaluation (I.R.ID.E) project [5], which has adopted ORCID as the national research identifier. This will require all those who operate in Italian

research to create an ORCID record, for use in ANVUR's VQR. Future extension to other research and educational purposes are now being considered. For this project, Cineca created an ORCID Hub, which is now integrated in IRIS. Cineca has now released this Hub software as open source on GitHub [6] to facilitate other countries' efforts with ORCID.

When locally installed, this ORCID Hub allows easy and one point management for researchers to login, and de/authorize their local profiles for use in several applications that can interact with ORCID. This will then allow with the use of a spreadsheet the research office in the university of those researchers, to quickly manipulate data to be uploaded to ORCID, or extracted from ORCID. Another important feature of the ORCID Hub is the ability to monitor, using the ORCID Premium API, all the profiles of Italian researchers. When new data becomes available in an ORCID profile, this data is automatically propagated to the ORCID profiles of the other Italian researchers involved (co-authors, co-investigators, etc.). IRIS profiles of the co-researchers will then also be automatically updated using the ORCID notify service.

In Italy, the personal ORCID identifier now replaces the usage of MIUR identifiers and logins, as well as the Italian tax code ID, for researcher identification. Outside Italy, many countries have similarly adopted, or are in the process of adopting ORCID as the national research identifier. These include Sweden, Denmark, Australia, United Kingdom, Finland, Spain (Catalonia), Austria (FWF), Portugal and Norway.

With all researchers having a unique ORCID that will follow them throughout their careers, many new activities become possible. IRIS can allow external collaborators of local researchers, to login with their ORCID to edit their data[7]. IRIS scripts can track researchers and their research projects in new employment, after they leave the local employment. In this regard, ANVUR will begin tracking the research projects and employment of PhD students three years on after graduation, to determine the impact of Italian PhD education.

3.5. Serendipity

With the addition of all of the above new functionality, IRIS can provide solutions that were not initially planned. For example, in the past faculty office staffs were called upon to search on many different internal and external sources, and tediously compile results for the following purposes. However local IRIS operators can now setup scripts to provide these reports and displays regularly or ad hoc.

- The Dean wishes to give an award to those researchers who have published in peer reviewed journals with an impact factor above "x", and with citations received above "y".
- Visitors from "x" university to the faculty wish to meet with researchers who have studied at, or worked at "x" university.

- The Library wishes to identify local editors of journals, to give orientation on the merits of Open Access.
- New researchers in the faculty want to know the major research topics of research projects and publications done in the faculty.
- The faculty wishes to have each researcher make a web page showing publication lists, research projects etc.
- The PVC Research wishes to know the total USD amount of grants received by faculty researchers in the previous year.
- The PR Office wishes to make a press release of all the awardees of honours, awards and prizes in the previous year.
- The PVC Research wishes a list of overseas universities with which faculty "x" researchers have collaborated.

4. Conclusions

In every developed country, the amount of money spent on research, and the expectations that it deliver quality results are huge. Rather than leave research management, evaluation, impact and reputation to haphazard and random execution, the institution is better served with an enterprise wide solution that will maximize ROI in every respect. Cineca's IRIS can provide this solution.

Appendix

Sites Installed with DSpace-CRIS

- University of Hong Kong (HKU)
- Hong Kong Polytechnic University (HK PolyU)
- Hong Kong Institute of Education (HKIEd)
- Spanish National Research Council (CSIC – Consejo Superior de Investigaciones Cientificas)
- Consorci de Serveis Universitaris de Catalunya (CSUC)
- CSUC uses DSpace-CRIS to collect research information from its 10 member universities
- PTCRIS
- Portugal national project collects information from more than 40 universities. DSpace-CRIS will be installed in 2 pilot Universities to test the upgrades of all repositories.
- Autonomous Region of Friuli-Venezia Giulia
- Implementing a DSpace-CRIS based portal to collect information from five different institutions.
- 63 IRIS systems of Cineca
- DSpace-CRIS embedded in each IRIS system

REFERENCES

[1] List of Italian institutions installed with IRIS.

https://wiki.u-gov.it/confluence/display/UGOVHELP/List+of+IRIS+Installations+-+Elenco+delle+installazioni+IRIS.

[2] Palmer, DT., Bollini, A., Mornati, S., Mennielli, M. (2014) DSpace-CRIS@HKU: Achieving Visibility with a CERIF compliant Open Source System. Procedia Computer Science. 33, 118-123. http://dx.doi.org/10.1016/j.procs.2014.06.019. http://hdl.handle.net/10722/198431.

[3] https://github.com/Cineca/dspace-cris.

[4] http://orcid.org/.

[5] I.R.ID.Ehttp://www.anvur.org/index.php?option=com_content&view=article&id=829&Itemid=600&lang=it.

[6] https://github.com/Cineca/OrcidHub)

[7] This is now enabled in euroCRIS DSpace-CRIS. http://dspacecris.eurocris.org/.