THE INSIDE OUT LIBRARY: CRO OPENDOCUMENTS DIGITAL INSTITUTIONAL REPOSITORY

Mauro Mazzocut[1], Jordan Piscanc[2], Nicolas Gruarin[1], Roberto Ricci[3], Romano Trampus[2], Ivana Truccolo[1]

[1] *CRO Aviano National Cancer Institute, Scientific and Patients library, Italy*

[2] *University of Trieste, IT Services, Italy*

[3] *CRO Aviano National Cancer Institute, IT Services, Italy*

Corresponding author: Mauro Mazzocut, mmazzocut@cro.it

Introduction

CRO OpenDocuments (1) is the new institutional repository of CRO National Cancer Institute IRCCS. The repository is developed in collaboration with the University of Trieste, owner of the open repository OpenStarTS (2). CRO OpenDocuments is the first institutional repository in Italy for an IRCCS (Istituto di Ricovero e Cura a Carattere Scientifico).

IRCCS are typical Italian health institutions that combine basic research, population sciences, clinical research and patient care and continuing medical education to promote enhancement in prevention, diagnosis, therapies and rehabilitation of different kind of diseases. In Italy there are 49 IRCCS focused on different health fields: 21 public and 28 private(3). CRO is a public IRCCS focused on cancer research and cure, and it is located in the North East of Italy.

Institutional repositories are well known in the Italian academic field. Since 2001 the implementation of OA institutional repositories has constantly increased. In 2004, rectors of 30 Italian universities signed the so called "Messina declaration" (4). This document, inspired by the "Berlin Declaration" (5), committed the Italian universities to foster the diffusion of Open Access policies and publishing, and encouraged the implementation of open institutional repositories. Some of them have been funding under regional spending, included the institutional repository of University of Trieste. Generally speaking, today almost each Italian university has its own institutional repository (6).

If we focus on the biomedical field, the scenario will be different.

Searching in OpenDOAR (7) and ROAR (8) directories for Italian Health and Medicines disciplinary repositories, only 7 over 75 Italian repositories can be found. This result is biased, because the content of many university repositories are described as "multidisciplinary", hiding those of the Medicine faculties or departments. Except from universities, there are only three health research institutes owning their own institutional repository: E-ms Eprints Open Archive in Social Medicine and related fields (which is currently not available), OpenPub (the research registry of the Edmund Mach Foundation) and DSpace ISS, the institutional repository of the Italian National Institute of Health (ISS).

This last institution was a pioneer in open access for the health and medicine field. Italian National Institute of Health has developed its own repository in 2006, and has been the first Italian Institution to adopt an internal open access policy in 2008 (9). At the same, time they made the first attempt to coordinate the implementation of a national repository for biomedical scholarly works produced by Italian health research centres.

This attempt was preceded by a pilot survey (10) that explored the system used by Italian cancer research institutes to archive their institutional research outputs. The results published in 2009 showed that existing repositories were mostly implemented using MS Excel or, sometimes, citation management software such as Refworks or Reference Manager. Rarely they were available online. An XML metadata schema based on Dublin Core metadata standard and a submission workflow were then developed by ISS. This aimed to allow partner institutions to supply their data and make them available online via DSpace ISS. Unfortunately, this initiative to create a national repository of the Italian biomedical research was unsuccessful. After this first attempt, nowadays no IRCCS has developed its own open institutional repository.

Together with lack of a government mandatory, the 2009 survey identified another threat to the successful development of open institutional repository in biomedical field. The existent repositories were mainly intended for evaluation purposes in view of the annual activity report and for assigning funding to research investigation. But these archives were not properly used for their information richness meant to provide high visibility to the scientific literary or to search for scientist competences and specializations (10). And this is particularly true for all the non-scientific publications produced by biomedical research centres. In fact, because of their translational nature, research centres like IRCCS generate knowledge expressed in many other ways: learning materials of Continuing Medical Education addressed to professionals; free press, multimedia and educational resources addressed to patients; technologies available for commercial use, an so on.

CRO Scientific and Patients Library has developed different repositories overtime, to fit with all these different needs. Altogether, these repositories fit with their own purpose and represent the overall knowledge developed at multiple levels within our institution. However different software, metadata standards and online publishing strategies have been adopted over the years, and a lack of overall integration and interoperability finally emerged.

Every repository has its own access point, but displayed in different parts of our institutional website. Some publications are exclusively available via institutional website, conversely some are not. Some publications are published using a online publishing platform such issuu.com. Sometimes archives overlap. Furthermore, different metadata standards have been used and sometimes their quality is low. This caused an underrated access and use of the resources, an objective difficulty to measure the overall impact of our translational research, and also affects the persistence of the resources and their metadata overtime.

Material and Methods

CRO OpenDocuments project was born to face a problem of persistence: the shutdown of our grey literature archive due to obsolescence of the CDS ISIS software. The grey literature archive collects the learning resources used for the Continuing Medical Education since 2004, and made all these resources available to the courses attendants.

In our Institute quality paths (ISO 9001:2008 and Accreditation Canada), is up to the library to warrant the online availability of the learning resources.

We replaced the CDS ISIS software with an installation of DSpace-CRIS customized by CINECA and made available by the University of Trieste. DSpace-CRIS features enhance the metadata standardization (Dublin Core), manage, collect and expose data about all the Research aspects (papers, journals, people, organization units, prize, project and grants). Following the OpenAIRE Guidelines the interoperability with other archives is granted using international standards (OAI-PMH, CERIF, OpenSearch, OpenUrl, Orcid, DOI, etc.). DSpace CRIS module offers, eventually, also a connection with Altmetrics.org and a set of usage statistics that allows to measure the impact of a single resource or a single entity such as authors, units and departments, grants, and so on.

Content type

Since a need of rationalization of the different repositories emerged, we decided to exploit the DSpace features to gather all scientific and educational resources collected over the years in different repositories, and make it available in a unique digital environment. We identified four existing archives to be migrated in CRO OpenDocuments:

- **Grey Literature Archive**: it collects over 1700 didactic resources (mainly slides in pdf format) of the Continuing Medical Education courses and 67 graduation or PhD thesis which have been conducted at CRO from 2004 to present. These latter are available only as metadata description, because we don't have a digital version of the thesis. Data have been exported from CDS ISIS in ISO 2709 format. Most of the resources are full-text, but the authors may decide to make resources available only to users of CRO, consequently the full-text is accessible only through the institutional intranet.

- **Scientific Publication Archive**: This archive collects the scientific production of the Institute since 1994, there are about 7500 records of journal articles, conference proceedings and unpublished communications at conferences and congresses. Full-text are available since 2003. At the moment the Scientific Publication Archive is powered using the bibliographic management software Reference Manager. Reference Manager has been deeply customized in order to manage the reporting of scientific productivity of the Institute and to allow the measurement of the Impact Factor in accordance with the Italian ministerial directives. For this reason, the archive of scientific publications will continue to be managed with Reference Manager, and will be exported each month to CRO OpenDocuments.

- **Technology Transfer**: one of the most important output of the scientific research is the transfer of technologies and facilities to the market. CRO has a Technology Transfer Office with the purpose to strengthen the collaboration with private companies by moving results of translational research into societal use. The Technology Transfer Office has drawn up descriptions forms of technologies developed by CRO scientists and available for commercial use or joint development. Nawadays twenty-six descriptions forms in pdf format are available, for technologies such as: diagnostic tools, e-health home automation facilities, antibodies, drug dose calculators, patient support systems, and so on.

- **CROnews Archive**: the archive contains 534 articles of CROnews, free press quarterly magazine of information on the CRO institutional activity since 2007, addressed to patients and citizens in general. It is a very important and appreciated way to inform people about the life inside our Institute, the initiatives taken in it, the people who work in it, the services offered to patients and their families by voluntary associations. Currently the archive is hosted on Refworks provided by Bibliosan, the Italian consortium of health & biomedical libraries. The database is available from the institutional website, but it has important limitations in the customization of its features. When the data migration will be completed all the full text will be freely accessible through CRO OpenDocuments. CROnews

To these pre-existing, we added two new archives.

- **Farmaci & Tumori (Drugs & Cancer)**. Farmaci & Tumori is the first independent bulletin on drugs used in the treatment of cancer in Italian language. It is directed at both patients and citizens and health professionals. Contents are based on the application of evidence-based medicine methodology. Farmaci & Tumori is funded exclusively by funds for scientific research and donations.

- **Patient Education & Empowerment:** The CRO Patient Library (1998) is a pilot project lead in Italy in the cancer patient information field. In 2010, CRO Scientific Directorate established a Patient Education & Empowerment Group (PEEG) charging the health librarian with the role of technical team and programme coordinator. The PE programme is defined by different activities as relating to three specific areas: research, education and information and communication. Moreover these are centred around the Scientific and Patient Library. Other services involved in these activities include the Continuing Education Office, the Pharmacy and all the health facilities. The PEE activities include:

    - Classes: hour-long meetings where professionals and volunteers, as teachers, speak to other patients and their relatives about relevant health topics during the daily activity hours in the hospital setting.
    - Patient education handouts about relevant topics are written by health care workers but are accurately revised by a subgroup of the PEEG, comprising psychologists, librarians, drug experts, patients and laymen.
    - Narrative Medicine Programme: it is an approach to medicine that recognises the value of patient narratives in clinical practice, research and education

Output of these activities are slides, handouts, booklets, books, bulletins.

Our Institute was also involved as project manager of a 3-year-long multi-centre collaborative project (2013-2015) granted by the National Health Authority in Italy, entitled "Extending Comprehensive Cancer Centers expertise in patient education: the power of partnership with patient representatives". This project involved eight Italian research centres and patient organizations. The output of this national project are multimedia educational resources addressed to patients, and also specific resources addressed to professionals who wants to develop or enhance PEE initiatives in their hospitals (11).

Archive Structure

DSpace organizes records in "community", "collections" and "items". A "community" is a group of "collections". "Collections", in turn, are groups of one or more "items". The "item" is composed of the attached resource – or beatstream - and its Dublin Core metadata description. "Community" and "collections" may contain respectively "sub community" and "sub collections" of lower hierarchical level.

Existing archives can be identified by instances "community." Any homogeneous subsets of records can be collected into "sub-communities". For instance, this feature allows the arrangement of events or Continuing Medical Education that are repeated over time in subsequent editions by resorting to a "sub collection". In this way we can preserve the specificity of the original files, and make them visible in the search interface. So it will be possible to cross search and to surf between different archives and to display the entire scientific and popular production by author, structure, department, laboratory, funds. Furthermore, the CERIF compliant features of DSpace CRIS module allow us to structure the database around some "entities" identified as: author, department, lines of research and grants. Every entity has its own set of descriptive information. For instance:

- "Author" Entity includes fields to enter the preferred name, variant names, biography, work groups, affiliation, research interests, ORCID and other ID's, mail, personal site, publication, and so on.
- "Department" Entity includes fields to enter organization name; description; director; researchers; projects; publications
- "Project" Entity includes fields to enter title; code; principal investigator; co-investigators; start and expiring dates; abstract; keywords

This will allow to have an overview of the impact of a research grant or fund not only on the scientific literature, but also on popular publications addressed to the population.

Policies

- **Submission policies**: Two trained librarians dealing with the implementation of the database and the consistency of metadata. A workflow with different cataloguing profiles and hierarchical levels of authority were implemented: the "light cataloguer" may submit a proposal for a resource input; "advanced cataloguer" can edit metadata; the "administrator" will check and authorize the proposed entry. He can also work on the creation and deletion of communities and collections. So, in the future it will be possible to allow the non-librarian personnel to enter their records, especially as regards the self-archiving of personal communications or educational resources. After the first submission made by the "light cataloguer", "advanced cataloguer" or the "administrator" will take care of validation and publication of the resource in CRO OpenDocuments, ensuring the quality of the bibliographic description.

- **Access policies:** All metadata description will be available online and exposed to the web. We planned several ways to access the full text, depending on the license available for each content. Full text protected by copyright - such as scientific articles - will be available only within the institute's intranet network. Free and Open Access licensed articles will be available online. Furthermore DSpace allow us to set a blackout period for Open Access publications subject to the embargo. Also, a part of the teaching resources used in the Continuing Medical Education will be accessible only through the institute's intranet. This is because some of the contents used are protected by copyright or preliminary results or innovative methods are discussed. The authors are required to sign a release for the online publication of the slide. On the other hand, CRO OpenDocuments will make freely available many other resources. At first the Technology Transfer descriptions forms, because of their own promotional purposes. Generally speaking, all the resources addressed to patients and laymen will be available online: Drug & Cancer free bulletin, CROnews quarterly free press, CROinforma booklets. The slides of the Patient Education & Empowerment classes may require a registration in DSpace. This depends on two reasons: the information are very tailored or otherwise need frequent updates; to activate a privileged channel of communication for patients attending the classes. The same will be for the resources addressed to the professionals of the Patient Education & Empowerment.

- **Preservation policies**: CRO OpenDocuments data are hosted by University of Trieste. The University adopt standard backup/restore and disaster recovery plans. Dspace software itself has some preservation tools like Checksum Checker, that check regularly the consistence of the archived material.

Results:

State of the art today is:

- **Grey Literature Archive**: the migration from CDS ISIS to CRO OpenDocuments is complete and the archive is fully operational.

- **Scientific Publication Archive**: we are involved in the migration of the previous years. Particularly we are testing now the migration of years 2014 and 2015. We used the Reference Manager Tab preset to export the bibliographic metadata

- **Technology Transfer description forms**: We aim to upload these forms into CRO OpenDocuments in the near future.

- **CROnews archive**: Archive has been exported from Refworks in End Note format; subsequently uploaded into Reference Manager and finally exported with the Tab export preset. We are currently checking the quality of the data, and testing the import on DSpace.

- **Farmaci & Tumori archive**: two numbers has been published and the full text is freely available. Drugs & Cancer is published with an Attribution-NonCommercial-ShareAlike 4.0 International licence

- **Patient Education & Empowerment**: some CROinforma series booklets have been uploaded. Future booklets will be licensed with Creative Commons.

Finally we are setting entities instances for authors, departments, grants and institutional research area.

Discussion: outcomes for the institute organization

We think that CRO OpenDocuments can have relevant outcomes for our institute organization.

From a management point of view, CRO OpenDocuments allows us to:

- Reduce the number of applications used: Refworks, ISIS, the web interface of Reference Manager, the subscription to the web publishing service ISSUU may expire.
- optimize the archives implementation and maintenance workflow
- independently manage the data through very flexible import and export facilities based on MS Excel (12,13)
- if necessary, manage a multi-centre cataloguing
- reduce the overall costs of the archives in terms of licenses, management, maintenance

From the point of view of the services provided, CRO OpenDocuments allows us to:

- enhance the visibility and accessibility of CRO's scientific and popular production
- make available validated educational and informative resources addressed patients
- expose standardize and rich metadata to different web services and archives such as Search Engines, Google Scholar, ORCID, Research Gate, and so on. We aim also to register CRO OpenDocuments in OpenAcces repositeories like: PLEIADI , ROAR, OpenDOAR, Driver, OpenAire, OAIster
- integrate the repository with the institutional website through permanent link
- manage multi-level access rights to resources
- display the updated scientific and educational output of each researcher and department
- access the whole knowledge produced by CRO scientists via a unique and responsive interface

Last, but not least, another relevant outcome will be the disclosure among professional of copyright and creative commons issues.

As seen before, CRO produces a lot of digital and traditional contents addressed to professionals and laymen, and it is also a publisher for many non-scholar publication. Often digital content published by the CRO in the past did not report any indication of copyright, which means that all rights are reserved. Most of the times, authors are not aware enough about the implications of the digital rights management issues: terms and condition on publishers contracts about the intellectual property of an article, risks of plagiarism or misuse of non scholarly contents by third parties, but also our author's misuse of other authors digital contents (such as pictures, graphics, and so on).

CRO OpenDocuments will be an opportunity for a redefinition of an institutional policy for Open Access and Digital Right management. We are now working on a document inspired by the ISS policy published in 2008 which will be published during the official debut of CRO OpenDocuments next September.

Conclusions

In answer to the findings of the survey of 2009 about the main threats to the development of open institutional repositories, CRO OpenDocuments was not born only for the purposes of the scientific output evaluation and funding assignment.

On the contrary, it has been developed in a "inside out" library perspective, exploiting the information richness of previous archives to offer an open display and access to all the knowledge developed by the CRO National Cancer Institute both devoted to the scientific and non-scientific community.

At the same time CRO OpenDocuments allows us to foster a deeper knowledge of the open access and digital right management issues among CRO professionals.

Besides the advantages, many issues are still to be faced.

Evidence in current literature suggests that the role played by repositories is still not adequately considered. Many institutional repositories are not being widely used by researchers with overall deposit levels remaining relatively low (14). For instance, OA articles were published as part of subscription journals published by scholarly societies (15). Often the green road places the burden of publishing costs basically on authors (16). According to Piwowar (17) the authors of studies on cancer and human subjects were least likely to make their datasets available. These results suggest research data sharing levels are still low and increasing only slowly, and data is least available in areas where it could make the biggest impact. Strong leadership in an organization combined with awareness of users about the practical utility of the tool and institutional advocacy programs promoting the use of repositories will be needed (14).

BIBLIOGRAPHY

1. CRO National Cancer Institute, University of Trieste. CRO OpenDocuments [Internet]. CRO OpenDocuments. 2016 [cited 2016 May 30]. Available from: http://opendocuments.cro.it/cod/

2. University of Trieste. OpenstarTs [Internet]. OpenstarTs. [cited 2016 May 30]. Available from: http://www.openstarts.units.it/dspace/

3. Italian Minister of Health. Istituti di Ricovero e Cura a Carattere Scientifico - IRCCS [Internet]. 2007 [cited 2016 May 30]. Available from: http://www.salute.gov.it/portale/temi/p2_6.jsp?id=794&area=Ricerca%20sanitaria&menu=ssn

4. Dichiarazione_Messina [Internet]. cab.unime.i. 2004 [cited 2016 May 30]. Available from: http://cab.unime.it/decennale/wp-content/uploads/2014/03/Dich_MessinaITA.pdf

5. Berlin Declaration [Internet]. Open Access Max-Planck-Gesellschaft. 2003 [cited 2016 May 30]. Available from: http://openaccess.mpg.de/Berlin-Declaration

6. Cassella M, Gargiulo P. Open Access in Italy. In: Anglada L, Abadal E, editors. OA report in Southern Europe [Internet]. FECYT; 2010 [cited 2016 Apr 15]. p. 63–82. Available from: http://eprints.rclis.org/15140/

7. University of Nottingham. OpenDOAR - Directory of Open Access Repositories [Internet]. OpenDOAR. 2006 [cited 2016 May 30]. Available from: http://www.opendoar.org/

8. EPrints.org network, University of Southampton. Registry of Open Access Repositories [Internet]. Registry of Open Access Repositories. [cited 2016 May 30]. Available from: http://roar.eprints.org/

9. Garaci E. Politica Istituzionale per il Libero Accesso alle Pubblicazioni Scientifiche. 2008 [cited 2016 May 30]; Available from: http://dspace.iss.it/dspace/handle/2198/352

10. Poltronieri E, Truccolo I, Di Benedetto C, Castelli M, Mazzocut M, Cognetti G. Science, institutional archives and open access: an overview and a pilot survey on the Italian cancer research institutions. J Exp Clin Cancer Res CR. 2010;29:168.

11. Truccolo I. Providing patient information and education in practice: the role of the health librarian. Health Inf Libr J. 2016 Jun 1;33(2):161–6.

12. Nash JL, Wheeler J. Desktop Batch Import Workflow for Ingesting Heterogeneous Collections: A Case Study with DSpace 5. - Lib Mag [Internet]. 2016 Jan [cited 2016 May 30];22(1/2). Available from: http://www.dlib.org/dlib/january16/nash/01nash.html

13. Silvis J. Batch Ingest into DSPACE based on EXCEL [Internet]. Jeff Silvis's Blog. 2015 [cited 2016 May 30]. Available from: https://web.archive.org/web/20150620112213/http://blog.lib.umn.edu/silvi003/codenotes/2010/04/batch_ingest_into_dspace_based.html

14. Pinfield S., Salter J., Bath PA., Hubbard B., Millington P., Anders JHS., et al. Open-access repositories worldwide, 2005-2012: Past growth, current characteristics, and future possibilities. J Assoc Inf Sci Technol. 2014;65(12):2404–21.

15. Matsubayashi M, Kurata K, Sakai Y, Morioka T, Kato S, Mine S, et al. Status of open access in the biomedical field in 2005. J Med Libr Assoc JMLA. 2009 Jan;97(1):4–11.

16. Abadal E. Gold or green: the debate on open access policies. Int Microbiol Off J Span Soc Microbiol. 2013 Sep;16(3):199–203.

17. Piwowar HA. Who shares? Who doesn't? Factors associated with openly archiving raw research data. PloS One. 2011;6(7):e18657.