

Scholarly communication in a digital world: the role of the digital repository at the Raman Research Institute

Girija Srinivasan, Y.M. Patil and Jacob Rajan

Raman Research Institute,
Sadashivanagar, Bangalore 560 080, Karnataka, India.
{girija, ympatil, jacobr}@rri.res.in

Abstract. Information and communication technologies have changed the landscape of information handling activities. The proliferation of electronic resources and a paradigm shift from print to electronic format of journals, has led to another crisis due to the escalating costs of electronic resources, dwindling library budgets and cuts in journal subscriptions. Libraries have tried to overcome this situation by a dual approach, viz. consortia deals on one hand to access more information with optimal payment and open access scholarly publishing and communication on a worldwide basis on the other hand. In the last few years, the Open Access Movement has been gaining momentum and many journal publishers are also supporting this cause. With the availability of open source software for creating repositories, many institutes have established institutional repositories in support of open access. Realizing the importance and the benefits of such a repository at the institute level, RRI Library has set up the RRI Digital Repository. In this paper, we describe the growth and evolution of the repository, the problems encountered, the solutions explored, the strategies formulated to add relevant content and finally lessons learnt while executing the project.

1 Introduction

With increase in prices of journals, shrinking library budgets, cuts in library subscriptions and shift from print to electronic form over the years, there has been a big challenge before the Indian library professionals to cope with the proliferation of electronic information resources. In order to overcome such piquant situations, there have been sporadic efforts by different groups of libraries to form consortia at different levels in order to have more access to more information with optimum payment. There have been different deals for consortia negotiations at different levels and many consortia models have been evolved. These are all endeavors towards enhancing access to more information at affordable costs. This is one of the approaches to facilitate access to information in a library environment and mostly such efforts are librarian initiated (Patil, 2004, & Patil, et al. 2006). As a matter of fact, the

fundamental problem remains the same. The increase in the cost of electronic resources continues to be much higher than the annual increase in the library budgets

On the other hand, due to the ever-changing scholarly publishing sector and the emergence of the open access movement, the scholarly information communication landscape has been evolving. In the process of scholarly communication, there are certain barriers in the developing countries, viz. pricing barriers and permission barriers. Pricing is the one related to the subscription rates of journals, which are escalating year by year with the result that libraries have to cancel subscriptions to some titles every year. Permission barriers relates to online journals which need licensing for access or pay-for-view for full text access and copyright comes in the way of access. Due to this kind of external forces and inherent internal budgetary constraints, the concept of institutional repository has emerged in order to capture, disseminate and preserve institute's research output in digital format (Chan, Kwok & Yip, 2005).

There are various ways and means to access information published in journals. One is through subscriptions – either individual, institutional or through consortia. All cannot have this at their disposal for accessing information, which has a limited domain, and access is ensured only for those who subscribe to the required journals. The other way for access to articles is through *Open Access*. Open Access can be achieved by authors publishing in open access journals - i.e. journals which allow their articles to be read for free. It can also be achieved by self-archiving i.e. each author hosts his research output on his home page or employer's site or on a central discipline based archive like arXiv so that his research is widely disseminated. All authors may not do this. Hence, library professionals have to take a lead to make the institute's research output in digital format available for access to all through the institution's digital repository and support the Open Access Movement.

1.1 Current scenario

The recent trend for access to information is "*free to read*" based on payments to either publish a paper and make it open-access after publication or sponsor payment for articles already published in any journal.

The American Physical Society has recently introduced a payment scheme for articles called FREE TO READ and any one can sponsor it – including individuals, philanthropists, or institutes. Funding for such projects could also be parents, grandparents, etc.

Similarly, some publishers have provided an option to publish articles against author payment and papers are accessible from the publishers' site for free.

What is the impact of such charges on the subscription models and revenue loss to publishers? Martin Blume, the Editor-in-Chief, APS states that "APS is a financially stable organization willing to take risks to support the community" and it is with the community in mind that APS is offering FREE TO READ. Of course, with this initiative, APS will be able to sustain cost-wise. Access to article is ensured without additional features like links to references and citing articles, which remain part of subscription and are not available through FREE TO READ.

The mandate of the Open Access Initiative is that all those who are involved in handling digital products at the institute's level should go for creating an Institutional Repository projecting the institute's research output and other gray literature and facilitate its access worldwide through the Internet. Keeping the Open Access movement in view, RRI Library initiated action for developing and implementing an Institutional Repository and the details of this are given below.

2 Birth of the RRI Digital Repository

The idea for a digital repository for the Raman Research Institute was born out of the following need:

- We have a large collection of newspaper clippings about Raman dating back to the 1930s. They were in a fragile state due to age and we had to restrict access to them. Therefore, we had all of them scanned so that handling of the originals could be minimized. Since they are valuable resources for historians of science, we were wondering how best we could make them more widely accessible.
- A few years ago, we had begun the practice of collecting all our authors' papers in pdf files as that was very convenient for storing and printing. We also scanned all the other papers that were from older journals/conference proceedings so that we had a complete set of all our scientists' papers in digital format. We were keen to make these searchable and more easily accessible not only on our Intranet but also on the Internet so that we could also support the Open-Access movement (Chan, 2004). So we started looking for a suitable software to achieve this objective.

2.1 Selection of software, trial and customization

The release of two open-source software – Eprints (www.eprints.org) in 2001 and Dspace (www.dspace.org) in 2002 accelerated the open-access movement. After being exposed to them through workshops conducted by NCSI, Bangalore (2002) and by IIT, Chennai (2004), we installed them on a

test server and tried them out. We also started learning about the process by reading about and examining other institutional repositories. Some documents and sites that we found useful are listed in the references.

Eventually, we selected Dspace as it satisfied many of our requirements. The test installation gave us the confidence that we can set up RRI's Institutional Repository and showcase its scholarly work globally.

After moving the trial installation to a dedicated server, we wrote the online help for submission, designed our homepage, and customized the submission page and the item display. One of the authors (Jacob Rajan, RRI's systems administrator) was responsible for installing and configuring the system and also the system maintenance. It took us about a year from the time of installing the test server to launching our repository on the Internet.

3 Policies and guidelines

"Decisions, policies and guidelines have a life of their own. They grow and evolve over time"(Chan, 2005). So it has been with us.

We wanted our repository to have different types of documents – so that not only does it collect, disseminate and preserve the research output of our institute, but it also serves as an information centre about our institute and its founder, C.V. Raman. We felt that setting up an omnibus repository of this kind was important as we could then make available a number of historical documents in a central place which otherwise would be scattered in different publications. The following documents are now available in our repository (as on 28th October 2006):

- Papers by RRI authors in journals/conference proceedings (645)
- Theses (To be added later)
- Annual Reports of RRI (28)
- Newspaper clippings from the Raman archives (707)
- Gandhi Memorial Lectures (12)
- C.V. Raman's collected papers (To be added later)
- C.V. Raman's lectures and miscellaneous writings (9)
- Biographical sketches and articles on Raman (5)
- Photo gallery (To be added later)

Any faculty member or student may contribute documents. We wrote an online document "Submission Guide" available on our Repository homepage to help them. It was decided that the project team would check all submissions for copyright clearance, correct format, edit the metadata if necessary and then archive it. Our policy was to include the full-text of the article in the format which was permitted by the publisher. Where it was not possible for the full text of the publication to be made available in the repository, we decided to make only the bibliographic details available. If

possible, we would provide a link to an alternate location, e.g. to a publisher's site. In this case, only subscribers to the journal in question will be able to follow the link and access the full text. In the case of theses, we plan to make only those open access where the author gives us permission. The others will be restricted to access via Intranet.

Regarding articles in astronomy and physics, many authors routinely submit their articles to arXiv before publication. While discussing the inclusion of these preprints/postprints into our repository, they felt there was no point in duplicating what was available on arXiv. The authors felt that giving a link to the arXiv site and journal site was sufficient.

It was also decided to involve all the library staff in this project and contribute to its growth. Guidelines were set down for the project team so that metadata could be created with consistent standards. The newspaper clippings proved to be the most difficult to upload as we not only had to give keywords – but also had to write an abstract for each one ourselves – which meant that we had to read each clipping. For journal articles, author names were rendered in a standard format and we decided to take the keywords given by the authors themselves in the article. A paper may have authors from different departments – this was solved by mapping them to each department.

4 Content development: the biggest challenge

“Recruitment of content, not technology, is the greatest barrier to success” (Gibbons, 2004). This points to a worldwide phenomenon where content recruitment is a slow and painful process. Experience of other repositories (Chan, 2005, Foster, 2005, Mackie, 2004, Mark & Shearer, 2005) has indicated that relying on self-submission from the authors alone is not sufficient. The reasons for poor response from authors are:

- Authors are not willing to take on additional work;
- They are afraid of violating copyright. They do not read the copyright transfer form fully before they sign and are not aware of their rights;
- Fear of plagiarism;
- Benefits not immediately obvious – why deposit in an institutional archive when their papers can be found on arXiv?

As a result, a multi-pronged approach is advocated and practiced by many. Mark and Shearer (2005) have identified these strategies and classified these into 6 categories.

- General promotional activities
- Mediated depositing services
- Content harvesting
- Researcher bibliographies
- Usage information

- Self-archiving policies

4.1 General promotional activities

Promotional activities are important to make authors aware about our service. “Academics have to hear about your institutional repository service many times, over a period of time, and from several sources (print, online, in person). A good rule of thumb is that someone needs to have been exposed to your service seven times before they are fully aware of your service. Be sure to outline explicitly the benefits of your service to academics.” (Barton, 2004). In other words, we must do some marketing to advertise the service on campus.

After we had a test installation running, we spoke to members of the library committee and also to several authors to keep them aware about this project and to seek their feedback. They encouraged us and felt that this would be a very useful service. So we decided to go ahead with the repository and started adding all the newspaper clippings and journal articles from the Indian Academy of Sciences Journals. Thus, when we made it available on the Internet, we already had a critical mass of content with which to market our repository. We also wrote online documents (available on the repository homepage) which explained the purpose and benefits of the repository and gave details about how authors could submit documents.

We then announced the repository to everyone at RRI through email and notice board and requested faculty to participate in submitting their papers. We offered to help them with copyright issues and also formally requested their consent to host all their older papers if the publisher permitted it. The response to this was quite good – many wrote back giving their permission and appreciated this new service. However, after this initial enthusiastic response, very few have actually submitted their new papers. This is because a cultural change in the mindset of the authors is necessary to make this a part of their academic life. (Nixon, 2002)

We have been following this up with periodic emails giving a report on the progress made and asking for postprints of all the articles published by the researchers. The library maintains a database of our research publications. So, whenever we come across a new paper which can be hosted in our repository, we do it ourselves if the publisher’s pdf is permitted to be archived. Otherwise, we personally ask the authors for their postprints. We will have to keep up this interaction with the faculty as long as it is necessary.

4.2 Mediated deposit

At RRI, we have been relying more on mediated deposit by the project team for most of the content. In this process, the project staff ensure copyright clearance for the full text, add metadata and the appropriate copyright notice

and publisher statements to each record before making them open-access on our repository.

When we first set up our DSpace installation, we tested it by adding our newspaper clippings. In the meantime, we wrote to the publishers of the newspapers and received their permission for hosting it on our repository. So when our server was ready, we had a ready set of documents which we could host on our repository.

After we had uploaded all our newspaper clippings, we turned to the journal articles. As mentioned earlier, we had a complete set of pdf files of all our research articles from 1972 onwards. We first decided to identify the publishers who give permission to host the publisher's pdf on repositories. We were keen to host this version as this is the definitive version and also the only one which we had. It seemed appropriate for us to start with the papers in the journals of the Indian Academy of Sciences (founded by Raman) who readily gave us their permission. Incidentally, all their journals are open access. We then took up the articles published in journals of the American Physical Society, the American Institute of Physics and IEEE who also gave us permission to do so.

Understanding and interpreting the copyright policies of publishers requires a lot of time and patience as each one is different. The SHERPA/ROMEO list is very useful, but we still went to the publisher's website to ensure that there has been no change in their policy. We have been adding documents to our repository according to their policies keeping in mind their embargo conditions. For example, the American Astronomical Society publications have a three year embargo. So papers prior to 2003 have been uploaded and every year we plan to add incrementally one more year's papers.

4.3 Content harvesting

This simply means searching for articles which are already open-access in the web and adding them to the repository. For example, many authors have personal websites and post their articles on them. We have added some articles in this way – but we found that many were from publications that did not allow self-archiving or whose copyright policy was not clear. So in these cases, we have added the relevant metadata and simply linked to the author's websites for the full-text.

There are some publishers who have not given us permission to host their version on our repository. However, the older articles in their journals are available free to download from the publisher's website or NASA ADS website. In these cases we add the metadata and link to the relevant website for the full-text. Some conference proceedings are also available at the NASA

ADS site – we plan to treat them in the same way as it is very time-consuming to ask each publisher for permission for one or two articles.

As mentioned elsewhere, many articles from RRI are available open access in arXiv also. We plan to take advantage of this and add all these through linking in future.

4.4 Usage/citation information

If we could show the authors how many times their paper has been downloaded from the repository, the benefits of adding their papers will immediately become obvious to them. Dspace offers a facility for tracking the number of downloads for each paper but we are not very happy with it. We would like to know how many times each paper is downloaded and by which country. We plan to implement the statistics add-on developed by the University of Minho at some future date. Our present statistics for the highest download seems to be for the newspaper clippings – a little surprising but it justifies the trouble we took to host them.

4.5 Self-archiving policies

A few universities have implemented ‘self-archiving policies’ or ‘open-access’ policies to support the use of institutional repositories. However, this is a very complicated issue as we cannot dictate in which journals authors should publish. Fortunately for us there is a strong self-archiving tradition established in the fields of physics and astronomy in centralized arXiv since August 1991. This means that many RRI articles are already available open-access in some form or the other.

5 Conclusions

Institutional repositories are still comparatively new but a number of important benefits have already become apparent. Several studies have shown that they facilitate more timely open access to research and that they maximize the research impact of archived publications. Thus they play a major role in the modern scholarly communication process. In our case, since astronomy and physics are major fields for us, there is already a strong tradition of self-archiving in arXiv. Astronomers also have a fantastic resource in the NASA ADS which has scanned and made open-access all the older articles from the major astronomy journals and conference proceedings. Hence, we do not know if making astronomy papers open-access in our repository will have much impact on the scholarly communication scene. But our repository has a number of papers which are not available open-access normally. And we hope that they will grow in the years to come and make an impact on the field due to their visibility and accessibility. Briefly, we can sum up the benefits of the RRI Digital Repository thus:

- It supports the open access movement and plays a role in the scholarly communication process;
- It collects and preserves all the publications of the institute in a central place - thus making them available to our students and researchers whenever needed. It thus enhances the teaching, learning and research of the institute;
- It showcases all the research publications and gray literature of the institute and enhances the prestige of the institute and the visibility of the faculty;
- It adds one more dimension to the role of the library in making available this collection to its users. By providing access not only to the research publications of our institute, but also to the historical records of the institute (Annual reports, newspaper clippings) and providing information about our founder and his work, we believe that we are providing an additional and important specialized information service.

Our repository is still in the early stages and the benefits of this are not yet obvious to everyone. But we are sure that once they become clear to our researchers, their initial apathy will disappear and they will participate actively in building and sustaining it.

We also believe that the Institutional Repositories address a very strong need in today's world and therefore there will be a steady growth of such repositories in the country and in the world in the days to come. And perhaps they will be instrumental in completely changing the scholarly communication landscape.

References

- [1] Barton, Mary R.(2004). *Creating an Institutional Repository: LEADIRS Workbook*. Cambridge, MA: MIT. <http://dspace.org/implement/leadirs.pdf>
- [2] Chan, L.(2004). Supporting and enhancing scholarship in the digital age: The role of open-access institutional repositories, *Canadian Journal of Communication*, 29, 277-300. <http://hdl.handle.net/1807/2786>
- [3] Chan, Diana L. H., Kwok, Catherine S. Y. & Yip, Steve K. F. (2005): Changing Roles of Reference Librarians: The Case of the HKUST Institutional Repository. *Reference Services Review* 33, no. 3268-282. <http://repository.ust.hk/dspace/handle/1783.1/2039>
- [4] Dspace: (<http://www.dspace.org/>)
- [5] Eprints: (<http://www.eprints.org/>)
- [6] Foster, Nancy Fried, & Gibbons, Susan (2005). Understanding Faculty to Improve Content Recruitment for Institutional Repositories. *D-Lib Magazine* 11, no. 1. <http://www.dlib.org/dlib/january05/foster/01foster.html>
- [7] Gibbons, S. (2004). Benefits of an institutional repository, *Library Technology Reports*, Vol.40, No.4, p11-16.
- [8] Lynch, Clifford A. (2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL*, no. 226. <http://www.arl.org/newsltr/226/ir.html>
- [9] Mackie, Morag. (2004). Filling Institutional Repositories: Practical Strategies from the DAEDALUS Project. *Ariadne*, no. 39. <http://www.ariadne.ac.uk/issue39/mackie/>
- [10] Mark, T. & Shearer, K. (2005). Institutional Repositories: A Review of Content Recruitment Strategies, *72nd IFLA general conference, Seoul*. <http://www.ifla.org/IV/ifla72/index.htm>
- [11] Nixon, William J. (2002). The Evolution of an Institutional E-Prints Archive at the University of Glasgow. *Ariadne*, no. 32. <http://www.ariadne.ac.uk/issue32/eprint-archives/>
- [12] Open Archives Registry: <http://archives.eprints.org/>
- [13] Patil, Y.M. (2004). Managing change: Consortia efforts in IT environment. In *Dr. P.S.G. Kumar Festschrift – Library and information profession in India, Vol.1. (Pt. II): Reflections and redemptions*. Eds. C.P. Vaishishth & M.P. Satija (p465-486). New Delhi: B.R. Publications.
- [14] Patil, Y.M., Birdie, Christina, Bawdekar, Nirupama, Barve, Sunita, Anilkumar, Nishta. (2006). Indian consortia models: FORSA libraries experiences. *Paper presented at the LISA V Conference: Common challenges, uncommon solutions*. Cambridge, Massachusetts, U.S.A., June 18-21.