

Setting-up an institutional repository of biomedical literature authored by Italian researchers

Elisabetta POLTRONIERI

Istituto Superiore di Sanità. Publishing activities, Rome, Italy

elisabetta.poltronieri@iss.it

Moreno CURTI

Policlinico S. Matteo, Pavia, Italy

curtim@smatteo.pv.it

Roberto RICCI

Centro di Riferimento Oncologico, Aviano, Italy

ricci@cro.it

Paolo ROAZZI

Istituto Superiore di Sanità. Data management, Rome, Italy

paolo.roazzi@iss.it

Albina RUMEO

Istituto Superiore di Sanità. Publishing activities, Rome, Italy

arumeo@iss.it

Ivana TRUCCOLO

Centro di Riferimento Oncologico, Aviano, Italy

itruccolo@cro.it

INFORUM 2006: 12th Conference on Professional Information Resources
Prague, May 23-25, 2006

Abstract. This paper aims to describe the creation of an institutional repository of scientific literature at the Istituto Superiore di Sanità (ISS, Italian National Institute of Health), the leading centre for biomedical research in Italy. This initiative is the core of a Project aimed at aggregating the published scientific works of Italian researchers in the fields of public health and related areas, as part of a programme of cooperation between the various biomedical research bodies that operate within the National Health Service. To this end OAI-compliant open-source software (DSpace) has been implemented at ISS and initially uploaded with data sourced from a bibliographic database that assembles internal scientific publications. Together with a library of publications produced by ISS as publisher, the repository will hold data on scientific literature authored by partner institutions such as Scientific Institutes for Research, Hospitalization and Health Care (Centro di Riferimento Oncologico di Aviano and Policlinico S. Matteo di Pavia). ISS is currently engaged in customising the Dublin Core metadata element set, also with a view to adopting other data sets in order better to identify the resources described in the repository. The objective is to gain access to scholarly publications within the framework of the Open Access environment and to programme the Institution's repository services.

ISS profile

Istituto Superiore di Sanità (ISS, Italian National Institute of Health) is the leading research organisation in Italy in all fields related to public health.

Three Departments of the Institute, the Library and the Documentation and Publishing activities departments, provide the necessary information support services for the Institute's research laboratories and offices, which in turn supply these structures with details of their separate activities. This system makes it possible to promote cooperation between multidisciplinary teams (librarians, documentalists, health professionals and researchers) so as to meet user needs and develop appropriate facilities.

As a publisher and owner of the copyright in all material published by the Institute (*Annali dell'Istituto Superiore di Sanità*, a quarterly journal also indexed in PubMed, *Notiziario dell'Istituto Superiore di Sanità*, a monthly bulletin, and series of technical reports), ISS already provides free online access to its publications through its web site (<http://www.iss.it/>).

The ISS research staff includes 700 scientists and the Institute's growing research output amounts to approximately 1600 publications per year, mostly journal articles. All works are indexed in a bibliographic database - based on a relational SQL/server database - which generates and updates the list of researchers' publications on the ISS web site. Bibliographic items recorded in the database are also included in the annual official report of ISS activity.

Steps towards adopting an Open Access strategy in ISS

The Open Access (OA) standard protocol¹ is widely applied in Italy, mainly by academic institutions which are making their resources available in open digital archives, in recognition of the role of such organisations as open access data and service providers.

In this fast-moving environment, where Internet and digitisation initiatives are closely merged, institutional repositories are rapidly becoming the privileged containers of research output. According to a recent study published by the European Commission², their number has been globally increasing over recent years in parallel with publishing in OA journals, another major outlet for freely available research literature, and with open-source software developments. By harvesting research results, open-access archives are a valuable tool for evaluating research in institutions.

Moreover, research-funding organisations such as the National Institutes of Health (NIH, USA) advocate that publicly-funded medical research should be deposited in OA archives³.

OA paradigm compliance (top-down approach)

In order to comply fully with the principles of the open access movement, ISS has recently signed the "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities" a milestone on the road to further declarations in favour of OA by national and international scientific organisations (152 signatories up to April 2006, including 71 Italian institutions, mostly universities)⁴. International conferences to discuss the OA principles established by the so-called Berlin Declaration are held annually⁵.

Exploring attitudes and perspectives regarding the OA paradigm (bottom-up approach)

In 2005 ISS launched a Project aimed at aggregating the biomedical resources (mainly scientific publications) produced by Italian research institutions.

This initiative represents a stimulus for the Institute to define an internal policy in accordance with the philosophy of free access to research results. The first step was to raise OA awareness among ISS research staff. To this end and in a similar manner to other initiatives undertaken at national and international levels in order to track authors' attitudes to OA issues^{6 7 8}, an internal survey was recently carried out to identify researchers' opinions towards the publication and distribution of academic material, following both the traditional and OA models.

The results revealed essentially that the highest response to the survey was received from more senior ISS scientists, mainly those responsible for leading research projects, who are also more closely involved in publication activities. The benefits offered by the creation of an ISS institutional repository have already been widely acknowledged within the research community. In particular: 97% of those surveyed declared they would deposit the published version of their works, while just 20% would archive pre-referred or referred papers; it is important to note that 54% agreed to deposit their works on condition that an formal ISS policy be pre-defined, while 49% would submit their works without any access or usage restrictions.

Towards the establishment of an ISS repository

Istituto Superiore di Sanità is already engaged in establishing an OAI-compliant institutional repository. This digital archive is designed to provide both data and services regarding primarily research articles published by ISS researchers. For this purpose efforts are being made to maximise the potential of the internal bibliographic database, which maintains a regular record of the research output of ISS researchers.

This in-house archive has been designed to provide a detailed description of each item entered (e.g. journal article, book chapter, congress paper, report). Therefore, besides the elements traditionally stored – e.g. author, title of the work, title of the publication, volume, issue, pages – additional data are entered in the database to identify a given paper more fully. Special importance will be given to each author's affiliation, making it possible to sketch a map of the different institutions that cooperate with ISS at national and international level in particular areas of research.

This process of identifying the co-authors of ISS researchers through their affiliations will enable an assessment to be made of the ability of ISS to attract resources by stipulating agreements with partner institutions for common research projects.

Further elements for resource identification envisaged by the internal database among its bibliographic fields are the MeSH (Medical Subject Headings) descriptors, the terms adopted by the National Library of Medicine (Bethesda, USA) for indexing and retrieving medical literature in PubMed. MeSH terms, as supplied by the published version of a work, are entered in the database and both English version and an Italian translation by ISS (<http://www.iss.it/site/Mesh/>) are available.

The process of rethinking the functions and objectives of the existing database within a wider framework will produce the following benefits for both the ISS community and a broader forum of Italian biomedical experts: visibility of research results resulting from the permanent storage of material held in the repository, a flexible submission process and an easy search interface.

DSpace software package customization

In order to set up its own institutional repository, ISS has adopted the DSpace open-source platform, version 1.3.2, and has installed it on a dedicated server.

DSpace (<http://www.dspace.org>) is a digital library system consisting of a database and a web interface developed jointly by Hewlett Packard and the Massachusetts Institute of Technology (MIT) of Cambridge (USA). It has been conceived to store digital information and index it by assigning descriptive metadata (titles, authors, identifiers such as ISSN and ISBN, etc.), to keep research material accessible and to preserve content in a safe archive (avoiding loss of data in the long term). The web interface is used for browsing, search, viewing and for submitting items. The web interface is used for browsing, search, viewing and for submitting items. DSpace user group (DSUG) meetings are regularly held to share experiences and gather the feedback of participants from individual institutions⁹.

ISS currently holds about 1350 items from a selection of full-text ISS publications on the DSpace platform. The next step on the agenda of the repository implementation plan will be the provision of public visibility on Internet of material held in the DSpace archive. All records stored

in the ISS bibliographic database will then be processed for gradual migration to the DSpace platform. This will allow metadata to be harvested and globally disseminated in order to increase the research impact of ISS literature production.

Technical staff of the ISS Data management sector are now addressing the problem of matching the metadata element set supported by the DSpace program with information recorded in the internal database. ISS is also developing alternative customised versions of the original software in order to prove that the repository will be able to interact with other IT applications (i.e. generate publication lists in different file formats suitable for compiling annual reports, project proposals, statistics of research output)

The issue is crucial for two reasons: on the one hand the application of standardised metadata (e.g Dublin Core) guarantees interoperability among OAI-compliant digital archives, but on the other hand the adoption of a pre-defined metadata set carries the risk of failure to respect the peculiarities of data recorded for describing the resources. A procedure will therefore be implemented to customise the metadata in order to include the identification and assessment of publications produced in partnership with ISS. In technical terms, the main efforts are directed towards enhancing the modularity of DSpace in order to include additional features. The tasks facing ISS staff therefore comprise testing and investigating solutions to create an integrated system (management of the editorial workflow, dissemination and long-term preservation of research output) that will make DSpace attractive to internal research staff.

Partnership as a tool for enhancing OA repository services

The final goal of aggregating the scientific works published by Italian researchers will be accomplished within the framework of a cooperation network linking the biomedical research centres operating in the Italian National Health Service. Partner institutions will be able to supply their data on a current basis and browse their own collections hosted as separate items on the central ISS DSpace server. By setting-up and managing such a digital archive ISS aims to create a stable location where scientific information produced by partner institutions can be preserved and disseminated.

A pilot uploading project has been carried out covering items dated 2004 and 2005 supplied by two *Scientific Institutes for Research, Hospitalization and Health Care*, the Centro di Riferimento Oncologico, Aviano, and Policlinico S. Matteo, Pavia. These institutions are part of a network of Italian centres of excellence engaged in research and medical care within the National Health Service. Their main target is translational, i.e. scientific investigation that results in practical outcomes of clinical relevance and vice versa. They produce scientific literature in their respective areas of competence and act as service providers for their own communities.

Institution profile with regard to scientific information services provided

- Centro di Riferimento Oncologico (CRO), Aviano

The Centro di Riferimento Oncologico (<http://www.cro.it>) is one of seven Italian oncology institutes officially designated by the Italian government as Scientific Institutes for Research, Hospitalization and Health Care (IRCCS). The centre was established in 1984, it has 150 beds, about 8000 patients per year between outpatients and inpatients, and 600 staff, about 150 of whom are clinicians and researchers. The CRO and other Italian oncology institutes recently joined the "Alliance Against Cancer" (Alleanza Contro il Cancro, ACC), a network created expressly to share projects, finances and resources in the field of oncology. One of these projects, Azalea, is a multicentre database offering free access to quality information for cancer patients, their relatives and the general public, available in print, electronically or on the Internet. The scientific publications of CRO staff - about 200 per year - are managed by the Institute's scientific library through a specific bibliography software and made available through the Institute's Intranet system. Most of them (about 90%) are published in JCR Impact Factor journals. In the last three years, the total mean value of impact factor publications exceeded 700 points.

-Policlinico S. Matteo, Pavia

The Policlinico S. Matteo (<http://www.sanmatteo.org>) represents a key institution in the history of health care services in the area of Pavia and dates back to the 15th century, when it was founded by a hospital confraternity. At present it consists of more than 50 departments, laboratories and services performing experimental and clinical research in a wide area of scientific fields and disciplines: infectious diseases, transplantations, biotechnology, biomedical engineering, medical informatics and health care management systems including medical statistics and management models. Institutional activities are currently performed in cooperation with a wide network of national and international research institutions.

Research literature produced by S. Matteo research staff, as indexed in Medline and/or ISI Thomson citation databases, is searchable on the Institute's web site. Single journal article citations, together with the impact factors of the relevant journals, are displayed in response to a query entered on a bibliographic database.

The average number of scientific papers published annually by S. Matteo research staff is about 400 items. As for publications (435) issued in 2005, the impact factor amounts approximately to 2058,319 points.

According to the objectives of ISS Project of resources aggregation, two data sets for a total of 634 bibliographic descriptions (respectively 217 relating to 2005 for CRO and 417 relating to 2004 for Policlinico S. Matteo) have been uploaded on the ISS server and made available for browsing through the ISS DSpace interface. A routine procedure will soon be implemented to transfer data automatically from partner institution servers to the central ISS DSpace system.

Future implementations

ISS Researchers should be the first to participate in the process of building the ISS DSpace collection. To do this they need to be able to speak directly to the system so that they can feel that their needs are met. To this end it may be useful to identify early users of the system among research staff, who will need to feel confident of the support offered by repository experts:

- 1) system technicians who maintain and upgrade the DSpace platform and optimise the system by customising its interface in order to make it easy for ISS researchers to archive their works
- 2) information specialists who act as a help-desk for users placing their works in DSpace in digital formats, through the practice of self-archiving.

Self-archiving of post-prints is desirable and is indeed to be expected once the repository has been set up. To achieve this a policy will be needed to regulate placement by authors of their research results in the repository. The ideal option would appear to be a formal requirement for researchers to place their research material in the archive, accompanied by effective author support policies¹⁰. Energetic campaigns to increase authors' awareness are the key to enabling repositories to progress from the planning stage to actual services that will allow scientists to enhance their research impact as never before.

To conclude, the concept to be borne constantly in mind is that the real challenge in setting up an institutional repository is multifaceted, involving as it does a blend of technological, managerial and cultural issues.

References

¹ Open Archive Initiative. Available from: <http://www.openarchives.org> Last visited: 19 April 2006.

² European Commission, DG-Research. Study on the Economic and technical evolution of the scientific publication markets in Europe. Final report - January 2006. Available from: http://europa.eu.int/comm/research/science-society/pdf/scientific-publication-study_en.pdf Last visited: 19 April 2006.

³ National Institutes of Health. Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research. Available from: <http://publicaccess.nih.gov/> Last visited: 19 April 2006.

⁴ Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Available from: <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html> Last visited: 19 April 2006.

⁵ The last follow-up meeting “Berlin 4 Open Access: from promise to practice” was held in Potsdam-Golm, March 29-31, 2006. Available from: <http://berlin4.aei.mpg.de/index.html> Last visited: 19 April 2006.

⁶ Pelizzari E. Academic staff use, perception and expectations about Open-access archives. A survey of social science sector at Brescia university. 2003. Available from: http://eprints.rclis.org/archive/00000737/01/Academic_staff_perception_about_Open_archives.htm Last visited: 19 April 2006.

⁷ Swan A, Brown S. Open access self-archiving: an author study. Key Perspectives:Truro (UK); 2005. Available from: <http://eprints.ecs.soton.ac.uk/10999/> Last visited: 19 April 2006.

⁸ Swan A, Brown S. JISC/OSI Journal authors survey. Report. Key Perspectives:Truro (UK); 2004. Available from: <http://cogprints.org/4125/01/JISCOAreport1.pdf> Last visited: 19 April 2006.

⁹ DSpace User Group Meeting. Bergen, 20-21 April 2006. Available from: <http://dsug2006.uib.no> Last visited: 19 April 2006.

¹⁰ Sale A. Comparison of content policies for institutional repositories in Australia. First Monday 2006;11(4). Available from: http://firstmonday.org/issues/issue11_4/sale/index.html Last visited: 19 April 2006.