

The potential of institutional repositories in fulfilling current research communication needs

B Jayashree^{1*} and JDH Keatinge²

1. International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502 324, Andhra Pradesh, India
2. AVRDC – World Vegetable Centre, PO Box 42, Shanhua, 74199 Tainan, Taiwan

*Corresponding author: b.jayashree@cgiar.org

Citation: Jayashree B and Keatinge JDH. 2008. The potential of institutional repositories in fulfilling current research communication needs. *Journal of SAT Agricultural Research* 6.

The way that scientific research is conducted today is quite different from a few years ago when digital communication was unavailable. Collaboration has become the driver of much of today's science. Research projects are now carried out amongst a network of geographically distant research groups who interact, communicate and share their data through internet technologies such as listservs, blogs, discussion forums, wikis and P2P (peer to peer) file sharing networks. This change in scholarly behavior has also resulted in a formal change in scholarly communication. The concept of refereed journal publications still dominates the researcher's definition of formal scientific communication. The research community understands that scientific information which is not peer reviewed may not be taken seriously; scientists discriminate between what is put on blogs and what is published in formal journals.

But, is traditional journal publishing sufficient in the context of today's research? It is now becoming clear that traditional journals have several deficiencies. They cannot adequately deal with large datasets or non-textual data that are communication units themselves. With a paper medium they are difficult to distribute, archive and duplicate, requiring redistribution points in the form of libraries. Besides, they are also much slower in publishing, given the peer review and printing process and delays between acceptance and publication. In a collaborative network based research scenario, early publication and dissemination of results speeds up the process of knowledge extraction through mining of datasets released. In today's context, a journal should allow preprints of publications, publishing of raw datasets, publication of simulation exercises and software besides postprints. A preprint is any version of an article prior to peer review and publication, usually the version submitted to a journal. Postprint articles could be of two types, those that have been peer reviewed but not copy edited and those that have been both peer reviewed as well as copy edited.

All of these could be regarded as independent units of communication and could be combined together as a compound formal unit of communication.

The Open Access (OA) movement focuses on some of these inadequacies of the traditional publishing system and offers two different solutions. The 'self-archiving' school strives for a scholar's right to make traditional journal publications freely available in an open repository. The other is the promotion and encouragement of journals that are free to all those who wish to access them. Open Access literature is digital, online, free of charge and free of most copyright and licensing restrictions. Open Access is also a form of license that allows other people to access it with less restrictions than normal more 'closed' access. What OA can allow an author is some copyright, peer review, print, preservation, prestige and career advancement and indexing similar to what one expects with traditional publication. The only substantive difference is that the reader does not face an access barrier and an intermediary may not get paid. When the copyright holder consents to OA, he/she can consent to allowing unrestrained reading, copying, sharing, storing, printing, searching and linking to the full text of their work. The OA movement itself has been controversial with vociferous proponents and opposition from the publishing industry. The Co-founder and Director of UK-based scholarly publishing consultancy Key Perspectives (KPL) Dr Alma Swan recently (March 2008) submitted a report to the Scholarly Communications Working Group as part of a JISC (Joint Information Systems Committee) commissioned study (Swan 2008). The JISC is a UK-based initiative funding a wide number of activities on information technology aspects in education and research. Swan's study states some interesting facts about researchers' acceptance of the OA movement. According to this report, researchers complain about restricted access to information yet their own dissemination behavior does not match up to their access expectations.

The reasons reported as to why researchers do not make their work OA are many.

- Either they are unaware of the concept; or are aware but uninformed or unsure of how to provide their work in OA.
- They may be ill informed and/or fear copyright and prior publication issues.
- There are those that think that placing their work on their websites is an adequate substitute for depositing it in a repository and have a poor appreciation of what institutional repositories are trying to achieve.
- Many do not understand data on the impact of their own work and how their performance measures up against their peers using new metrics.

Another important reason not mentioned in the report but contributing to the non-acceptance of the OA movement is that a scientist's performance assessor gives more value to traditional high impact peer reviewed journal publications.

This article is a continuation of the earlier educative narrative on the changing nature of scholarly communication systems (Jayashree and Keatinge 2007), with an emphasis on the role of institutional repositories as a mechanism to achieve OA. Opting for OA involves effort at two levels – operational and policy. Operationalizing OA involves adopting tools, software and processes that help make information OA. At the level of policy, the university or research institution puts in place the processes and procedures to fulfill its OA mandate, giving data producers some incentive to provide their research outputs OA and meeting funding agency OA requirements.

Journals vis-à-vis repository publishing

Both OA journals and OA archives are meant for preservation, dissemination and indexing of content. Major differences are that OA journals conduct peer review while an archive can contain peer-reviewed content. The OA archives may be discipline-based or institution wide (institutional repositories). The OA archives need not contain only journal publications (preprints or postprints), but also theses, course materials, data files, learning objects or any other kind of record. Since OA archives carry peer reviewed articles and their preprints, authors may have several fears about submitting their finished manuscripts into the public domain. Authors depositing preprints still hold the copyright for the articles that they have finished writing. If the author transfers copyright of the postprint to a journal, he/she needs the copyright holder's permission to deposit the article in the institutional archive. Authors can overcome problems

with copyright through use of an 'author addendum' – a legal instrument that allows the author to keep key rights to his/her articles. It is a free resource developed by SPARC in partnership with Creative Commons (<http://www.creativecommons.org>) and Science Commons (<http://scholars.sciencecommons.org/>) and can be used by the author when signing the copyright agreement with the journal publisher. Where the journal does not allow the author to deposit a postprint in a repository, the author can still archive the preprint and differences between preprint and postprint. Most journals now allow deposit into repositories of postprints prior to copyediting.

In a traditional publishing system, the scholarly process consists of registering a manuscript, certifying it, disseminating, archiving and rewarding the author(s) for communicating his/her research findings (Roosendaal and Geurts 1997, Van de Sompel et al. 2004). The process of registration records the date of receipt of the article, which allows claims for precedence for a scholarly finding. The peer review process carried out by the journal takes care of certification, namely verifying the claims made in the paper. When the article is published the process of creating awareness is fulfilled, and the journal itself is archived somewhere in the shelves of the library. The author derives his reward from the metrics associated with the class of journal in which he publishes. Authors write for 'impact' and any rewarding system is largely based on their performance in the journal system. Institutional repositories can play a very similar role as will become clear from the example of the arXiv archive – a preprint repository (www.arXiv.org; Ginsparg 2001) that has been used by the physics and mathematics community of researchers since 1991. This repository provides an implementation of most of the functions of the scholarly process and is hosted out of Cornell University, USA.

The arXiv archive allows scientists to deposit manuscripts and the deposit is registered as first claim on the part of the author. The manuscript in the archive is now open to peer review by the physics community at large; peers may review the article, and others may read and rate the article. The author can then communicate his/her manuscript to an established physics journal, which is mostly the case, and the manuscript now goes through the traditional peer review process. The author benefits from the opinions of his/her peers and other interested colleagues, which he/she can then incorporate before submitting to an established journal. This completes the certification and awareness function of the scholarly process. Search engines (google) index the content of arXiv and alerts are also sent to interested scholars. Journals that publish articles archived in arXiv enhance awareness; so also do citation services (citeBase) as well as overlay journals (SIGMA) that select manuscripts

from arXiv. Overlay journals are virtual journals that provide a list of evaluated and commented links to full text articles held elsewhere (in other words point to full text articles held in a repository) (Enger 2005). The arXiv archive is an important and premier example of an OA archive that has no direct paper precursor and is entirely scientist driven. Similar examples are cogPrints hosted by the University of Southampton in UK (<http://cogprints.org/>), focusing on papers in psychology, linguistics and neuroscience and modeled after arXiv. There are many more such as RePEc (for papers in economics) besides other archives for theses and dissertations.

Operationalizing the OA self-archiving mandate

Institutional repositories – set up by universities, research institutes and libraries – are created to register, disseminate and preserve their scholarly assets. There are several software tools available in the public domain to help institutions implement and manage their repositories without having to get into in-house technical development. These are tools developed by members of the OAI (Open Archives Initiative) (<http://www.openarchives.org/>) community. The OAI mandate

is to enhance access to e-print archives as a means of increasing the availability of scholarly communication. The technological framework and standards developed are meant to open up access to a range of digital resources. Thus, where an archive is OAI compliant, it allows harvesting software to locate articles for users from this repository without the user having to know where the repository is located or the nature of its contents (http://www.soros.org/openaccess/pdf/OSI_Guide_to_Institutional_Repository_Software_v2.pdf). An illustration of the interactions between the repository and the role of metadata harvesting service providers in enhanced access and dissemination of repository contents is given in Figure 1. Available OAI compliant tools include: Fedora, jointly implemented by the University of Virginia and Cornell University (Staples et al. 2003) that defines interfaces for administering the repository as well as accessing and disseminating digital content; Eprints.org (<http://www.eprints.org>), developed by the University of Southampton, provides software to run centralized, discipline-based as well as distributed, institution-based archives of scholarly publications; MIT's DSpace digital repository was developed to capture the intellectual output of interdisciplinary organizations and also to allow customization of the system to manage content, authorization and IP (Intellectual Property) issues of the institution implementing it. These are only a

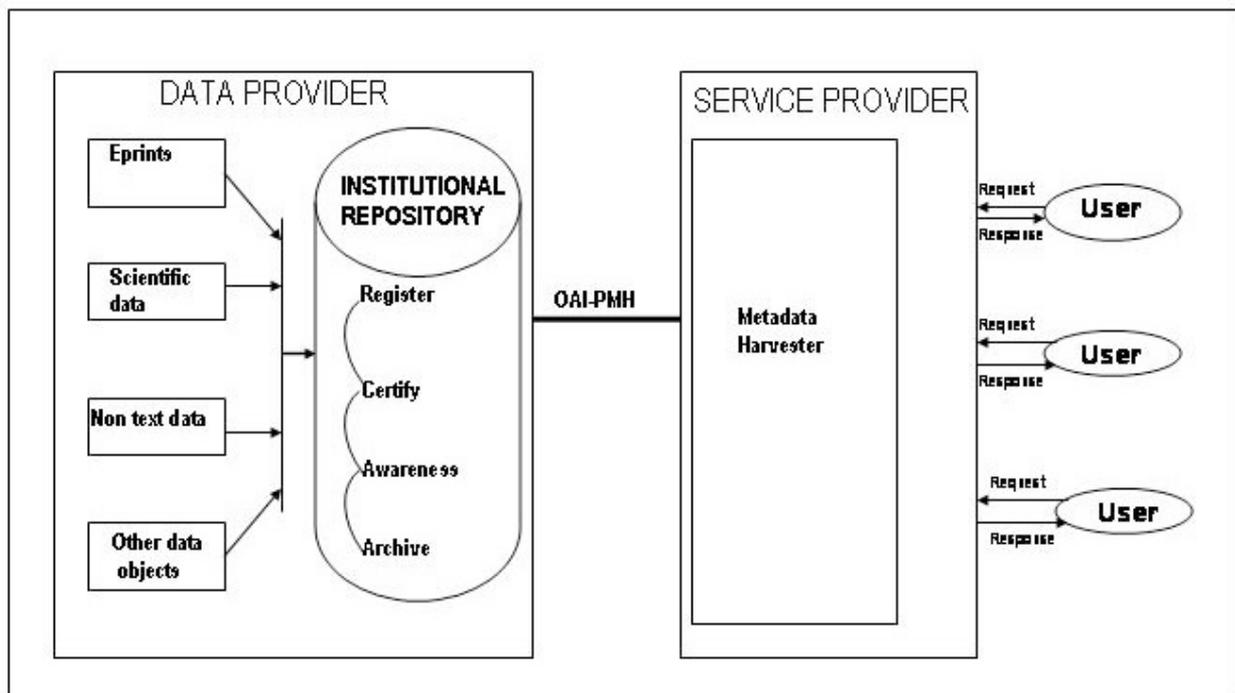


Figure 1. The flow of data, information and knowledge from the data provider to the user. The data provider deposits electronic prints of scientific articles, data, images, learning objects into the repository. An OAI compliant repository can then be linked to metadata harvesters operated by service providers to collect metadata from repositories. The repository is a network accessible server that can process OAI-PMH (OAI-Protocol for Metadata Harvesting) requests and performs most of the functions of the scholarly process.

few examples amongst a growing list of freely available software.

As can be seen from the arXiv example, articles deposited into an archive may also proceed through the other traditional publication process gathering peer review under the auspices of the journal. Institutional archives thus become geographically distributed hubs of scholarly information and depending upon their OAI-compatibility or facilitation with metadata harvesters and publishers, they could allow interoperability. Such interoperability is hoped to stimulate the transition of e-print systems into genuine building blocks of a transformed scholarly communication model. The ICRISAT (International Crops Research Institute for the Semi-Arid Tropics) repository (<http://ngl.icrisat.org/newgenlibctxt/jsp/aportal/cataloguing/index1.jsp>) is powered by the open source Newgenlib software and harvested by the OAIster Search engine. It currently contains a modest number of manuscripts and electronic versions of books and archival material.

Strategies to operationalize OA

Populating the ICRISAT repository requires staff involvement. This calls for a need to understand staff concerns and evolving a strategy that will provide them with a reason/motive/incentive to contribute. Contributions would include but not be limited to research findings and/or research publications that are not published in OA journals.

Repositories are an intersection of interests for different communities: research, digital libraries, e-science, publishing, information management, learning and preservation. Through managing a minimum set of services they can fulfill the data communication needs of today's research providing for faster research turnaround, improved quality through community curation, improved innovation and education, public engagement and improved visibility for the institution itself. This is a relatively new and growing system of research communication that can serve as a complement to traditional publishing methods, providing a means to trace back to the origin of research ideas or trends in a particular area of research. Providing OA does not affect copyright (though licenses must be addressed). Open archives exist in the context of IP legislation, and there may be more coming over the years to resolve ambiguity and dispute as the movement gains momentum, redefining the scope of copyright when it comes to providing access and removing barriers to knowledge for the community. In the coming years, institutional repositories will also help define new alternative metrics to assess the quality of scholarly information in repositories, where not only journal

publications but also documents and draft reports on institutional thinking and research plans are preserved. Self-archiving should be seen more as an opportunity to contribute to the global knowledge base and increase attribution. The aim of this article is to encourage deposit and participation from scientists in the agricultural stream.

Acknowledgments. The authors acknowledge the inputs received from Mr S Srinivas, Library and Information Services, ICRISAT. The authors also thank Dr D Hoisington, Deputy Director General (Research), ICRISAT for reviewing and commenting on the draft.

References

Enger M. 2005. The concept of 'overlay' in relation to the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Masters thesis in Documentation Science, University of Tromsø, Norway. (<http://eprints.rclis.org/archive/00005183/01/master.pdf>)

Ginsparg P. 2001. Creating a global knowledge network. Invited contribution for the Second Joint ICSU Press-UNESCO Expert Conference on Electronic Publishing in Science. (<http://people.ccmr.cornell.edu/~ginsparg/blurp/pg01unesco.html>)

Jayashree B and Keatinge JDH. 2007. Open access: a transformation in scholarly publishing. Published as Editorial in Journal of SAT Agricultural Research, Volume 5, Issue 1. (www.ejournal.icrisat.org)

Roosendaal H and Geurts P. 1997. Forces and functions in scientific communication: an analysis of their interplay. Cooperative Research Information Systems in Physics, August 31–September 4, 1997, Oldenburg, Germany. (<http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html>)

Staples T, Wayland R and Payette S. 2003. The Fedora Project: An open source digital object repository management system. D-Lib Magazine, Volume 9, Issue 4. (doi: 10.1045/april2003-staples)

Swan A. 2008. Key concerns within the scholarly communication process – report to the JISC Scholarly Communications Group. Document accessed from <http://www.jisc.ac.uk/media/documents/aboutus/workinggroups/topconcernsreport.doc> on 3 May 2008.

Van de Sompel H, Payette S, Erickson J, Lagoze C and Warner S. 2004. Rethinking scholarly communication. D-Lib Magazine, Volume 10, Issue 9. (doi: 10.1045/september2004-vandesompel)