

Second draft of Arun's paper for Berlin-4 [27 March 2006]

Open Access – Current Developments in India

Subbiah Arunachalam

Distinguished Fellow

M S Swaminathan Research Foundation

Chennai 600 113, India

arun@mssrf.res.in

Background

India, the second most populous nation in the world, is emerging as an important player in the world economy and geopolitics. In the nearly six decades since Independence, India has made considerable progress. A number of leading corporations, especially in the areas of automobiles, information technology and chemicals, have set up shop in India for manufacturing, business process outsourcing and R&D. Advanced countries look at India as a huge market to be tapped and a reservoir of English-speaking workforce that can be hired at a fraction of the cost they pay as wages in their home countries. About a million people work in software industry alone. And now India is increasingly looked up to for outsourcing R&D. In the past few months, many heads of states and governments – including President Bush - came calling and President Bush even spoke about the rather sensitive subject of cooperation in nuclear energy. Both the Vice chancellor of Oxford in the UK, the Rt Hon Chris Patten and the President of Harvard University Lawrence Summers in the US visited India recently and are keen to set up centres of excellence devoted to Indian studies. Indeed Harvard is planning to institute a dozen chairs in the new centre.

Despite a long history of science, scholarship and philosophical inquiry dating back to millennia before the emergence of modern European civilization, India is struggling to keep pace with the West in science and technology. Although there are about 300 universities, and about the same number of government funded research laboratories under agencies such as the Departments of Atomic Energy and Space and the

Ministries of Defence, Agriculture, Science & Technology, and Ocean Development, India's research output in science and technology, as seen from the *Web of Science*, is barely 2.5% of the world's journal literature. What is more, in none of the subjects Indian papers on the whole are cited as often as the world average. It will not be wrong to conclude that India is contributing to growth of knowledge in the sciences sub-optimally. There is a crying need for strengthening higher education (and, indeed, education at all levels) and promoting excellence and innovation in research. India is investing millions of dollars to set up three institutions of excellence in science on the lines of the Indian Institute of Science in Bangalore and six world class medical colleges and hospitals of the quality of the All India Institute of Medical sciences in underserved regions.

Information is a key input for both higher education and research. Both making available information produced worldwide to Indian researchers, teachers and students and disseminating information produced in Indian laboratories to the wider world are equally important. One is concerned with access to information and the other with increasing the visibility and impact of research performed in India. Fortunately, recent developments in technology, in particular the Internet, World Wide Web and affordable bandwidth, have made it possible to achieve both these goals. But we seem to be facing problems in moving from the realm of possibility to the realm of reality. As not many universities and research laboratories around the world are making their research publications freely available either through (central or distributed) archives or through publishing them in OA journals, and as Indian libraries are unable to subscribe to even a small fraction of the number of toll-access journals subscribed by well-endowed libraries in richer countries, access to relevant research information continues to be a major problem for Indian researchers. In 2002, the Indian Institute of Science, India's best-known higher education institution in science and engineering, subscribed to 1381 print journals of which around 200 titles were also accessible online. Currently, professors and students of the institute have access to 8,350 full-text journals through the INDEST consortium, and an additional 600 journals through library subscription. Certainly that is a big improvement, but if we compare universities in the West such as Harvard and Stanford, scientists in Bangalore (and elsewhere in India) are tremendously handicapped. And as a large proportion of papers from India are published in low visibility journals,

either Indian or foreign, they go unnoticed. Often scientists working in one Indian laboratory may not be aware of publications from another Indian laboratory in the same subject, although both of them are supported by the Indian taxpayer. Added to this is the Harvard vs. Hyderabad syndrome: If there are two equally good papers, one from Harvard and the other from Hyderabad, often people will prefer to read (publish, cite) the one from Harvard. It is all the more necessary then for scientists in India (and in other developing countries) to push even harder and make their work easy to access. That is why open access archiving is very important.

Current status

A few of us in India are active in OA advocacy and have joined OA advocates elsewhere in pleading for increased OA archiving around the world. Access to freely available information is proportional to what is available in OA archives. The oft-quoted figure that only around 15% of papers are being archived is not good news at all! We would like to see this figure rise rapidly. There is nothing much people in India and other developing countries can do to promote OA in the advanced countries of the world where the bulk of research is carried out, except join hands with OA advocates elsewhere and add their voice to the worldwide chorus.

But on the dissemination front, India can do a great deal. We can open up our journals and we can set up central and distributed archives. Unlike in the West, most journals in India are published by the government (CSIR, ICAR, ICMR), science academies (IASc, INSA, NAS), and professional societies (e.g. Geological Society of India). Journals published by commercial firms are fewer, but the number is increasing. The science academies get grants from the government and many professional societies can also obtain some government support for their journals.

OA journals - Currently close to a hundred journals are OA. Not one of them follows the author-pays model of BMC and PLoS. The journals published by the Indian Academy of Sciences (11), the Indian National Science Academy (4), and the Indian Council of Medical Research (1) are all open access. Besides, there are two agencies, which bring out OA versions of journals published by others, mostly professional societies. MedKnow Publications [www.medknow.com], a private firm in Bombay,

brings out both the print and electronic (OA) versions of about 30 journals in the area of medicine. The National Informatics Centre, a government outfit located in New Delhi, brings out the electronic (OA) version of 38 biomedical journals under the medIND project. Both these agencies are likely to add more journals to their portfolio. [NIC also has a service called IndMED, a database covering prominent peer reviewed Indian biomedical journals that facilitates quick and easy access to Indian literature.] The Calicut Medical College, Kozhikode, is home to a few OA medical journals. The OA version of a few Indian journals are brought out by Bioline International. While some of these OA journals are pretty easy to use and highly user-friendly (e.g. MedKnow journals), others are less so. For example, there is considerable scope for improving the ease of access of the journals published by the Indian Academy of Sciences, a body of India's leading performers of scientific research, but despite repeated efforts I have not been able to persuade them to improve the web presence of their journals.

MedKnow has carried out some very useful studies. For example, the *Journal of Postgraduate Medicine* (JPGM) gets close to a hundred thousand hits every month and an equal number of article downloads, the circulation of the print version has increased, number of papers submitted for publication (especially from outside India) has increased manifold, and above all its papers are getting cited far more often now than before and if it is indexed today in the *Web of Science* and *Journal Citation Reports* it will have the highest impact factor among all Indian journals indexed. These studies need to be publicized widely, and other Indian OA publishers should be encouraged to carry out such studies.

Journals published by the government sector have a long way to go. In the past, three journals published by CSIR were made available electronically by Bioline International, probably against subscription, but CSIR has discontinued this agreement. The ICAR journals are available only in print. All these journals have a very poor circulation.

OA archives – The Institute of Mathematical Sciences, Chennai, was the first Indian institution to have actively taken part in open access archiving. Years ago they set up a mirror site for the then Los Alamos-based arXiv. The Indian Institute of Science was

the first in India to set up an institutional open access archive. Why did OA happen first at IISc and not anywhere else in India?

The National Centre for Science Information located at the Institute and chaired by Prof. N Balakrishnan (who was also instrumental in setting up the Super Computer Education and Research Centre of the Institute) had the necessary skills and infrastructure and a dedicated OA champion (the late Dr T B Rajashekar). Prof. Balakrishnan, who had since become Associate Director of the Institute, wielded considerable influence and he was ready to support the OA initiative. In fact, he was also moving ahead with the Million Books Digital Library project in collaboration with the Carnegie-Mellon University. Dr Rajashekar, unlike most others in the library profession in India, believed in action rather than merely talking. He had a research-oriented mind and was methodical. He took one job at a time and was keen to complete it. He was supported by the Department of Scientific & Industrial Research through its National Information System for Science and Technology (NISSAT) programme, and he was training about half a dozen science librarians every year in the application of information technology in information services. About five years ago, I conducted two three-day workshops on electronic publishing with Leslie Chan and Barbara Kirsop as the main faculty with the support of NCSI, IISc and the Indian Academy of Sciences at IISc, with financial support from IDRC, British Council and INASP. We invited professors of IISc and the Indian Statistical Institute to speak at these workshops. One of the participants (Dr D K Sahu of MedKnow Publications), I am happy to report, turned to be outstanding. He is not only bringing out 30 OA journals but also is carrying out studies that prove the value and advantage of going OA for developing country journals.

Today the GNU Eprints archive at IISc has more than 3,700 papers. Unfortunately, not all of them are full text. Also, many of them are uploaded by the NCSI staff and project assistants and not by the authors, many of them not really interested or even indifferent. This apathy of scientists seems to be a universal problem.

Today about twenty institutions in India have an archive of their own. The National Chemical Laboratory has an archive, but populated mostly with doctoral and masters

dissertations. Only a few scientists have come forward to place their papers in the NCL archive. NCL also has a repository of data on industrial micro organisms and an OA biodiversity database. The National Institute of Technology at Rourkela has an archive and the Director and Librarian of NIT-R are seeking the consent of the faculty for mandating OA for all research papers published from the institute. The Indian Institute of Management at Thiruvananthapuram has a Dspace repository, but author reluctance seems to be rather high. The Raman Research Institute in Bangalore has a repository which has research papers, dissertations, press clippings and a few other categories all in one place. The National Aerospace Laboratories in Bangalore also has a mixed-bag repository. Although such omnibus repositories are inferior to archives dedicated to research papers, one has to be content with the thought some archive is better than none! The Indian Institutes of Technology at New Delhi and Bombay have started some work.

The Documentation Research and Training Centre at the Indian Statistical Institute in Bangalore has set up a central Dspace archive for papers in the field of library and information science. Dr A R D Prasad, an Associate Professor at DRTC, is on the governing council of Dspace and is perhaps among the most knowledgeable Dspace experts in the world. He has conducted many workshops on setting up Dspace repositories. Another institution providing centralized services is the National Informatics Centre in New Delhi. Specialising in biomedical information, NIC has three services: As mentioned earlier medIND provides open access to journals. Besides they have IndMED, a bibliographic database for biomedical sciences and OpenMED, a central archive for biomedical papers. Mr Sukhdev Singh of OpenMED is a frequent contributor to the 'oa-india' dgroups discussion list. OneWorld South Asia, an NGO, has set up an archive for collecting papers and reports in the field of ICT-enabled development.

In mid-March NCSI-IISc created CASSIR, a cross archive search service for Indian repositories [<http://casin.ncsi.iisc.ernet.in/oai/>]. This service will harvest metadata as per the OAI-PMH protocol (<http://openarchives.org>) from the registered OA repositories in India, and provide a web-based search/browse service over harvested metadata. As of 27 March 2006 the tally was close to 7,850 papers in 13 archives. Not all of the papers are full text though.

For a country of the size of India, what we have done so far is utterly inadequate.

Digital libraries – The Million Books Digital Library project [<http://dli.iiit.ac.in>] in which Prof. Balakrishnan of IISc leads the Indian part is the most important programme in this category. More than twenty Indian institutions, including the Presidential Palace, are involved and books, both old and new in many languages including Sanskrit and Tamil, are being digitized. So far, more than 57,000 books in 14 languages have been scanned. The project is also developing software for transliteration and translation into many languages so one can get to know the content of a book in a language one is not familiar with.

The University Grants Commission has funded a project at Mysore University to digitize theses. Called Vidya Nidhi, this project is led by Prof. Shalini Urs who learnt about thesis digitization when she was a Fulbright Fellow at Virginia Tech. Currently Vidyanidhi has more than 100,000 metadata records in its database including 1,000 records in Kannada and 7,000 records in Hindi. So far only seven universities and a research institute have joined this programme.

Secondary service for OA material – A few weeks ago Informatics India, a Bangalore-based company, came up with Open J-Gate, a search service for all material available via open access. It covers about three thousand serials, about 1,500 of them S&T journals. The company also has another priced product called J-Gate and it covers many thousands more journals. DOAJ, Lund, is already in discussion with Informatics India on possible cooperation.

Mahiti, a FOSS based ICT support agency for civil society, academic and government organizations, has developed an offline text only version of Wikipedia.org available on DVD for Rs 2000/- per copy for the Indian market under FOSS license. In other words you can buy one copy, make 100 copies and distribute for free to your friends. Offline Wikipedia is a DVD that contains the contents of entire Wikipedia.org. Wiki contents can be browsed offline without actually connecting to Internet through a PC. It is very useful for academicians, researchers, schools, NGOs, colleges or any one who wants the entire Wiki for their personal usage. Also in places where Internet bandwidth is low this would be of great help.

Features :

- * All the Articles on Wikipedia are available on the DVD (Text only, No images)
- * Wikipedia contents need not be copied into hard disk to browse the contents
- * Simple configuration steps to view the contents of Wiki either directly from DVD or from local hard disk
- * Very basic System requirements and works on PCs, laptops which runs Windows Xp, 2000 or NT
- * Any browser can be used to view the contents IE, Firefox, Mozilla, etc.

Mahiti has also brought out a CD containing Eprints, Dspace and Fedora software and it is given away free.

Discussion

One is happy to note that

- many independent journals are opting to become OA and the project “Scientific Journal Publishing in India: Indexing and Online Management” carried out by NCSI-IISc with financial support from IDRC is covering them.
- Mahiti is supporting open access by distributing a CD containing three major OA software.
- No OA journal in India is charging author side fees
- Informatics India, a for-profit company, is providing free access to Open J-Gate, a service for searching and accessing articles in OA journals
- Interest in setting up OA archives is picking up; now there are at least 17 and a few more are on the anvil
- Success of OpenMED may lead to setting similar central subject-specific archives
- Many experts – both Indian and foreign – are ready to give their time generously to promote OA. Two years ago Leslie Carr of Southampton and Leslie Chan of Toronto came all the way and conducted two three-day hands-on workshops on Eprints. Alma Swan toured India for a week and met key people in Bombay, Hyderabad, Bangalore and New Delhi and also gave the keynote address at the special session on open access held as part of the 93rd Indian Science Congress

Meeting. Jean-Claude Guedon came to India to deliver the annual lecture of Informatics India and to formally launch Open J-Gate.

- Early this week both Mr Sam Pitroda and Dr Pushpa Bhargava, chairman and vice-chairman of India's Knowledge Commission, have promised to look into our suggestion to mandate OA for publicly-funded research.
- INSA is a signatory to the Berlin declaration and a Fellow of INSA, Dr Kishan Lal is an active supporter of OA for research publications and data. He is influencing policy at CODATA.

At the same time, one is concerned that

- No funding agency or research institution in the country has so far mandated OA
- Author are still not coming forward to self-archive; most archiving is done by intermediaries (such as library staff and project assistants)
- There is still reluctance among institutions, scientists and librarians to adopt open access archiving; there appears to be a mental block more than technical hitches.
- There appears to be apprehensions in the minds of scientists about copyright and IPR. Many scientists are unable to distinguish between publishing and archiving.
- There appears to be much talk and very little action.

We need to continue our advocacy programmes aimed at both the scientists and the policymakers. We need to encourage editors, publishers and scientists to carry out studies similar to the ones carried out by D K Sahu on the impact of OA on journals, their visibility, subscription and impact. We should enlarge the membership of the electronic discussion list 'oa-india'. Right now most members are from the library and information science background. We need to reach out to scientists in the laboratories.

It will be useful to hold international OA conferences in India in collaboration with the InterAcademy Panel, TWAS and Unesco. Indeed, a high-level meeting of policymakers from India, China and Brazil will be worth pursuing. If these three leaders of the developing world come up with strong OA policies, it will help us win support for OA elsewhere in the world.